Journal of Scientific Computing and Machine Learning, vol. 1, pp. 37–42, 2025
Submission Date: January 15, 2025 Acceptance Date: January 26, 2025

PM 2.5 Advection-Diffusion with Multiple Sources and LSTM Neural Network Surrogate Model Optimization

Kevin Yotongyos[1] and Somchai Sriyab[1,2]

[1]Ph.D Degree Program in Mathematics, Faculty of Science, Chiang Mai University
[2]Department of Mathematics, Faculty of Science, Chiang Mai University
kevinyotongyos@gmail.com

Abstract

PM 2.5 (particulate matter with a diameter of 2.5 micrometers or smaller) has become a significant environmental issue, particularly in many Asian countries. This research employs the advection-diffusion equation to model the dispersion and transport of pollutants, accounting for multiple sources within the same region with varying emission time windows. Optimizing emission schedules to minimize the maximum concentration at any given time using the advection-diffusion equation is computationally intensive. To address this, an LSTM neural network surrogate model is used to efficiently search for the optimal emission schedule. The best emission schedule is applied back to the advection-diffusion model, and its performance is compared to the initial unoptimized schedule. Results verify that the surrogate model is effective for optimizing emission schedules and reducing maximum concentrations.

Keywords PM 2.5 · advection-diffusion · LSTM · neural network · optimization

1. Introduction

In recent years, PM 2.5 (particulate matter with a diameter of 2.5 micrometers or smaller) has become a widespread problem across many nations, particularly in Asia. Major sources include agricultural and industrial practices, with contributions from transportation and residential activities. PM 2.5 smog not only reduces air quality but is also a major cause of respiratory illnesses, including decreased lung function, bronchitis, and lung cancer.

Several models have been developed to address PM 2.5, each tackling different aspects of the issue. Chemical Transport Models (CTMs) simulate the transport, chemical transformation, and deposition of PM 2.5 in the atmosphere [1, 2]. Statistical and machine learning models leverage historical data and observable trends for short-term predictions [3, 4, 5, 6, 7]. Gaussian dispersion models focus on predicting local dispersion patterns.

Many studies have modeled the spread of PM 2.5 or similar pollutants, often examining their dispersion and transport under varying atmospheric conditions and terrains [8, 9, 10]. This research builds on such studies by using a two-dimensional advection-diffusion model to simulate PM 2.5 dispersion, emphasizing multiple continuous sources with distinct locations and emission time windows.

To develop a faster surrogate model for predicting and assessing pollution severity based on known data, a long short-term memory (LSTM) recurrent neural network is trained on randomized data. This data includes source locations, emission time windows, and meteorological conditions such as sunlight and wind. The goal is to create a trained neural network capable of predicting maximum concentrations over a simulation period, yielding results comparable to solving

the advection-diffusion equation but in a much faster and efficient manner.

2. Method

This research utilizes two models. The first is the advection-diffusion equation, a classical approach for modeling the dispersion and transport of pollutant particles in the atmosphere. It is used to simulate PM 2.5 concentrations over a spatial domain with multiple sources and varying emission time windows.

The second model is a long short-term memory (LSTM) recurrent neural network, trained on the same randomized data used for the advection-diffusion model. The LSTM serves as a surrogate model, capable of predicting maximum concentrations at each time step without repeatedly solving the advection-diffusion equation, significantly improving computational efficiency.

2.1. Advection-Diffusion Equation

The advection-diffusion model is a linear hyperbolic partial differential equation used to describe the dispersion and transport of atmospheric pollutants. It accounts for diffusion caused by meteorological factors like sunlight and turbulence, as well as advection driven by wind. The two-dimensional equation models the temporal spread of pollutant concentration over an infinite spatial domain.

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} - u_x \frac{\partial C}{\partial x} - u_y \frac{\partial C}{\partial y} + S(x,y,t), \quad x,y \in \mathbb{R}, \quad t \geq 0 \tag{1}$$

Let $C(x,y,t)$ denote the concentration with respect to spatial coordinates and time, D_x and D_y be the diffusion coefficients, u_x and u_y be the wind velocity in the x and y directions respectively, with $S(x,y,t)$ as the source term.

2.1.1. Analytical Solution

Keeping in mind that the objective of this research is to consider multiple different sources simultaneously within the same domain, the analytical solution is preferred over its numerical counterpart. Performing iterative methods for solutions with respect to multiple source terms while complying to the appropriate numerical stability conditions condition quickly proves to be highly computationally expensive. The homogeneous advection-diffusion equation (case where $S(x,y,t) = 0$) may be solved using the Gaussian kernel:

$$G(x,y,t,x_0,y_0,t_0) = \frac{1}{4\pi\sqrt{D_x D_y}(t-t_0)} \left[\exp\left(-\frac{(x-\xi(t,t_0,x_0))^2}{4D_x(t-t_0)} - \frac{(y-\eta(t,t_0,y_0))^2}{4D_y(t-t_0)} \right) \right] \tag{2}$$

Suppose the angle of the direction θ is known, with respect to the positive x-axis, so that the wind velocity components may be written with respect to wind direction, such that $u_x = |\vec{u}|\cos(\theta)$ and $u_y = |\vec{u}|\sin(\theta)$, where the magnitude is defined as $|u| = \sqrt{u_x^2 + u_y^2}$. To condense the notation, we denote the following characteristic equations which account for the transport due to advection in the x and y directions respectively.

$$\xi(t,t_0,x_0) = x_0 + |\vec{u}|\cos(\theta)(t-t_0) \tag{3}$$

$$\eta(t,t_0,y_0) = y_0 + |\vec{u}|\sin(\theta)(t-t_0) \tag{4}$$

In this model, the sources are designed to emit continuously over a designated time span rather than at a single instance. Therefore, the concentration must account for the source term over the entire time span, which can be achieved through temporal convolution. The concentration can be expressed as:

$$C(x,y,t,x_0,y_0,t_0) = \int_{t_0}^{t} S(x,y,\tau,x_0,y_0,t_0) G(x,y,\tau,x_0,y_0,t_0)\, d\tau \tag{5}$$

When discussing the dispersion of air pollutants in the atmosphere, stability classes are often used to determine the degree of mixing caused by weather factors such as turbulence and sunlight. As a result, the diffusion factor in the solution can be rewritten using stability Briggs parameters:

$$\sigma_x = \sqrt{2D_x(t-t_0)}, \quad \sigma_y = \sqrt{2D_y(t-t_0)}.$$

Briggs parameters are selected based on meteorological conditions, such as ground-level wind speed and sunlight, which are classified using Pasquill stability classes ranging from A to F, where A represents highly unstable conditions and F represents highly

stable conditions. See **Appendix A** for details on Pasquill stability classes.

After determining the appropriate stability classes for the region of interest based on wind and sunlight conditions, it is assumed in this research that diffusion in the x direction is the same as in the y direction. Consequently, the Briggs parameters can be calculated using the following equations:

$$\begin{aligned}\sigma_x(t,t_0,x_0) &= \alpha\xi(t,t_0,x_0)(1+0.0004\xi(t,t_0,x_0))^{-\frac{1}{2}},\\ \sigma_y(t,t_0,x_0) &= \alpha\eta(t,t_0,y_0)(1+0.0004\eta(t,t_0,y_0))^{-\frac{1}{2}},\end{aligned} \tag{6}$$

where the value of α is determined according to the stability classes, as outlined in **Appendix B**.

2.1.2. Multiple Sources and Parallel Computing

By the superposition principle observed in linear partial differential equations such as the advection-diffusion equation, the concentration from each source contribution may be represented as a sum in order to yield the overall concentration in the spatial domain.

$$C(x,y,t) = \sum_{i=1}^{N} C_i(x,y,t,x_{0,i},y_{0,i},t_{0,i}), \tag{7}$$

Since the concentration from each source is independent, concentrations can be computed in parallel for efficiency. In this model, sources are treated as point sources emitting continuously from $(x_{0,i}, y_{0,i})$ within $t \in [t_i^{\text{act}}, t_i^{\text{deact}}]$, with the source term expressed as $S_i(x,y,t) = S_i(t)\delta(x-x_0)\delta(y-y_0)$. Sources follow switch conditions: $S_i = 0$ for $t \notin [t_i^{\text{act}}, t_i^{\text{deact}}]$ and $S_i = \text{const.}$ for $t \in [t_i^{\text{act}}, t_i^{\text{deact}}]$, implemented via Heaviside functions.

Heaviside functions determine activation and deactivation times. For $t \geq t_i^{\text{act}}$, $H(t-t_i^{\text{act}}) = 1$ activates the source; otherwise, its contribution is zero. Before t_i^{act}, $C_i = 0$, achieved by multiplying $H(t-t_i^{\text{act}})$ to all terms. For $t > t_i^{\text{deact}}$, a second integral accounts for contributions after deactivation, ensuring the full emission time window and elapsed time are included. Combining all the parts mentioned earlier, the concentration as a function of space and time in this model can be written as

$$\begin{aligned}C(x,y,t) = \sum_{i=1}^{N} H(t-t_i^{\text{act}})\Bigg[H(t_i^{\text{deact}}-t)\\ \int_{t_i^{\text{act}}}^{\min(t,t_i^{\text{deact}})} \frac{S_i(\tau-t_i^{\text{act}})\delta(x-x_{0,i})\delta(y-y_{0,i})}{2\pi\sigma_{x,i}(\xi(\tau,t_i^{\text{act}},x_{0,i}))\sigma_{y,i}(\eta(\tau,t_i^{\text{act}},y_{0,i}))}\\ \exp\left(-\frac{(x-\xi(\tau,t_i^{\text{act}},x_{0,i}))^2}{2\sigma_{x,i}^2(\xi(\tau,t_i^{\text{act}},x_{0,i}))}-\frac{(y-\eta(\tau,t_i^{\text{act}},y_{0,i}))^2}{2\sigma_{y,i}^2(\eta(\tau,t_i^{\text{act}},y_{0,i}))}\right)d\tau\\ +H(t-t_i^{\text{deact}})\int_{t_i^{\text{act}}}^{t_i^{\text{deact}}} \frac{S_i(t-\tau)\delta(x-x_{0,i})\delta(y-y_{0,i})}{2\pi\sigma_{x,i}(\xi(\tau,t,x_{0,i}))\sigma_{y,i}(\eta(\tau,t,y_{0,i}))}\\ \exp\left(-\frac{(x-\xi(\tau,t,x_{0,i}))^2}{2\sigma_{x,i}^2(\xi(\tau,t,x_{0,i}))}-\frac{(y-\eta(\tau,t,y_{0,i}))^2}{2\sigma_{y,i}^2(\eta(\tau,t,y_{0,i}))}\right)d\tau\Bigg]\end{aligned} \tag{8}$$

By inputting source functions S_i, along with their respective spatial coordinates $x_{0,i}$, $y_{0,i}$, activation time t^{act_i} and deactivation times t_i^{deact} for $i = 1, 2, ..., N$, where N is the total number of sources, the overall pollutant concentration $C(x,y,t)$ within the spatial and time domains may be computed. The deactivation time for each source is computed by $t_i^{\text{deact}} = t_i^{\text{act}} + t_i^{\text{emission}}$. Both t_i^{act} and t_i^{deact} must be contained within the simulation time interval, so that $t_i^{\text{act}}, t_i^{\text{deact}} \in [t_0, T]$ for $i = 1, 2, ..., N$. For purposes of this study, there are 500–3000 continuous point sources with emission rates S_i that are randomly generated to be between 1000-5000 $\mu g/cm^3/s$ in a $10,000m \times 10,000m$ spatial region for $T = 86,400s$ or one day, starting at $t_0 = 0$. The magnitude of wind speed $|\vec{u}|$ is randomly generated to be within the range 1–10 m/s and uniform throughout the entire spatial domain considered.

2.2. Optimization Via LSTM Neural Network Surrogate Model

The maximum concentration is often used as a key metric for assessing PM 2.5 or other pollutant concentrations, as it provides insight into the worst-case scenario. However, determining the maximum concentrations when considering multiple sources with different emission time windows requires computing the solution for all time steps within the simulation time span. Although using the analytical solution reduces computational resources, calculating concentrations at every time step remains time-intensive.

Optimizing emission schedules to minimize the maximum concentration across all time steps is challenging, as testing different schedules by repeatedly solving the advection-diffusion equation is impractical. Therefore, an alternative approach is necessary to achieve optimal results more efficiently. It is important to note that the objective is to minimize the maximum concentration over the entire spatial region and time span, without focusing on the specific location where the maximum concentration occurs.

2.2.1. LSTM Neural Network Design

The LSTM neural network designed used in this study is relatively simple. The input layer processes sequences with features including number of sources, emission rates, durations, wind speed, and concentrations. The first hidden layer is an LSTM with 32 units and 'sequence' output mode, capturing temporal dependencies. A dropout layer with a 0.1 rate follows to prevent overfitting by deactivating 10% of neurons during training. Next, a second LSTM layer with 16 units and 'last' output mode condenses sequence information into a single vector, summarizing learned patterns. This output feeds into a fully connected layer, mapping features to a scalar for regression. The final regression layer computes the mean squared error (MSE) loss for training.

2.2.2. Optimizing Maximum Concentration

Determining the optimal emission schedule to minimize the overall maximum concentration is challenging due to the infinite number of possible schedules. While emission schedules are allowed to vary, source locations, emission durations, and rates remain fixed. For simplicity, 10,000 randomly generated candidate emission schedules are generated in order to be tested. As each source's emission time serves as a separate parameter, the complexity of parameter estimation is significantly increased, thus rendering basic optimization techniques unsuitable.

Solving the advection-diffusion equation for each schedule is computationally intensive. Therefore, a surrogate model, based on an LSTM, is used to predict maximum concentrations for optimization. Predictions from the LSTM model for the 10,000 emission schedules are stored, and the schedule most likely to yield the lowest global maximum concentration is applied in the advection-diffusion model. The optimization efficiency is defined as the ratio of maximum concentrations before and after optimization:

$$\mathscr{E} = \frac{\max_{x,y\in\mathbb{R}^2,\, t\geq t_0} |C(x,y,\tau,t^{\text{act}}_{\text{optimized}},t^{\text{deact}}_{\text{optimized}})|}{\max_{x,y\in\mathbb{R}^2,\, t\geq t_0} |C(x,y,\tau,t^{\text{act}}_{\text{unoptimized}},t^{\text{deact}}_{\text{unoptimized}})|} \tag{9}$$

Ideally, the value of $\mathscr{E}$ should be minimized to demonstrate that the optimized emission schedule effectively maintains lower maximum concentrations.

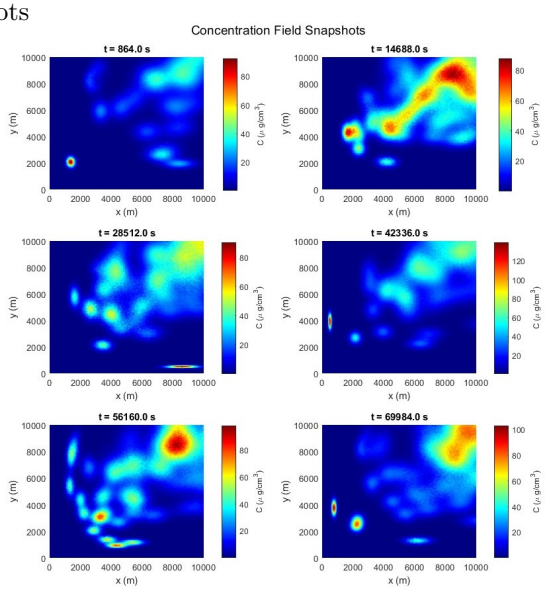
3. Results

The analytical solutions to the advection-diffusion equation are presented below to illustrate the concentration field when considering multiple sources within the same region with different emission times. Subsequently, the maximum concentrations from an initial simulation, prior to emission schedule optimization, are shown alongside the maximum concentrations obtained when the optimal emission schedule, derived from the LSTM surrogate model, is applied.

3.1. Advection-Diffusion Equation Solutions

Sample simulation snapshots of two-dimensional concentration plots generated by random source locations and emission schedules, as described earlier are displayed in **Figure 1**. As simulation time progresses, the concentration evolves due to diffusion and wind. With multiple source locations and varying emission times, the plumes inevitably overlap, resulting in higher maximum concentrations when two or more sources are sufficiently close to one another.

Figure 1: PM 2.5 concentration map in six time snapshots

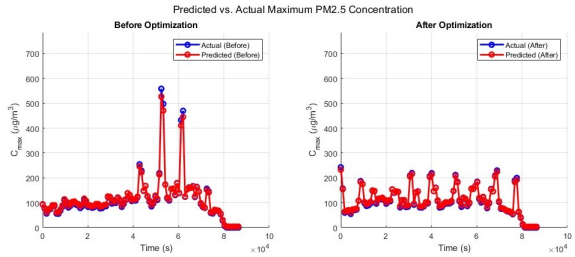


From the concentration maps in **Figure 1**, it is evident that predicting maximum concentration values at every time step solely from initial conditions and parameters, without computing the concentration field at each time step, is highly challenging. This difficulty has therefore motivated the development of a neural network surrogate model that is more computationally efficient to work with, especially during optimization routines.

3.2. Optimization Results

After verifying that the surrogate model's predicted maximum concentrations closely match the results from the advection-diffusion model, it is deemed suitable for use in optimization. Using 10,000 random candidate emission schedules, the most optimal emission schedule is applied to the advection-diffusion equation and compared to the maximum concentrations from the initial emission schedule.

Figure 2: Plot comparing maximum concentrations before and after emission schedule optimization



As shown in **Figure 2**, the maximum concentration plots obtained from both the neural network prediction and the analytical solution comparing the optimized emission schedule to the original unoptimized emission schedule yielded optimization efficiency ratio of $\mathscr{E} \approx 0.43$. This demonstrates that the method is able to significantly reduce the global maximum concentration compared to the original. Since the optimization method employed here relies on brute-force testing of emission schedules, integrating more sophisticated optimization techniques into the model could further reduce the cumulative maximum concentration and speed up the optimization process.

4. Conclusion

The advection-diffusion equation is an effective tool for modeling the dispersion and transport of air pollutants such as PM 2.5. With the inclusion of multiple sources, the analytical approach to solving the advection-diffusion equation becomes a compelling choice. However, the randomization and vast number of source locations, emission times, and atmospheric conditions (e.g., wind and sunlight) makes the optimization of emission schedules challenging and computationally expensive. The use of an LSTM neural network surrogate model to estimate maximum concentrations throughout the simulation significantly reduces computational cost and time. By identifying the approximated optimal emission schedule, only this schedule needs to be applied to the advection-diffusion model, streamlining the optimization process while maintaining accuracy.

Acknowledgments

This research is supported by Ph.D Degree Program in Mathematics, Faculty of Science, Chiang Mai University, under the CMU Presidential Scholarship and Chiang Mai University.

References

[1] C. Chemel, B. E. A. Fisher, X. Kong, X. V. Francis, R. S. Sokhi, N. Good, W. J. Collins, and G. A. Folberth. Application of chemical transport model cmaq to policy decisions regarding pm2.5 in the uk. *National Centre for Atmospheric Science, Centre for Atmospheric & Instrumentation Research*, 2013.

[2] J. Liu and J. Xing. Identifying contributors to pm2.5 simulation biases of chemical transport model using fully connected neural networks. *Journal of Advances in Modeling Earth Systems*, 15, 2023.

[3] S. A. Inam, A. A. Khan, T. Mazhar, N. Ahmed, and T. Shahzad. PR-FCNN: A data-driven hybrid approach for predicting PM2.5 concentration. *Discover Artificial Intelligence*, 3(1):1–12, 2023.

[4] T. Thongyen, N. Thongyen, and N. Thongyen. Statistical PM2.5 prediction in an urban area using vertical meteorological factors. *Atmosphere*, 14(3):589, 2023.

[5] Y. Zeng and Others. Extended random forest for multivariate air quality forecasting. *International Journal of Machine Learning and Cybernetics*, 14(5):1234–1245, 2023.

[6] F. Jiang, C. Zhang, S. Sun, and J. Sun. Forecasting hourly PM2.5 based on deep temporal convolutional neural network and decomposition method. *Applied Soft Computing*, 113:107988, 2021.

[7] H. Chen, K. T. Putra, C. Weng, and J. C. W. Lin. A novel predictor for exploring PM2.5 spatiotemporal propagation by using convolutional recursive neural networks. *Journal of Internet Technology*, 23(1):165–176, 2022.

[8] B. Choi and M. A. Hummel. Spatiotemporal air quality prediction using stochastic advection–diffusion model for multimodal data fusion. *Environmental Research Letters*, 20(1):014065, 2025.

[9] M. Khamlich, G. Stabile, G. Rozza, L. Környei, and Z. Horváth. A physics-based reduced order model for urban air pollution prediction, 2023.

[10] J. Yin. Analytical solutions of the pm2.5 diffusion model and its application. *Nature Environment and Pollution Technology*, 13(3):583–588, 2014.

Supplementary Resources

Appendix A: Pasquill Stability Classes

Wind Speed (m/s)	**Sunlight**			
	Strong	**Medium**	**Weak**	**Night**
< 1.9	A	A-B	B	D
$2 - 2.9$	A-B	B	C	D
$3 - 4.9$	B	B-C	C	D
$5 - 5.9$	C	C-D	D	D
> 6	C	D	D	D

Appendix B: Briggs Parameters for Dispersion Coefficient

Stability Class	α **value**	Description
A	0.32	Very unstable
B	0.32	Unstable
C	0.22	Slightly unstable
D	0.16	Normal
E	0.11	Slightly stable
F	0.11	Stable