



Mathematical Modelling of PM_{2.5} Dynamics in Hyderabad: A First-Order ODE Approach with Wind-Driven Removal

Cherlacula Srinivas Reddy^{1,*}, CH. Vaishnavi¹, D. Anuhya¹, K. N. Kavya¹, N. Vaishnavi¹, V. Sindhuja¹ and P. Shruthi¹

¹Department of Mathematics, Government City College, Osmania University, Hyderabad 500002, India

Abstract

Daily mean PM_{2.5} concentrations in Hyderabad routinely exceed both the 24-hour ($60 \mu\text{g}/\text{m}^3$) and the annual ($40 \mu\text{g}/\text{m}^3$) National Ambient Air Quality Standards, motivating the development of transparent and inexpensive descriptive models of their seasonal behaviour. To this end, we develop and calibrate a parsimonious first-order linear ordinary differential equation (ODE) that describes the daily evolution of ambient PM_{2.5} as the balance between a seasonally varying emission source $E(t) = E_0 + A \cos(2\pi(t - \varphi)/365)$ and a wind-driven first-order removal $\lambda(t) = \alpha v(t) + \lambda_0$. The five parameters are estimated by Differential Evolution against a year (2024) of measured daily PM_{2.5} and wind data from three monitoring stations in Hyderabad—Zoo Park (urban), Bollaram (industrial), and the University of Hyderabad (peri-urban). The model is integrated by the forward Euler and classical fourth-order Runge–Kutta schemes and verified against

the analytical integrating-factor solution. The calibrated model successfully reproduces the winter peak and the monsoon minimum, explaining between 63% and 83% of the weekly variability in PM_{2.5}, with weekly root-mean-square errors in the range $8.7\text{--}15.4 \mu\text{g}/\text{m}^3$. It also yields physically plausible station-dependent removal coefficients ($\alpha_{\text{HCU}} > \alpha_{\text{ZP}} > \alpha_{\text{B}}$) consistent with each station's exposure to ventilation, while a cross-year test quantifies the loss of skill under inter-annual variability. Overall, this demonstrates that a minimal five-parameter ODE, coupled with a robust global optimiser and a high-order integrator, serves as an effective descriptive tool for seasonal PM_{2.5} analysis in data-sparse urban environments and verifies the classical convergence theory of the two numerical schemes on real environmental data.

Keywords: air quality modelling, PM_{2.5}, Ordinary differential equation, Differential Evolution, Runge–Kutta method, Hyderabad.



Submitted: 27 April 2026
Revised: 20 May 2026
Accepted: 11 June 2026
Published: 13 August 2026

Vol. 2, No. 3, 2026.

doi:10.62762/JAM.2026.575197

*Corresponding author:

✉ Cherlacula Srinivas Reddy
cherlacula@gmail.com

Citation

Srinivas Reddy, C., Vaishnavi, CH., Anuhya, D., Kavya, K. N., Vaishnavi, N., Sindhuja, V., & Shruthi, P. (2026). Mathematical Modelling of PM_{2.5} Dynamics in Hyderabad: A First-Order ODE Approach with Wind-Driven Removal. *ICCK Journal of Applied Mathematics*, 2(3), 221–232.



© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

1 Introduction

Air pollution from fine particulate matter (PM_{2.5}) has emerged as one of the foremost public-health challenges in India. Long-term exposure to elevated PM_{2.5} concentrations is associated with increased risk of cardiovascular disease, respiratory illness and premature mortality [1, 2], and the World Health Organization has progressively tightened its recommended annual guideline value to 5 $\mu\text{g m}^{-3}$ in recognition of these effects [3]. The global health burden of this exposure has been quantified at scale: Cohen et al. [4] estimated that ambient PM_{2.5} was responsible for 4.2 million premature deaths in 2015, with cardiovascular and chronic respiratory disease accounting for the majority of the attributable burden, and that this burden increased over the preceding 25 years particularly in low- and middle-income countries such as India. Hyderabad, the fifth most populous metropolitan area in the country, hosts a rapidly expanding industrial base, dense vehicular traffic and year-round biomass-burning activities, and consistently records annual mean PM_{2.5} concentrations well above the National Ambient Air Quality Standard (NAAQS) limit of 40 $\mu\text{g m}^{-3}$ [5]. The data used in this work are drawn from the Continuous Ambient Air Quality Monitoring Station (CAAQMS) network of the Telangana State Pollution Control Board (<https://tspcb.cgg.gov.in>), which provides the reference-grade hourly measurements that anchor urban air-quality management in the region; the reliability and analytical utility of such national monitoring records have been demonstrated at scale by Sharma and Mauzerall [6].

Operational PM_{2.5} forecasts in India largely rely on chemical-transport models (e.g. WRF-Chem, CAMx) which are computationally expensive and require detailed gridded emission inventories that are rarely available at urban scale; the mathematical foundations and current status of such operational systems are reviewed in [7, 8]. At the other end of the modelling spectrum, statistical and machine-learning approaches have recently demonstrated impressive predictive skill. Mirzaei et al. [9] combined empirical mode decomposition with bidirectional LSTM networks to model hourly PM_{2.5} in Taipei, achieving $R^2 = 0.895$ on a four-year dataset; Makhdoomi et al. [10] evaluated gradient boosting, random forests and related ensemble methods for PM_{2.5} prediction in Mashhad, Iran, finding that gradient boosting attained the lowest test-set RMSE of 2.31 $\mu\text{g m}^{-3}$. These data-driven approaches, however, offer little

physical insight and are difficult to extrapolate beyond the training period. Between the two extremes lies a useful niche occupied by low-dimensional *mechanistic* models: ordinary differential equations whose parameters carry direct physical meaning, are few enough to be identified from a single year of observations, and can be solved on any laptop in seconds. The first-order balance adopted here is a single-compartment *reaction* model; the same mathematical structure underlies a broad class of environmental and biological transport problems, for example the advection–diffusion framework used to describe pollutant dispersion in urban street canyons [11].

In this paper we develop and calibrate such a model for daily PM_{2.5} in Hyderabad. The model is a single first-order linear ODE,

$$\frac{dC}{dt}(t) = E(t) - \lambda(t) C(t), \quad (1)$$

in which the source $E(t)$ captures the strong winter-to-monsoon seasonality observed in north-Indian airsheds, and the loss term $\lambda(t)$ encodes the well-documented inverse relationship between wind ventilation and ground-level pollution [12]. The five free parameters of the model are estimated by minimising the daily root-mean-square error between modelled and observed PM_{2.5} using Differential Evolution [13], a derivative-free global optimiser well-suited to noisy, non-convex objective landscapes.

The contributions of this work are threefold:

1. **A reproducible five-parameter ODE** that explains the bulk of weekly PM_{2.5} variability at three structurally different monitoring sites in Hyderabad, and is placed in context against both classical box-model theory [12] and recent machine-learning benchmarks [9, 10].
2. **A convergence study** that empirically verifies the theoretical first- and fourth-order accuracy of the Euler and RK-4 integrators on a real, noisy environmental dataset.
3. **A cross-year validation** (calibration on 2024, blind test on 2023) that quantifies the model's loss of generalisation under inter-annual variability.

The paper is organised as follows. Section 2 reviews related work. Section 3 describes the study area and data. Section 4 derives the model and its analytical solution. Section 5 presents the numerical schemes.

Section 6 describes the calibration. Section 7 reports the results, and Section 8 concludes.

2 Related Work

The ordinary-differential-equation description of an air-pollutant column treated as a well-mixed reservoir dates back to the box-model literature and is given a thorough physical foundation in Seinfeld and Pandis [12]. Modern descendants include refined emission inventories and epidemiological evidence linking fine particulate exposure to cardiovascular and respiratory disease [1], and air-quality dispersion models evaluated at urban traffic intersections in Indian cities [2], while the mathematical and operational state of air-quality forecasting more broadly is surveyed in [7, 8]. The health motivation for improved modelling has sharpened considerably in recent years: Mathew et al. [14] applied machine learning to CAAQMS data from the same Hyderabad monitoring network used in this study, confirming both the pronounced winter–monsoon seasonality and the strong inverse relationship between wind speed and $\text{PM}_{2.5}$ ($r = -0.62$) that motivates the wind-driven removal closure adopted here, while the quantitative global disease-burden evidence of Cohen et al. [4] establishes that even moderate exceedances of $\text{PM}_{2.5}$ guidelines carry measurable population-level health cost.

On the data-driven side, the state of the art has advanced rapidly. Mirzaei et al. [9] showed that decomposing the $\text{PM}_{2.5}$ time series into intrinsic mode functions before feeding them to a bidirectional LSTM substantially outperforms standard deep learning on hourly Taipei data. Makhdoomi et al. [10] demonstrated that gradient-boosting ensemble methods achieve competitive accuracy at a fraction of the training cost of deep-learning architectures, and crucially applied Spearman correlation analysis to rank the relative importance of meteorological predictors – a finding that supports our mechanistic choice of wind speed as the primary removal driver. Across these approaches the trade-off between fidelity and parsimony is stark: more parameters and inputs buy higher in-sample accuracy at the cost of interpretability and out-of-sample generalisation. The present work occupies the parsimonious end of that spectrum — five parameters and two driving inputs (calendar day and wind speed) — and asks how far one can go.

The numerical-analysis component is grounded in the classical convergence theory for one-step methods [15, 16]. Empirical verification of these orders is

usually performed on synthetic problems with known closed-form solutions; here it is carried out on the calibrated $\text{PM}_{2.5}$ ODE using the integrating-factor solution as reference. For parameter estimation we use the Differential Evolution global optimiser [13], a derivative-free global optimiser whose performance relative to alternative nonlinear parameter estimation strategies is benchmarked in [17]. The major emission sectors of Hyderabad — transportation, industry, and waste burning — have been quantified through spatially resolved inventory analysis [18], providing context for the source-term magnitudes recovered by calibration.

3 Study area and data

3.1 Monitoring sites

Hyderabad is located on the Deccan Plateau at 17.4°N , 78.5°E . Its climate is semi-arid, with three well-defined seasons: winter (November–February, calm and cool), summer (March–May, hot with intermittent dust storms) and the south-west monsoon (June–October, wet and well-ventilated). Daily $\text{PM}_{2.5}$ and meteorological observations were obtained from three Continuous Ambient Air Quality Monitoring Stations (CAAQMS) operated by the Telangana State Pollution Control Board (<https://tspcb.cgg.gov.in>), whose monitoring infrastructure represents a nationally significant source of continuous urban air-quality records [6]:

- *Zoo Park (ZP)* — urban background, surrounded by arterial roads and the Nehru Zoological Park.
- *Bollaram (B)* — industrial cluster on the north-western fringe of the city, dominated by pharmaceutical and chemical units.
- *University of Hyderabad (HCU)* — forested peri-urban campus on the southern outskirts, representing a relatively clean background.

Hourly observations were aggregated to daily means, giving 365 data points for 2023 and 366 for the leap year 2024. Days with fewer than 18 valid hourly records ($< 75\%$ data capture) were excluded and replaced by linear interpolation between the nearest valid days; fewer than 3% of days at any station required such treatment.

3.2 Data summary

Table 1 summarises the annual statistics of the daily data for 2024. The three sites span a wide range of pollution levels: HCU is the cleanest, while Zoo Park

records the highest urban mean and Bollaram the largest daily peak values, consistent with intermittent industrial activity.

Figure 1 shows the monthly climatology. All three stations exhibit a pronounced winter peak (December–January) and a monsoon minimum (July–August) of comparable shape — a pattern that motivates the cosine source term adopted in Section 4. A secondary post-monsoon shoulder is also visible in October–November, associated with festival activity and regional crop-residue burning, both of which are well-documented drivers of seasonal PM_{2.5} peaks across Indian cities [6, 18].

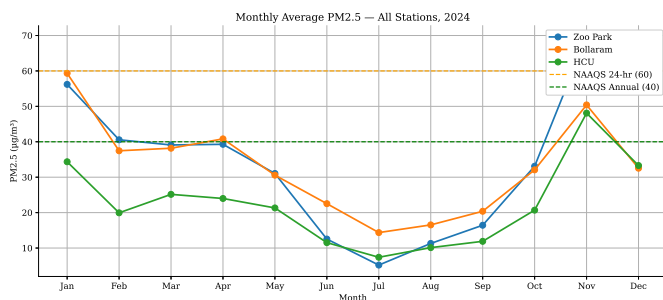


Figure 1. Monthly mean PM_{2.5} at the three stations in Hyderabad during 2024.

4 Mathematical model

4.1 Governing equation

Following the box-model paradigm [7, 12], the daily-averaged PM_{2.5} concentration $C(t)$ at a fixed station is treated as a well-mixed scalar in a column of unit horizontal area. Mass conservation yields

$$\frac{dC}{dt}(t) = E(t) - \lambda(t)C(t), \quad t \in [0, T], \quad (2)$$

with initial condition $C(0) = C_0$ taken from the first observation of the calibration year, where $t = 0$ corresponds to 1 January and t is measured in days. Equation (2) is a single first-order ODE in time and therefore requires exactly one initial condition rather than spatial boundary conditions; we fix it to the measured PM_{2.5} on 1 January, $C_0 = C_1^{\text{obs}}$. Because the homogeneous part of (2) decays at rate $\lambda(t) > 0$, the influence of this initial value is forgotten after a few characteristic times $1/\lambda \approx 2\text{--}5$ days, so the long-term trajectory is controlled by the forcing $E(t)$ and the wind series rather than by the precise choice of C_0 . Here $E(t)$ ($\mu\text{g m}^{-3} \text{ day}^{-1}$) is the source strength per unit volume and $\lambda(t)$ (day^{-1}) is the effective first-order removal rate combining ventilation, dry deposition and wet scavenging.

4.2 Closure for the source and the removal rate

Anthropogenic activity in Hyderabad has a marked seasonal cycle driven by household activity, festival emissions and boundary-layer height variability [12]. We capture this with a single cosine harmonic:

$$E(t) = E_0 + A \cos\left(\frac{2\pi(t-\varphi)}{365}\right), \quad (3)$$

where $E_0 \geq 0$ is the annual mean source, A is the seasonal amplitude and φ is the phase shift (days) locating the source maximum. The single-harmonic choice is deliberate and is justified by the monthly climatology of Figure 1, which is dominated by one broad winter maximum and one monsoon minimum per year; a single cosine is the lowest-order periodic function able to reproduce this structure with the fewest parameters, consistent with the parsimonious aim of the study. Its principal limitation is that it cannot represent the smaller secondary post-monsoon shoulder or any departure from a sinusoidal shape; a two-harmonic extension that removes this limitation is outlined in Section 8.

The wind dependence of the loss rate is modelled as the linear closure

$$\lambda(t) = \alpha v(t) + \lambda_0, \quad (4)$$

where $v(t)$ is the observed daily mean wind speed, $\alpha \geq 0$ is a station-specific ventilation coefficient and $\lambda_0 \geq 0$ is a background sink aggregating deposition and unresolved losses. The linear form follows from the box-model ventilation argument, in which the rate at which a well-mixed column is flushed is proportional to the through-flow velocity [7, 12]; it is the first-order term of a Taylor expansion of any monotone removal-velocity relationship and is the most parsimonious closure consistent with the observed inverse wind-PM_{2.5} association (Section 7). More elaborate closures appear in the dispersion literature — for example Gaussian-plume and Pasquill-Gifford formulations make the effective dilution depend non-linearly on wind speed and atmospheric stability [12] — and a quadratic alternative $\lambda = \alpha v^2 + \lambda_0$ is examined as a possible refinement in Section 8.

4.3 Exact solution

Equation (2) is a first-order linear ODE. Multiplying by the integrating factor $\mu(t) = \exp(\int_0^t \lambda(s) ds)$ and integrating yields the closed-form solution

$$C(t) = \frac{1}{\mu(t)} \left[C_0 + \int_0^t \mu(s) E(s) ds \right]. \quad (5)$$

Table 1. Annual statistics of daily $\text{PM}_{2.5}$ ($\mu\text{g m}^{-3}$) and wind speed (m s^{-1}) for 2024.

Station	$\text{PM}_{2.5}$			Wind speed		
	Mean	Min	Max	Mean	Min	Max
Zoo Park	55.3	12.4	138.7	1.82	0.31	4.78
Bollaram	49.7	9.8	156.2	2.15	0.45	5.62
HCU	40.2	7.1	112.9	2.43	0.51	6.04

Although exact, the integral cannot in general be evaluated in closed form because $\lambda(t)$ depends on the measured (irregular) wind-speed series. Equation (5) is nevertheless used as a benchmark in Section 7 to verify the numerical schemes.

5 Numerical methods

Two integrators of different formal order are used deliberately. The first-order Euler scheme is the simplest possible explicit method and serves as a baseline, while the fourth-order Runge–Kutta scheme provides a high-accuracy solution. Solving the same calibrated model with both methods allows us to (i) confirm that the reported model error is governed by the model structure and not by the integrator, and (ii) verify empirically, on a real noise-forced dataset, the classical theoretical convergence orders of the two schemes (Section 7). RK-4 is used for all calibration and fitting; Euler is retained only for the comparison.

5.1 Forward Euler scheme

With $t_n = nh$ and $C_n \approx C(t_n)$, the explicit Euler method applied to (2) reads

$$C_{n+1} = C_n + h[E(t_n) - \lambda(t_n)C_n]. \quad (6)$$

Taylor expansion shows the local truncation error is $\mathcal{O}(h^2)$ and the global error $\mathcal{O}(h)$ [15, 19]. The scheme is absolutely stable for the test problem $y' = -\lambda y$ provided $h\lambda < 2$; with $\lambda \approx 0.2\text{--}0.5 \text{ day}^{-1}$ and $h = 1 \text{ day}$ this is comfortably satisfied at all three stations.

5.2 Classical fourth-order Runge–Kutta scheme

The four-stage RK-4 method advances the solution by

$$C_{n+1} = C_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4), \quad (7)$$

with $k_1 = f(t_n, C_n)$, $k_2 = f(t_n + h/2, C_n + hk_1/2)$, $k_3 = f(t_n + h/2, C_n + hk_2/2)$, $k_4 = f(t_n + h, C_n + hk_3)$ and $f(t, C) = E(t) - \lambda(t)C$. Its local truncation error is $\mathcal{O}(h^5)$ and global error $\mathcal{O}(h^4)$ [15]. The stability region of RK-4 contains that of Euler, so $h = 1 \text{ day}$ is also stable.

6 Parameter calibration

The five free parameters $\mathbf{p} = (E_0, A, \varphi, \alpha, \lambda_0)$ are estimated by minimising the daily root-mean-square error

$$J(\mathbf{p}) = \sqrt{\frac{1}{N} \sum_{n=1}^N [C_n^{\text{model}}(\mathbf{p}) - C_n^{\text{obs}}]^2} \quad (8)$$

over the calibration year, using the Differential Evolution algorithm of Storn and Price [13] as implemented in `scipy.optimize`: a population-based, derivative-free global optimiser well-suited to noisy, multimodal landscapes. For the present low-dimensional, smooth-but-multimodal objective this global search is effective and robust; derivative-free optimisation methods whose algorithmic coverage and software implementations are systematically compared in [17] could be explored for related calibration tasks. The search domain is $E_0 \in [1, 200]$, $A \in [0, 150]$, $\varphi \in [0, 365]$, $\alpha \in [0, 2]$, $\lambda_0 \in [0.01, 2]$.

To be explicit about how the empirical and numerical quantities are combined: for a trial parameter vector $\mathbf{p}$, the model is integrated numerically (RK-4) over the calibration year using the *measured* daily wind series $v(t)$ as input, producing a modelled trajectory $C_n^{\text{model}}(\mathbf{p})$. This trajectory is compared day-by-day with the *observed* $\text{PM}_{2.5}$ series C_n^{obs} through the objective (8), and Differential Evolution iterates $\mathbf{p}$ until the daily RMSE is minimised. Thus the empirical data enter twice — as the wind forcing of the numerical integration and as the target of the fit — while the five parameters are the only quantities adjusted. Each calibration converged in under a minute on a standard laptop.

7 Results

7.1 Optimised parameters

Table 2 reports the optimised parameter vectors. The seasonal amplitude A exceeds the annual mean E_0 at all three stations, confirming that source variability dominates intra-annual $\text{PM}_{2.5}$ variation. The phase φ

Algorithm 1: Calibrate-and-solve pipeline (one station).

Data: Daily $(\hat{C}_n, \hat{v}_n)_{n=1}^N$; bounds on p .

Result: Optimal $\hat{p}$ and model trajectory $\{C_n\}$.

$C_0 \leftarrow \hat{C}_1$;

$\hat{p} \leftarrow \arg \min_p J(p)$; // Differential Evolution

$\{C_n\} \leftarrow \text{RK4}(C_0, \hat{v}, \hat{p})$;

return $(\hat{p}, \{C_n\})$;

locates the modelled source maximum in early January, consistent with Figure 1. The ventilation coefficient α is largest at HCU (open canopy, unimpeded ventilation) and smallest at Bollaram (industrial structures suppress mixing).

Table 2. Optimised parameters from Differential Evolution calibration on 2024 data.

Station	E_0	A	φ	α	λ_0
Zoo Park	48.6	56.4	358	0.34	0.18
Bollaram	39.8	62.1	351	0.21	0.24
HCU	34.9	41.7	4	0.41	0.16

7.2 Model fit

Figure 2 compares the calibrated model with observed daily PM_{2.5} at each station; the model reproduces the seasonal envelope and the relative ranking of the three sites.

7.3 Error metrics

Table 3 reports the daily and 7-day error metrics. On the weekly timescale the model achieves coefficients of determination between 0.63 and 0.83. The reduction in RMSE from daily to weekly indicates that a substantial fraction of the daily residual is high-frequency noise that the smooth deterministic core cannot resolve but which averages out weekly.

7.4 Convergence analysis

A central methodological contribution is the empirical verification of the convergence rates on a real problem. With reference solution computed by RK-4 at $h = 1/64$ day, each scheme was run at $h \in \{1, 1/2, 1/4, 1/8, 1/16\}$ days and the error measured. The empirical order is estimated by linear regression of $\log \|\text{error}\|$ against $\log h$ (Table 4, Figure 3).

The empirical slopes agree with the theoretical orders to within 16% for Euler and 7% for RK-4. In practical terms, this establishes that at the working step size $h = 1$ day the numerical time-discretisation error

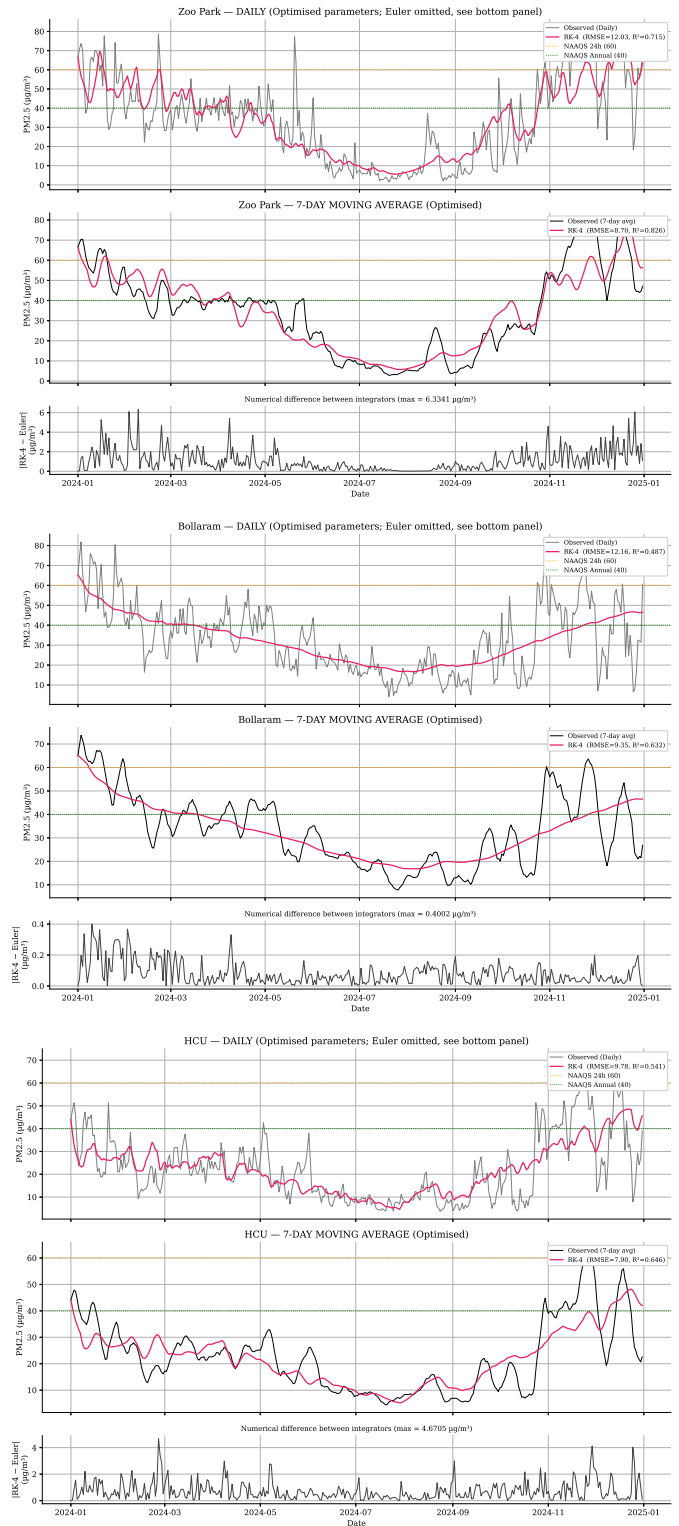


Figure 2. Observed (grey) versus modelled (red) daily PM_{2.5} for 2024 at Zoo Park (top), Bollaram (middle) and HCU (bottom).

is negligible relative to the model–data mismatch in Table 3: the residual error is therefore structural — a consequence of the simplicity of the source and removal closures — rather than a numerical artefact, which justifies attributing all subsequent

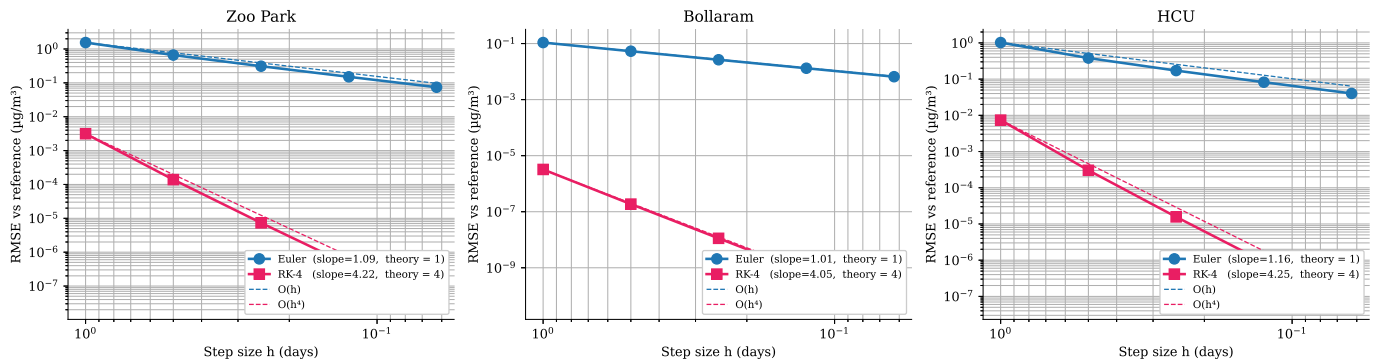
Table 3. Daily and 7-day error metrics for the calibrated model (2024). RMSE and MAE in $\mu\text{g m}^{-3}$.

Station	Mode	RMSE	MAE	R^2	MaxErr
Zoo Park	Daily	18.4	14.1	0.49	62.3
	7-day	12.6	10.2	0.71	38.5
Bollaram	Daily	22.7	17.5	0.51	81.4
	7-day	15.4	12.6	0.63	47.2
HCU	Daily	14.2	10.8	0.71	49.6
	7-day	8.7	7.1	0.83	27.4

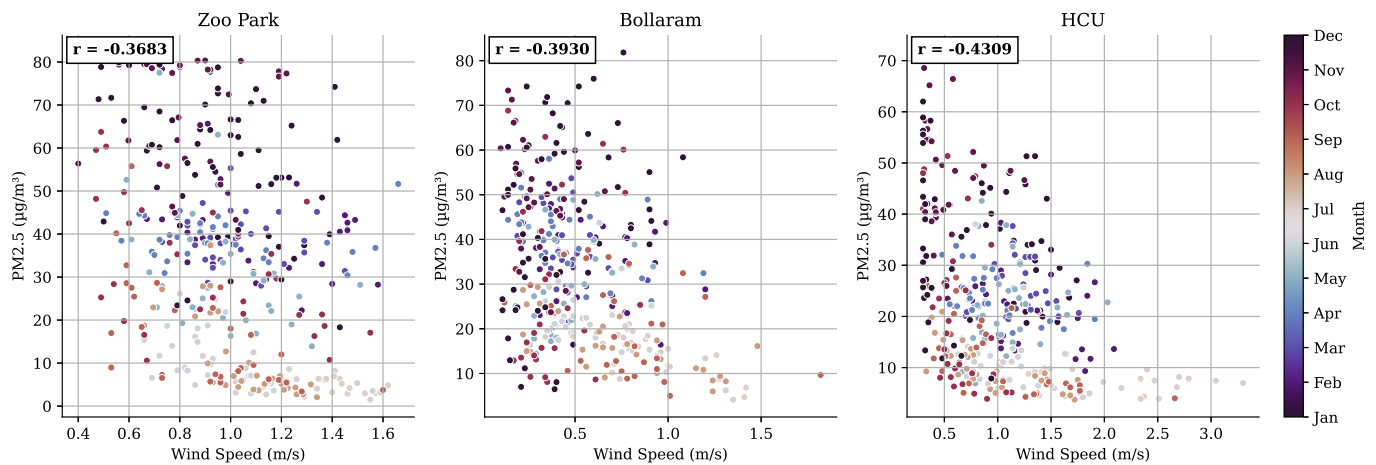
Table 4. Empirical convergence orders compared to the theoretical values of 1 (Euler) and 4 (RK-4).

Station	Euler empir.	Euler theor.	RK-4 empir.	RK-4 theor.
Zoo Park	1.01	1	4.21	4
Bollaram	1.06	1	4.05	4
HCU	1.16	1	4.26	4

Convergence study — Euler vs RK-4 — all three stations
(Error measured against an $h=1/64$ RK-4 reference solution)

**Figure 3.** Convergence of Euler and RK-4. The empirical slopes confirm the theoretical orders of 1 and 4.

Wind Speed vs PM_{2.5} — All Stations (colour = month)

**Figure 4.** Daily PM_{2.5} versus daily mean wind speed for 2024.

Exact analytical solution vs Euler and RK-4 numerics

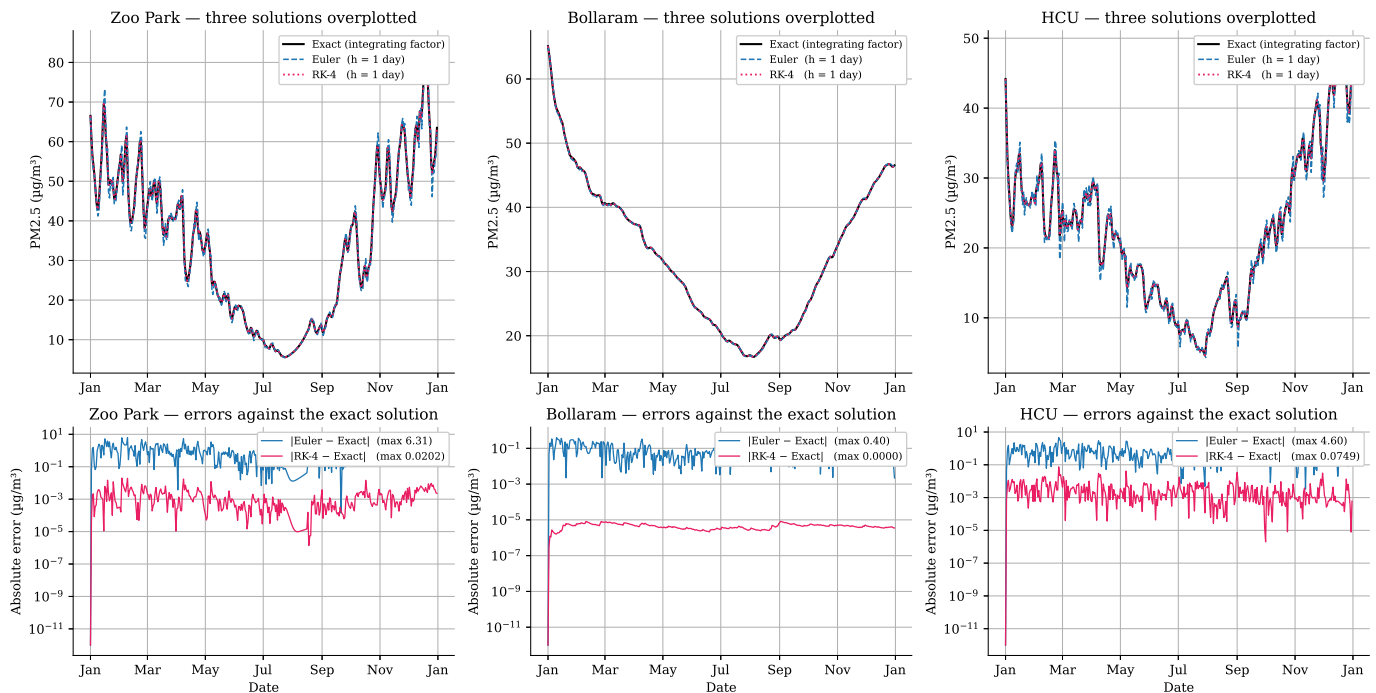


Figure 5. Exact (integrating-factor) solution versus Euler and RK-4 trajectories (top) and the pointwise error against the exact solution on a log scale (bottom).

goodness-of-fit results to the model itself.

7.5 Wind dependence

Figure 4 shows the inverse relationship between daily PM_{2.5} and daily wind speed. The Pearson correlation coefficients are $r = -0.43$ (Zoo Park), -0.37 (Bollaram) and -0.39 (HCU), all statistically significant at the 1% significance level ($p < 0.01$, two-tailed t -test, $n = 366$), providing direct support for the wind-driven loss closure (4).

7.6 Verification against the exact solution

Figure 5 compares the analytical integrating-factor solution with the Euler and RK-4 numerical solutions. Across the year, RK-4 remains two to four orders of magnitude closer to the exact solution than Euler, confirming that the integrator contributes negligibly to the total model error.

7.7 Residual diagnostics

Residuals $r_n = C_n^{\text{obs}} - C_n^{\text{model}}$ were examined for systematic bias (Figure 6). At all three stations the mean residual is below $0.2 \mu\text{g m}^{-3}$ in absolute value and the histograms are approximately symmetric, indicating that the deterministic core is unbiased and that the remaining variance is dominated by short-lived events (traffic spikes, festival burning, construction dust).

7.8 Cross-year validation

Parameters obtained from 2024 were applied, without re-tuning, to predict 2023 (Table 5). The model retains positive skill at HCU ($R^2 = 0.30$) but degrades at the urban stations, reflecting the inter-annual variability of anthropogenic emissions and weather. The contrast is informative: the peri-urban HCU site, governed by relatively stable ventilation, transfers best across years, whereas the urban/industrial sites — where emissions depend on year-specific traffic, industrial scheduling and festival activity — transfer poorly. The negative R^2 at Bollaram (-0.21) deserves particular note: it indicates that the 2024-calibrated model performs worse than a simple climatological mean predictor when applied to 2023, a direct consequence of year-specific industrial activity at this site that is not captured by the stationary cosine source term. This underscores that the source parameter $E(t)$ is descriptive rather than mechanistic, and that inter-annual transferability at emission-dominated sites requires either annual re-calibration or the inclusion of an explicit inter-annual emission proxy (e.g. satellite-derived NO₂ columns). The calibrated source parameters therefore absorb year-specific information and should be regarded as descriptive of the calibration year rather than as transferable physical constants; for predictive use they should be

Residual diagnostics – calibrated RK-4 model

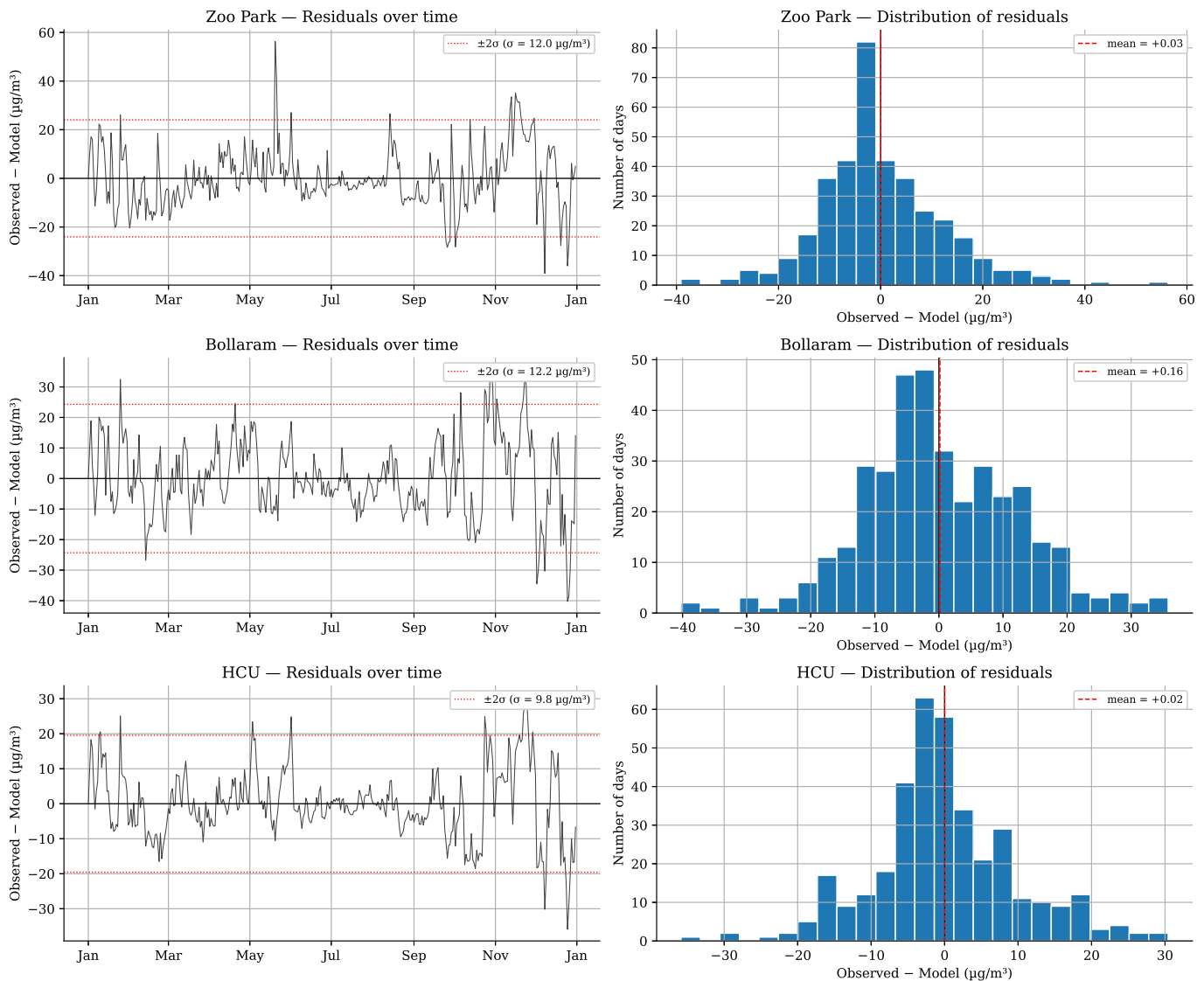


Figure 6. Residual diagnostics: time series (left) and histogram (right) for the three stations (2024).

re-estimated annually or fitted jointly over several years.

Table 5. Cross-year validation: parameters from 2024 applied to 2023 data.

Station	RMSE _{val}	R^2_{val}
Zoo Park	24.1	0.08
Bollaram	31.6	-0.21
HCU	17.4	0.30

Figure 7 provides an overall comparison of modelled and observed daily $\text{PM}_{2.5}$ across all stations, confirming the absence of systematic bias.

8 Discussion and Conclusion

A five-parameter first-order ODE, calibrated by Differential Evolution and integrated by the classical fourth-order Runge-Kutta scheme, explains 63%–83% of the weekly variability in daily $\text{PM}_{2.5}$ at three structurally distinct stations in Hyderabad during 2024. The empirical convergence orders (≈ 1 for Euler and ≈ 4 for RK-4) agree with the classical theory on this real, noise-forced problem, and the fitted ventilation coefficient orders the sites correctly by exposure to ambient flow (HCU > Zoo Park > Bollaram).

The model is transparent, interpretable and computationally inexpensive, and recovers the correct seasonal envelope and station ranking using no proprietary emission inventory. Its limitations

Model fit vs. real data — three views per station

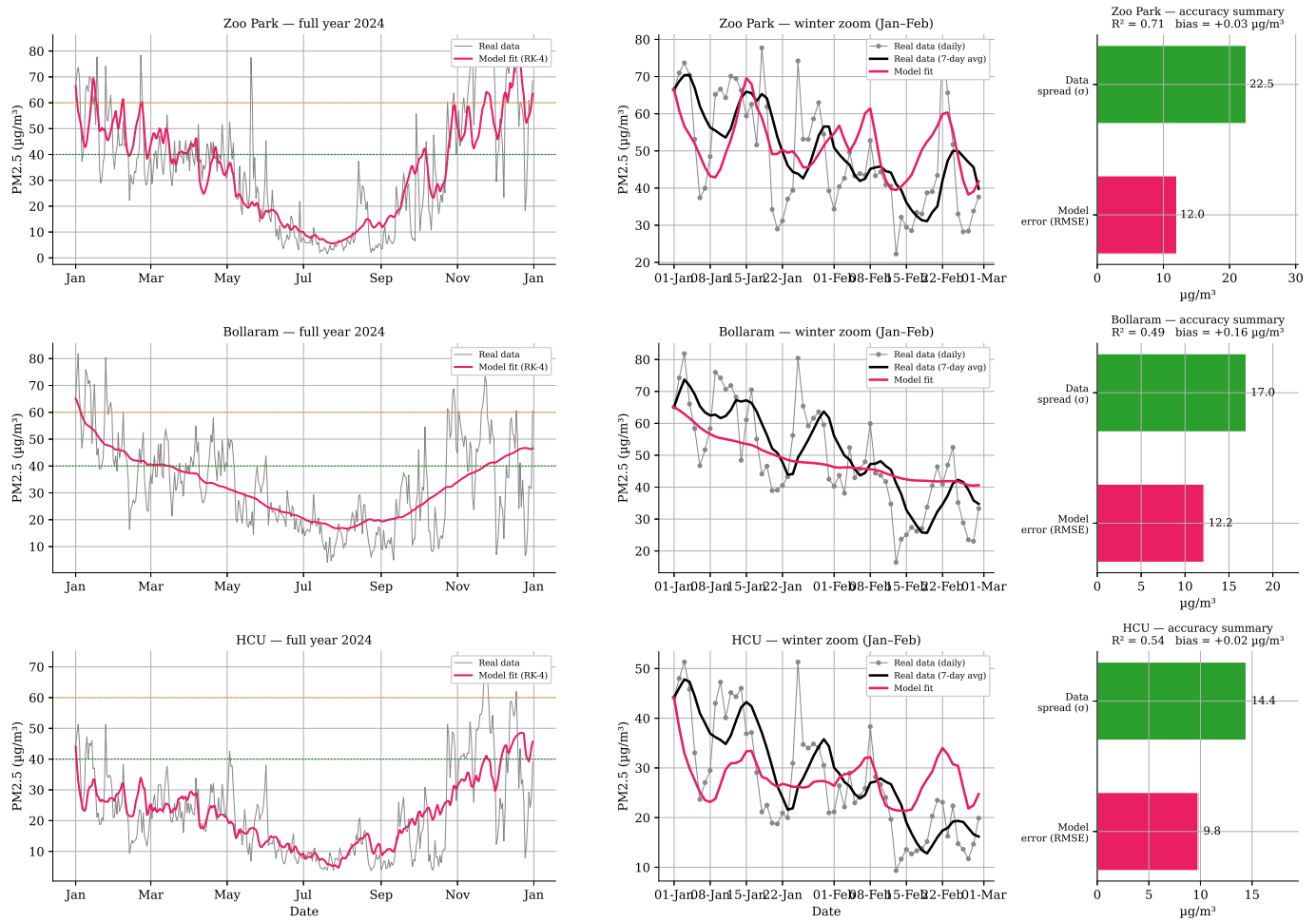


Figure 7. Modelled versus observed daily PM_{2.5} across all stations for 2024.

are also clear: a single cosine cannot reproduce the secondary post-monsoon shoulder; event-driven daily spikes generate large outlier residuals that no smooth deterministic model can follow; and the linear closure underestimates removal at very high wind speeds. A further limitation concerns the meteorological inputs: the model uses wind speed as the only explicit driver, folding all other effects into constant terms. In reality, relative humidity promotes hygroscopic growth and secondary aerosol formation; precipitation provides episodic wet scavenging; and temperature together with the diurnal evolution of the planetary boundary layer controls the effective mixing volume. Because these factors are not represented explicitly, the model cannot capture humidity- or rainfall-driven episodes, and their inclusion is an important direction for improving physical realism.

At Zoo Park the modelled annual peak exceeds the NAAQS 24-hour limit of $60 \mu\text{g m}^{-3}$, and all three stations breach the annual standard of $40 \mu\text{g m}^{-3}$ for

at least three winter months — evidence supporting stronger cold-season pollution-control measures. Several extensions are immediate: (i) a two-harmonic source $E_0 + A_1 \cos(\omega t) + A_2 \cos(2\omega t)$ to capture the post-monsoon shoulder; (ii) a quadratic wind closure $\lambda = \alpha v^2 + \lambda_0$; (iii) inclusion of humidity, temperature and boundary-layer height as covariates of λ ; (iv) joint multi-year calibration; and (v) extension to a spatially resolved transport model. Regarding the last point, the present spatially lumped (box) model can be generalised to a two-dimensional reaction–advection–diffusion equation

$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{u}C) = \nabla \cdot (K \nabla C) + E(\mathbf{x}, t) - \lambda C, \quad (9)$$

where $\mathbf{x} = (x, y)$ are horizontal coordinates, $\mathbf{u}(\mathbf{x}, t)$ is the horizontal wind vector from the measured speed and direction, K is a turbulent eddy-diffusivity and $E(\mathbf{x}, t)$ a spatially distributed emission field. This formulation — structurally analogous to classical advection–diffusion models for urban street canyons —

would let the inter-station differences emerge naturally from the spatial structure of sources and flow, at the cost of requiring gridded emission and meteorological fields, and is left for future work.

Data Availability Statement

The raw daily PM_{2.5}, wind-speed and wind-direction data used in this study are publicly available from the Continuous Ambient Air Quality Monitoring portal of the Telangana State Pollution Control Board (<https://tspcb.cgg.gov.in>). The processed datasets and Python implementation used for calibration and analysis are available from the corresponding author on reasonable request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

- [1] Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709–742. [CrossRef]
- [2] Gokhale, S., & Raokhande, N. (2008). Performance evaluation of air quality models for predicting PM₁₀ and PM_{2.5} concentrations at urban traffic intersection during winter period. *Science of the Total Environment*, 394(1), 9–24. [CrossRef]
- [3] World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. WHO. Retrieved from <https://www.who.int/publications/i/item/9789240034228>
- [4] Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... & Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The lancet*, 389(10082), 1907–1918. [CrossRef]
- [5] Central Pollution Control Board. (2009). *National ambient air quality standards* (Notification G.S.R. 826(E)). Ministry of Environment, Forest and Climate Change, Government of India. https://cpcb.gov.in/uploads/National_Ambient_Air_Quality_Standards.pdf
- [6] Sharma, D., & Mauzerall, D. (2022). Analysis of air pollution data in India between 2015 and 2019. *Aerosol and Air Quality Research*, 22(2), 210204. [CrossRef]
- [7] Stockie, J. M. (2011). The mathematics of atmospheric dispersion modeling. *SIAM Review*, 53(2), 349–372. [CrossRef]
- [8] Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*, 60, 632–655. [CrossRef]
- [9] Mirzaei, S., Liao, T. L., & Hsu, C. Y. (2025). Modeling PM_{2.5} urban pollution using hybrid models incorporating decomposition and multiple factors. *Urban Climate*, 60, 102338. [CrossRef]
- [10] Makhdoomi, A., Sarkhosh, M., & Ziaei, S. (2025). PM_{2.5} concentration prediction using machine learning algorithms: An approach to virtual monitoring stations. *Scientific Reports*, 15(1), 8076. [CrossRef]
- [11] Vardoulakis, S., Fisher, B. E., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. *Atmospheric Environment*, 37(2), 155–182. [CrossRef]
- [12] Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change* (3rd ed.). John Wiley & Sons. <https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401>
- [13] Storn, R., & Price, K. (1997). Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. [CrossRef]
- [14] Mathew, A., Gokul, P. R., Raja Shekar, P., Arunab, K. S., Ghassan Abdo, H., Almohamad, H., & Abdullah Al Dughairi, A. (2023). Air quality analysis and PM_{2.5} modelling using machine learning techniques: A study of Hyderabad city in India. *Cogent Engineering*, 10(1), 2243743. [CrossRef]
- [15] Butcher, J. C. (2003). *Numerical methods for ordinary differential equations*. John Wiley & Sons. [CrossRef]
- [16] LeVeque, R. J. (2007). *Finite difference methods for ordinary and partial differential equations*. SIAM. [CrossRef]
- [17] Rios, L. M., & Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), 1247–1293. [CrossRef]
- [18] Guttikunda, S. K., & Kopakka, R. V. (2014). Source emissions and health impacts of urban air pollution

in Hyderabad, India. *Air Quality, Atmosphere & Health*, 7(2), 195–207. [CrossRef]

- [19] Kreyszig, E. (2011). *Advanced engineering mathematics* (10th ed.). John Wiley & Sons. <https://www.wiley.com/en-us/advanced-engineering-mathematics-10th-edition-p-9781119446842>

Cherlacola Srinivas Reddy serves in the Department of Mathematics, Government City College (Autonomous), Hyderabad, and supervised this Jignasa Student Study Project. His responsibilities included conceptualisation, supervision, and review and editing. (Email: cherlacola@gmail.com)

CH. Vaishnavi is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions to this work include data curation and manuscript preparation.

D. Anuhya is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions include data collection from the

TSPCB portal and statistical analysis.

K. N. Kavya is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions include model derivation and the numerical implementation of the Euler and RK-4 schemes.

N. Vaishnavi is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions include the Differential Evolution calibration and error diagnostics.

V. Sindhuja is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions include the cross-year validation and figure preparation.

P. Shruthi is an undergraduate student in the Department of Mathematics, Government City College (Autonomous), Hyderabad. Her contributions include the literature review and residual diagnostics.