# Mechanistic Modeling of Fine Particulate Matter Dynamics at the Sihwa Industrial Complex Using Partially Observed Markov Process (POMP)


## Abstract

We develop a nonlinear POMP model for daily $PM_{2.5}$ (fine PM) concentrations at the Sihwa Industrial Complex industrial monitoring station in South Korea, spanning January 2021 to March 2025 (1,551 days). The latent state evolves through temperature-driven planetary boundary layer (PBL) mixing, humidity effects, wind-driven dispersion, and an emission rate that relies on seasonal variation. Parameters are estimated via iterated filtering (mif2), with a global search over 100 random starting points confirming that the local search MLE is reliable. The model achieves a log-likelihood of $-763.53$, comparable to a standalone ARMA(3,1) benchmark at $-761.26$, while providing physically interpretable parameters that the benchmark cannot. A regression + ARMA(3,1) benchmark at $-543.88$ achieves a higher log-likelihood through unconstrained flexibility, but CLL diagnostics show that a small number of anomalous days contribute to this gap, suggesting not all of it is a systematic misspecification. Key findings include that temperature-driven PBL mixing is negligible once seasonal variation is accounted for ($\alpha_T \approx 0.006$), resolving a strong temperature effect discovered in our previous midterm regression model that was due to seasonal confounding. Moreover, wind is the dominant removal mechanism ($\alpha_W \approx 0.33$) and humidity shows weak identifiability ($\alpha_H \approx 0.16$) with competing deposition and hygroscopic growth mechanisms that are not distinguishable from the data.


## 1 Introduction

Particulate matter (PM) is a critical indicator of air pollution that is made up of extremely small particles and liquid droplets containing acids, organic chemicals, metals, and soil or dust particles (Yoon et al. (2022)). Among air pollutants, PM is most strongly linked to human health risks (Fuzzi et al. (2015)). PM is typically categorized into $PM_{10}$ (coarse particles), which consist of particles with diameters between 2.5 and 10 micrometers and $PM_{2.5}$ (fine particles), which consist of particles 2.5 micrometers or less (Yoon et al. (2022)). While $PM_{10}$ is often filtered out by the respiratory tract, $PM_{2.5}$ can pass directly through it and penetrate deeper into the lungs, leading to cardiovascular or respiratory diseases that include stroke and lung cancer (Yoon et al. (2022); Zhou et al. (2018)) with prolonged exposure even at low concentrations leading to reduced life expectancy (Pope III & Dockery (2006); Fuzzi et al. (2015)). Given these public health implications, studying and understanding the behavior of $\mathrm{PM}_{2.5}$ and $\mathrm{PM}_{10}$ is a pivotal step toward effective air quality management and public health protection.

In our midterm project (Anonymous (2026)), we analyzed a dataset from the Asian Initiative for Clean Air Networks (AICAN), which maintains multiple air quality stations across South Korea (AICAN (2026)). Summary of the data and the preprocessing steps are detailed in the Supplementary (Section 5). In particular, we focused on the Sihwa Industrial Complex station, located near the coast in Gyeonggi Province. In the project, we modeled daily $PM_{2.5}$ at the industrial station using linear regression with ARMA errors. We analyzed the assocations between fine and coarse PM with meteorological variables such as temperature, relative humidity, air pressure, and wind speed. We selected ARMA(3,1) errors as our best fitting model and found significant association between temperature, humidity and wind speed relative to PM. However, our model suffered from near-unit-root errors and while we understood contemporaneous association, the mechanistic processes by which meteorology drove PM accumulation and dispersion were not fully captured. To understand the underlying mechanisms, we consider a partially observed Markov process (POMP) model(Ionides, 2026a; King et al., 2016) with a focus on fine PM at the industrial station. The inferential challenge of integrating over the unobserved latent trajectory is addressed via the particle filter, with parameters estimated by iterated filtering (Ionides, 2026b).

We therefore ask: can a mechanistic POMP model identify the relative contributions of meteorological variables to daily fine PM dynamics at the Sihwa Industrial Complex?

# 2 Model

## 2.1 POMP Framework

Let $X_t$ denote the true atmospheric $PM_{2.5}$ concentration ($\mu$g/m$^3$) at the industrial station on day $t$. This latent state evolves as a Markov process driven by daily-averaged meteorological covariates, temperature (°C), relative humidity (%), air pressure (hPa), and wind speed (m/s) respectively as per the standard atmospheric box model framework (Seinfeld & Pandis, 2006). The ground sensor yields a noisy daily-averaged observation $Y_t$.

## 2.2 Model Motivation

**Temperature** We examined the relationships between log-transformed daily $PM_{2.5}$ and each meteorological covariate to guide model structure. We modeled temperature with a softplus threshold rather than a linear term. According to (Stull (1988)), PBL mixing, which is an air mixing mechanism that creates circulation and removes near-surface $PM_{2.5}$, becomes prominent when temperature rises and solar heating is sufficient to destabilize the atmosphere. We see in our data that mean log $PM_{2.5}$ remains elevated and relatively flat for days below 10°C, then declines consistently as temperatures rise above 15°C, where active convective mixing is well established. This suggests a threshold effect around 5–10°C, where the PBL transitions from stable to convective. We can see therefore, that beyond a certain threshold, the atmosphere remains stably stratified and temperature has little additional removal effect. Above this threshold, we see a stronger effect of temperature in reducing fine PM due to the PBL effect.

Therefore, we considered that a linear temperature term would not fully be able to capture this behavior and considered the softplus function $\log(1 + e^{T-5})$, which provides a smooth, differentiable approximation of this threshold behavior (Dugas et al. (2001)), approaching zero for cold days and growing linearly for warm days. We fixed 5°C as our threshold, based on results from our data

discussed above and scientific evidence seems to suggest the same where at temperatures below this threshold, there is not enough heat flux to sustain a PBL strong enough to affect fine PM concentrations (Stull (1988)).

**Wind speed** We modeled wind speed as a linear term. We initially considered whether wind might increase fine PM at low speeds through dust resuspension, before reducing concentartions at higher speeds through dispersion, which would call for the same softplus threshold as in the temperature case. However, our data showed a monotonically decreasing relationship between wind speed and log-transformed fine pM across the full observed range as shown in Figure 1 when we computed mean log fine PM by wind speed bin. This is consistent with basic aerosol physics: mechanical resuspension requires particles larger than approximately 10 $\mu$m (Loosmore & Hunt, 2000; Seinfeld & Pandis, 2006). Fine PM, by definition is below 2.5 $\mu$m, and therefore may be unaffected by resuspension. Thus we consider wind to act purely as a removal mechanism through turbulent mixing and horizontal advection (Zhang et al., 2018). However, it is also the case that wind speed at Sihwa are generally low, with most observations clustered below 1.5 m/s, limiting the range of dispersion conditions captured in the data. We suspect therefore, that the wind may not play a strong role in removing fine PM at this station.

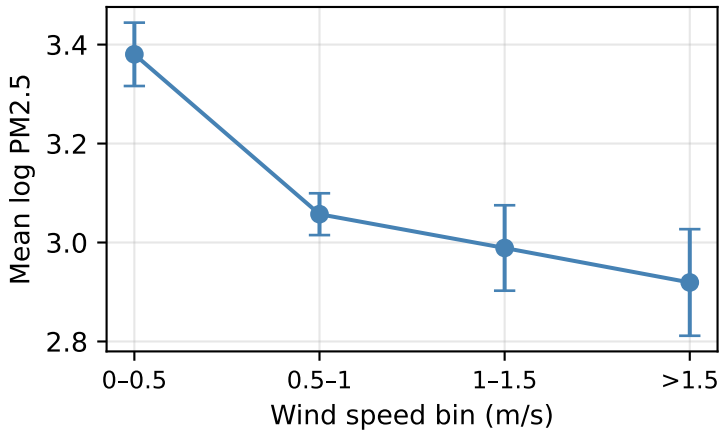


Figure 1: Mean log fine PM by wind speed bin, showing a monotonically decreasing relationship

**Relative humidity** We also consider a linear effect for relative humidity. Humidity affects fine PM in both positive and negative ways. On one hand, it causes wet deposition and on the other, increases particle mass through hygroscopic growth and causes settling (Tai et al. (2010)). Neither mechanism has a threshold analogous to the PBL effect, nor does it show any indication in the data, so a linear term is appropriate and the sign is left unconstrained to be determined by the data.

**Air pressure** Air pressure was excluded from the fine PM model. While we considered that intuitively, higher pressure would be physically associated with stagnant conditions in the atmosphere which may trap PM, we see that pressure is strongly collinear with seasonal tempearture patterns, making the two effects difficult to separate. We found that incorporating pressure may introduce identifiability concerns without a clear mechanistic gain beyond what the seasonal emission harmonics

already capture. Figure 2 illustrates this collinearity: monthly mean temperature and pressure track each other in near-perfect inverse phase, with pressure peaking in winter when temperature is lowest.

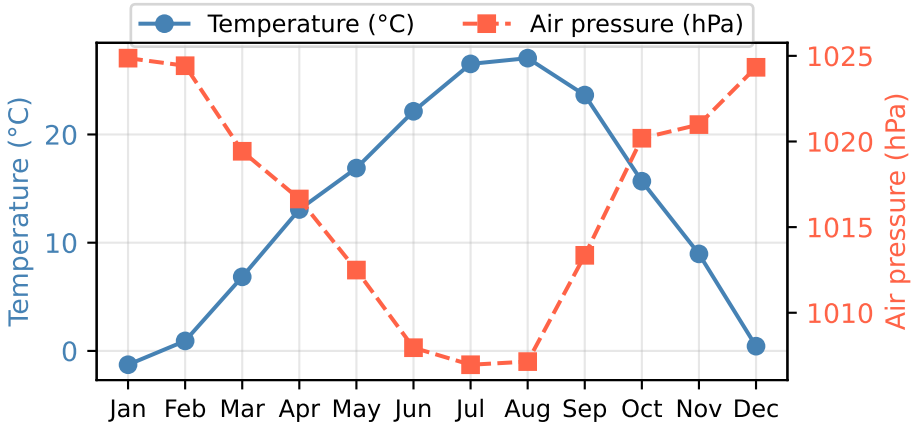


Figure 2: Monthly mean temperature and air pressure

**Seasonal pattern.** Monthly averaging over the data shows a clear peak in fine PM during late winter and early spring at the industrial station (Figure 3). Seasonality was also identified in our midterm project (Anonymous (2026)) but was not explicitly modeled there, a limitation we acknowledged at the time. Similarly, peer review of Project 17 (Winter 2025) noted that seasonality was identified but left unmodeled (STATS 531 course staff (2025b)). Motivated by the peer review suggestion, we explicitly incorporate seasonal variation through a pair of annual harmonic terms, a cosine and a sine with period 365 days (Shumway & Stoffer, 2017). This approach was based on what was covered in class expressing it as a linear combination of $\cos(2\pi t/365)$ and $\sin(2\pi t/365)$ (Ionides (2026a)). Using both terms together avoids fixing the peak timing in advance, using the data to determine where the peak occurs and how large it is. The amplitude of the seasonal cycle is $A = \sqrt{\delta_c^2 + \delta_s^2}$, and the peak day-of-year is $\phi = (365/2\pi)\arctan 2(\delta_s, \delta_c)$, which follows from the identity $a\cos\theta + b\sin\theta = \sqrt{a^2+b^2}\cos(\theta - \arctan 2(b,a))$ (Wikipedia contributors, 2026).

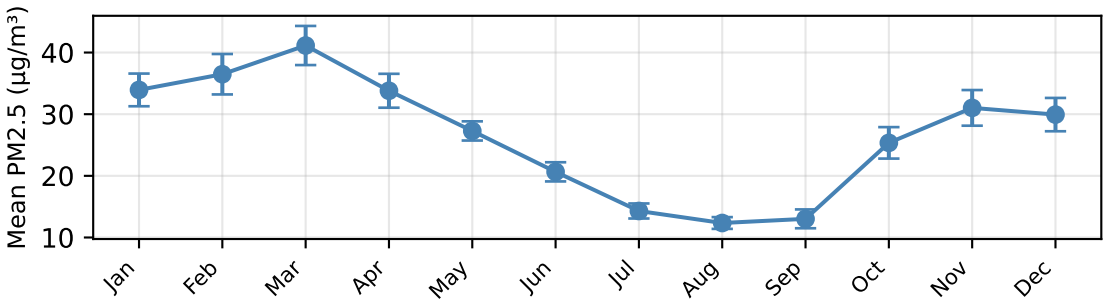


Figure 3: Monthly mean of fine PM (± 1.96 SE) at the station, showing a clear seasonal peak in late winter/early spring

## 2.3 Process Model

Considering each meteorological component as above, we designed our latent state as such

$$X_{t+1} = (X_t \cdot d_t + E_t)\, e^{\sigma \varepsilon_t}, \quad \varepsilon_t \overset{iid}{\sim} N(0,1) \tag{1}$$

The meteorological decay factor is a function of our meteorological variables to capture their role in atmospheric removal of fine PM:

$$d_t = \exp\Big(-\alpha_T \,\mathrm{softplus}(T_t - 5) - \alpha_H \tfrac{H_t}{100} - \alpha_W W_t\Big), \tag{2}$$

where $\mathrm{softplus}(x) = \log(1 + e^x)$. All meteorological covariates excluding air pressure were introduced in the decay term, capturing their role as removal mechanisms.

$$E_t = E \cdot \exp\big(\delta_c \cos(2\pi t/365) + \delta_s \sin(2\pi t/365)\big), \tag{3}$$

represents our seasonal emission rate and $E$ is the baseline daily emission rising from the industrial region. $\delta_c$, $\delta_s$ capture the seasonal pattern noted in Figure 3. The combined seasonal amplitude is $A = \sqrt{\delta_c^2 + \delta_s^2}$, with peak emission occurring at day-of-year $\phi = (365/2\pi)\arctan 2(\delta_s, \delta_c)$. We kept the emission term free of meteorological variables to avoid identifiability issues, thus designing our model in a way that industrial emission rates at Sihwa are driven by production schedules and seasonal demand rather than moment-to-moment weather.

The parameters were given the range where parameter $\alpha_T > 0$. Considering that temperature-driven PBL mixing can only remove PM2.5, $\alpha_T$ is constrained to be positive. The sign of $\alpha_H$ is left free. While high humidity can promote wet deposition of PM2.5, it also correlates with stagnant weather conditions that trap pollution (Tai et al. (2010)). The estimated sign reflects the dominant mechanism at this station. Parameter $\alpha_W > 0$ since wind can only remove fine PM through dispersion and cannot resuspend it, as shown in Figure 1.

**Multiplicative log-normal noise** ($e^{\sigma \varepsilon_t}$): PM2.5 concentrations are empirically log-normally distributed (Ott (1990)), motivating multiplicative noise that scales with concentration level, guarantees positivity, and reflects greater absolute variability during high-pollution episodes.

## 2.4 Observation Model

The daily sensor reading is log-normally distributed around the latent concentration, using the same reasoning as the process noise (Ott (1990)):

$$Y_t = X_t \cdot e^{\tau \eta_t}, \quad \eta_t \overset{iid}{\sim} N(0,1) \tag{4}$$

so that $\log Y_t \sim N(\log X_t,\ \tau^2)$. Working on the log scale is motivated by the same reasoning used in our midterm project: PM2.5 concentrations are right-skewed and the log transform stabilizes variance and produces an approximately normal distribution (Anonymous (2026)). The parameter $\tau$ captures

the discrepancy between the ground-level point measurement and the true fine PM concentration since the sensor is fixed at a single location and cannot fully observe the concentration.

## 2.5 Parameters

The mif2 algorithm perturbs parameters via additive Gaussian random walks on the estimation scale. Some of the parameters by design, we keep as strictly positive ($\alpha_T$, $\alpha_W$, $E$, $\sigma$, $\tau$) so we consider log-transformation to keep them above zero. Parameters that take either sign ($\alpha_H$, $\delta_c$, $\delta_s$), we leave as is on the natural scale. For reference, we summarize the parameters and their transformations in Table 1.

Table 1: Model parameters

| Parameter | Description | Transformation |
|---|---|---|
| $\alpha_T$ | Temperature PBL removal rate | log |
| $\alpha_H$ | Net humidity effect on removal | unconstrained |
| $\alpha_W$ | Wind dispersion removal rate | log |
| $E$ | Baseline emission rate ($\mu$g/m$^3$/day) | log |
| $\delta_c$ | Seasonal cosine amplitude | unconstrained |
| $\delta_s$ | Seasonal sine amplitude | unconstrained |
| $\sigma$ | Process noise log-normal sd | log |
| $\tau$ | Observation noise log-normal sd | log |

## 2.6 Inference

Parameters are estimated by iterated filtering (mif2) (Ionides, 2026b) as implemented in the `pypomp` package (pypomp developers, 2024). We ran 40 independent starts initialized near physically motivated default values with 10% multiplicative noise, each with 500 mif2 iterations, 8,000 particles, and a geometric cooling schedule with rate 0.2. Log-likelihood is evaluated at each endpoint using the particle filter with 8,000 particles averaged over 20 replicates (Ionides, 2026c). We consider our final result from the midterm project as our benchmark model (Anonymous (2026)), a linear regression on meteorological covariates with ARMA(3,1) errors since we seek to improve upon it.

We consider the initial values without too much scientific assumption, slightly basing it off the data. We start $\alpha_T$, $\alpha_H$, $\alpha_W$ all at 0.5. The baseline emission $E$ is initialized at 5 µg/m³, near the approximate mean based on our data. We initialize seasonal amplitudes at zero, making no prior assumptions. Noise parameters $\sigma$ and $\tau$ are initialized at 0.5. Finally, we fix the initial state $X_0$ to the first-week observed mean (~10 µg/m³) rather than estimated, because estimating an initial condition from a single trajectory is unreliable — fixing it to a data-informed value avoids this instability without meaningfully affecting inference on the remaining parameters.

# 3 Results

## 3.1 Convergence Diagnostics

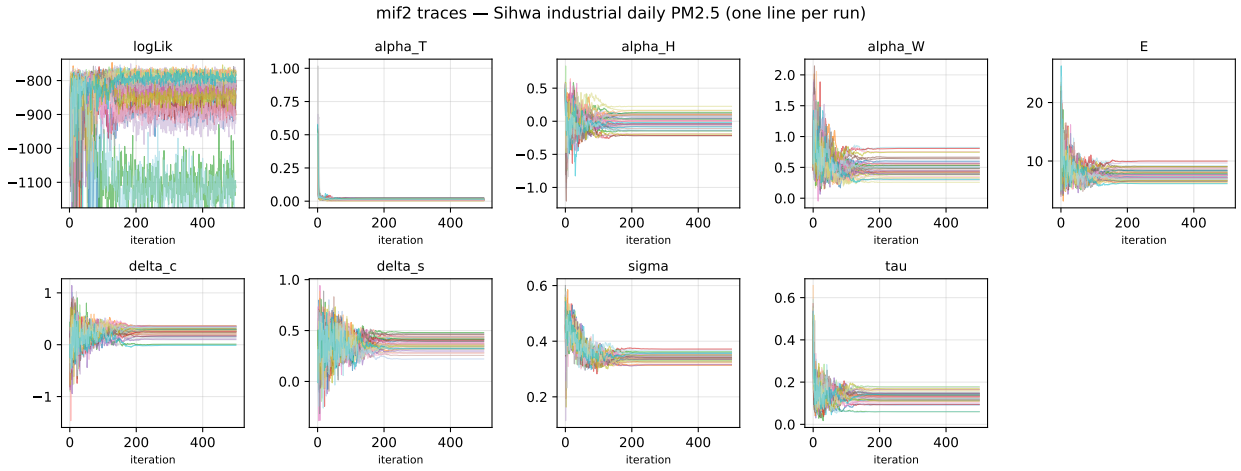


Figure 4: mif2 convergence traces for log-likelihood and all parameters across 40 local search runs. Each colored line is one independent run.

Figure 4 shows per-iteration mif2 traces for log-likelihood and all parameters across 40 local search runs. As a local search, each was initialized near the same initial values with slight perturbation of 10% multiplicative noise. As shown in the plot, the log-likelihood climbs in the first 100 iterations or so before stabilizing with most of them converging around $-763$ and $-800$. However, a minority fail to do so, likely due to unfavorable random initialization. Parameters $\alpha_T$, $\delta_c$, $\delta_s$, $\sigma$, $\tau$, and $E$ all seem to converge well into tight bands, where the progressive flattening are due to the cooling rate of 0.2, perturbation decaying to zero.

We see that $\alpha_H$ switches signs, some runs settle at positive values and others at negative. This may indicate that the net humidity effect is not well identified by the data. $\alpha_W$ shows a wider spread (~0.3–1.0) but with a clear central tendency around 0.3–0.5. $\alpha_T$ on the other hand, converges to near zero, suggesting a negligible effect of temperature-driven PBL mixing on fine PM at this station. The seasonal amplitudes $\delta_c$ and $\delta_s$ converge to stable values, consistent with a clear seasonal cycle peaking in early spring rather than mid-winter.

## 3.2 Global Search

To verify that the local search did not miss a better region of parameter space, we ran a global search over 100 random starting points drawn from a $\pm 0.5$ range around the initial parameter values. Figure 5 shows the endpoint log-likelihoods and parameter values across all runs.

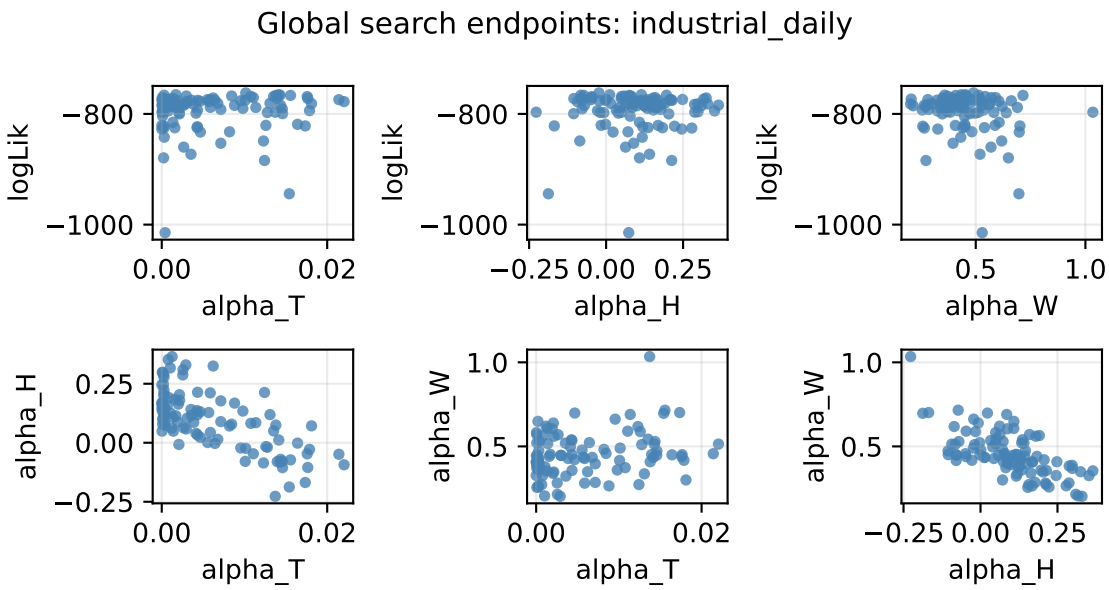


Figure 5: Global search endpoints: logLik vs each meteorological parameter (top row) and pairwise parameter scatter (bottom row).

The global search best log-likelihood is −762.09, within 1.4 units of the local MLE at −763.53, a difference well within Monte Carlo noise. This confirms that the local search did not miss a better region of parameter space and that the MLE is reliable.

### 3.3 Parameter Estimates

Table 2: Best-run parameter estimates (logLik = -763.53).

| Parameter | Estimate | Parameter | Estimate |
|---|---|---|---|
| $\alpha_T$ | 0.0056 | $\delta_c$ | 0.2598 |
| $\alpha_H$ | 0.1615 | $\delta_s$ | 0.4192 |
| $\alpha_W$ | 0.3274 | $\sigma$ | 0.3372 |
| $E$ | 6.3219 | $\tau$ | 0.1634 |

The best-run parameter estimates are shown above.

This contrasts with our earlier midterm regression + ARMA model, where temperature appeared as a significant predictor of fine PM (Anonymous (2026)). The POMP framework reveals that this was largely due to seasonal confounding. In the previous project, we had acknowledged the model's limitation that it had failed to take into account the seasonal variation. We now see here that the POMP model separates these effects by explicitly modeling the seasonal emission cycle through $E_t$, leaving $\alpha_T$ to capture only the PBL mixing effect, which is close to negligible.

We find that $\alpha_H$ switches between positive and negative, showing bimodality across runs, suggesting weak identifiability (Ionides (2026c)). Since there is both evidence that humidity could positively and negatively affect PM (Tai et al., 2010), we had wanted our model to explain the dominant effect, but we see that the data alone cannot distinguish the two competing effects. In contrast, $\alpha_W$ has a clear positive estimate: each 1 m/s increase in wind multiplies the removal rate by $e^{0.327} \approx 1.39$, confirming that wind disperses fine PM at this station rather than causing resuspension. The seasonal cycle has a clear amplitude and peak timing, consistent with elevated emissions during late winter and early March.

### 3.4 Model Fit and Limitations

We also fit a standalone ARMA(3,1) model to the fine PM series without any meteorological variables as a pure time series benchmark, achieving a log-likelihood of $-761.26$. Both the models evaluate on the log-transformed fine PM, so the log-likelihoods are directly comparable. The best POMP log-likelihood is $-763.53$, within 2.3 units of the standalone ARMA(3,1) benchmark, which shows comparable performance along with physical interpretability. However, the regression + ARMA(3,1) benchmark from our midterm project (Anonymous (2026)) at $-543.88$ achieves a substantially higher log-likelihood.

Furthermore, probe diagnostics comparing 200 model simulations against observed statistics show that the model systematically underestimates the mean, variability, and autocorrelation of fine PM. Full simulation diagnostics are provided in the Supplementary (Section 5).

This seems like a huge price to pay for physical interpretability and it may bring into question the reliability of the parameter estimates as well. A similar issue arose in Project 10 (Winter 2025), where the POMP model also failed to achieve competitive log-likelihood relative to the benchmark; peer review of that project highlighted the importance of CLL diagnostics for understanding where and why the mechanistic model underperforms (STATS 531 course staff (2025a)). We adopt this approach here to better understand the nature of the gap.

### 3.5 Conditional Log-Likelihood Diagnostics

To better understand where the POMP model underperforms relative to the benchmark model, we compute the per-day conditional log-likelihood (CLL) anomaly (Ionides (2026a)), the difference between the POMP one-step-ahead log density and the benchmark's at each observation. Figure 6 shows the anomaly and particle filter CLL variance over time.

Most days show anomalies near zero, indicating that POMP and the benchmark perform similarly under typical conditions. A few days of note however, show large negative anomalies, which are likewise reflected in the CLL variance, indicating that the particle filter struggles to compute a stable likelihood estimate on those days. These days correspond to sudden spikes in the fine PM, which cannot be explained by meteorological variables alone and are outliers that may have been caused by measurement error, industrial accidents, or long-range transport events. However, other than the small number of concerning days, most of the days seem to show near zero variance, which could suggest that the model is well-calibrated for the majority of the data and that the log-likelihood gap is caused by a small number of outlier events. The benchmark model may fare better in such cases since it can flexibly match the pattern in recent observations through its ARMA structure.

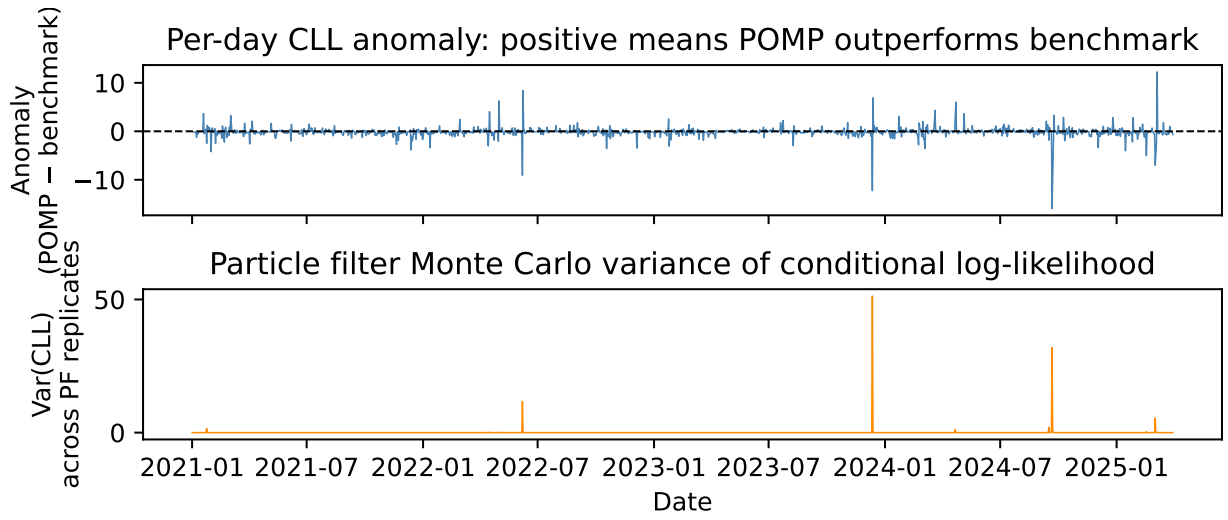


Figure 6: Per-day CLL anomaly (POMP minus benchmark, top) and particle filter CLL variance (bottom)

This allows it to handle outliers much better than the mechanistic POMP model, which cannot anticipate sudden spikes in fine PM that is not caused by its underlying mechanical process.

This explains a portion of the gap in the log-likelihood, which may suggest a need for model extension. We could consider adding a regional background state to capture long-range transport events. We could also consider measurement error, and look back to the raw data to see if there are any unnatural spikes that seem out of place and preprocess the data more thoroughly.

## 4 Conclusions

We fitted a POMP model to daily PM2.5 at the Sihwa Industrial Complex and estimated its parameters via iterated filtering, achieving a log-likelihood comparable to a standalone ARMA benchmark.

The key contribution over our midterm regression model is interpretability. The midterm model found a significant association between temperature and fine PM, which was in fact a seasonal variation, the POMP model was able to separate the effect, leaving $\alpha_T$ to capture only PBL mixing, which turned out to be negligible. Wind was the primary removal mechanism now expressed as a quantitative daily decay rate rather than a regression coefficient. These were questions that could not be answered by the midterm model.

**Limitations.** The model unfortunately underperforms relative to the regression + ARMA benchmark. It underestimates the mean, variability, and autocorrelation of fine PM as shown in the simulation diagnostics. The humidity effect $\alpha_H$ is weakly identified, with the data unable to distinguish wet deposition from hygroscopic growth. We also acknowledge that the cooling rate of $a = 0.2$ caused convergence around iteration 150 out of 500, leaving the remaining compute budget unused, however, we are fairly confident about the convergence of the parameters.

# 5 Supplementary

## 5.1 Data

Table 3: Metadata for the Sihwa Industrial Complex monitoring station.

| | Station | Latitude | Longitude | Height (m) | Installed |
|---|---|---|---|---|---|
| 0 | Industrial | 37.334244 | 126.726128 | 3.0 | 2019-12-16 |

Table 4: Summary statistics of daily PM2.5 (µg/m³) at the Sihwa industrial station, January 1, 2021 – March 31, 2025.

| | Industrial |
|---|---|
| N | 1551.00 |
| Mean | 27.17 |
| Std | 16.98 |
| Min | 1.02 |
| Q1 | 14.22 |
| Median | 23.41 |
| Q3 | 36.44 |
| Max | 106.96 |

Daily PM2.5 and meteorological measurements were obtained from the AICAN monitoring network for the Sihwa industrial station for the period January 1, 2021 through March 31, 2025, yielding 1,551 daily observations. Raw 10-minute data were first aggregated to hourly means, with short gaps of at most 3 consecutive hours filled by linear interpolation. Days with fewer than 18 valid hourly observations (75% coverage) were excluded before daily averaging, following US EPA data completeness criteria for ambient PM2.5 monitoring (U.S. Environmental Protection Agency, 2024). The resulting daily series were further interpolated across gaps of up to 7 days. Covariates used in the POMP model are daily-averaged temperature (°C), relative humidity (%), wind speed (m/s), and air pressure (hPa).

## 5.2 Simulation Diagnostics

Figure 7 shows probe diagnostics comparing the observed PM2.5 data against 200 model simulations at the best-fit parameters. Results consistently fall at the upper tail of the simulation distributions, indicating that the model systematically underestimates the mean concentration level, variability, and autocorrelation of fine PM.

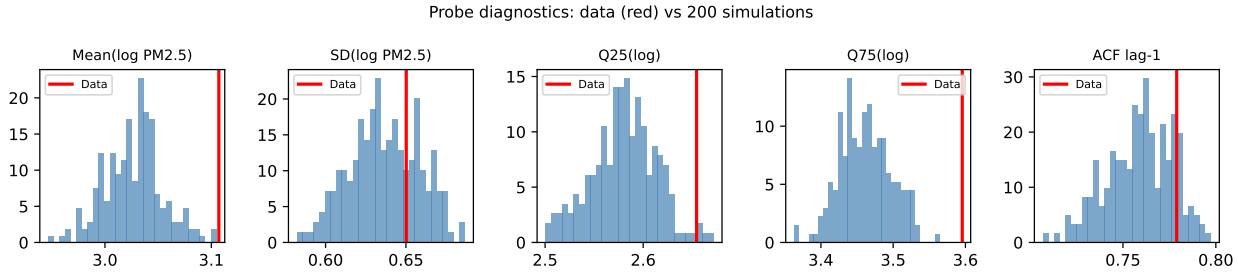


Figure 7: Probe diagnostics: observed statistics (red line) vs. distribution across 200 model simulations at the best-fit parameters. Observed values fall at the upper tail in all probes, indicating the model underestimates mean, variability, and autocorrelation.

## Acknowledgments

Data from the AICAN air quality monitoring network. Analysis conducted using `pypomp`, `JAX`, and `pandas`. Computations run on the University of Michigan Great Lakes cluster. AI (Claude) was used for debugging and code development in designing and formatting plots. Claude also suggested modeling the seasonal emission cycle using a pair of harmonic terms — $\cos(2\pi t/365)$ and $\sin(2\pi t/365)$ — which allows the peak timing and amplitude to be estimated from the data rather than fixed in advance. All AI-generated code and suggestions were reviewed and validated by the authors before inclusion, with final responsibility for the analysis resting with the team.

## Bibliography

AICAN. (2026). *Particulate matter data*. https://www.data.go.kr/data/15080316/fileData.do

Anonymous. (2026). *STATS 531 midterm project: PM2.5 analysis at the Sihwa industrial complex*. https://github.com/ionides/531w26/tree/main/midterm_project/project07

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C., & Garcia, R. (2001). Incorporating second-order functional knowledge for better option pricing. *Advances in Neural Information Processing Systems*, *14*.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y., Schaap, M., Slowik, J. G., Spracklen, D. V., Vignati, E., Wild, M., … Gilardoni, S. (2015). Particulate matter, air quality and climate: Lessons learned and future needs. *Atmospheric Chemistry and Physics*, *15*, 8217–8299. https://doi.org/10.5194/acp-15-8217-2015

Ionides, E. (2026a). *Notes for STATS 531, Modeling and Analysis of Time Series Data*. https://ionides.github.io/531w26/

Ionides, E. (2026b). *STATS 531 lecture notes: Iterated filtering*. https://ionides.github.io/531w26/

Ionides, E. (2026c). *STATS 531 midterm exam 2*. https://ionides.github.io/531w26/mt2/mt2-all-sol.pdf

King, A. A., Nguyen, D., & Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, *69*(12), 1–43. https://doi.org/10.18637/jss.v069.i12

Loosmore, G. A., & Hunt, J. R. (2000). Dust resuspension without saltation. *Journal of Geophysical Research: Atmospheres*, *105*(D16), 20663–20671. https://doi.org/10.1029/2000JD900271

Ott, W. R. (1990). A physical explanation of the lognormality of pollutant concentrations. *Journal of the Air & Waste Management Association*, *40*(10), 1378–1383. https://doi.org/10.1080/10473289.1990.10466789

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*, 709–742. https://doi.org/10.1080/10473289.2006.10464485

pypomp developers. (2024). *Pypomp: Partially observed Markov processes in Python*. https://github.com/pypomp/pypomp

Seinfeld, J. H., & Pandis, S. N. (2006). *Atmospheric chemistry and physics: From air pollution to climate change* (2nd ed.). John Wiley & Sons.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

STATS 531 course staff. (2025a). *Review comments on final project 10*.

STATS 531 course staff. (2025b). *Review comments on final project 17*. https://ionides.github.io/531w25/final_project/project17/comments.html

Stull, R. B. (1988). *An introduction to boundary layer meteorology*. Kluwer Academic Publishers.

Tai, A. P. K., Mickley, L. J., & Jacob, D. J. (2010). Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmospheric Environment*, *44*(32), 3976–3984. https://doi.org/10.1016/j.atmosenv.2010.06.060

U.S. Environmental Protection Agency. (2024). *Appendix N to part 50 — interpretation of the national ambient air quality standards for* $PM_{2.5}$ (Code of Federal Regulations 40 CFR Part 50, Appendix N). U.S. Environmental Protection Agency. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-50/appendix-Appendix%20N%20to%20Part%2050

Wikipedia contributors. (2026). *List of trigonometric identities*. https://en.wikipedia.org/wiki/List_of_trigonometric_identities

Yoon, S., Heo, Y., Park, C. R., & Kang, W. (2022). Effects of landscape patterns on the concentration and recovery time of PM2.5 in South Korea. *Land*, *11*, 2176. https://doi.org/10.3390/land11122176

Zhang, B., Jiao, L., Xu, G., Zhao, S., Li, J., Zhu, S., & Gu, Y. (2018). Influences of wind and precipitation on different-sized particulate matter concentrations (PM2.5, PM10, PM2.5–10). *Meteorology and Atmospheric Physics*, *130*, 383–392. https://doi.org/10.1007/s00703-017-0526-9

Zhou, L., Chen, X., & Tian, X. (2018). The impact of fine particulate matter (PM2.5) on China's agricultural production from 2001 to 2010. *Journal of Cleaner Production*, *178*, 133–141. https://doi.org/10.1016/j.jclepro.2017.12.204