From Statistical Analysis to Physics-Informed Reconstruction

A Diffusion-Based Framework for Surface PM2.5 Estimation

1. Summary of Published Academic Work

Journal 1: Exploiting the Results of Running the GEOS-CF Model to Evaluate PM2.5 Concentration in Near Real-Time in Vietnam. **Published:** *Journal of Hydro-Meteorology, 2024*

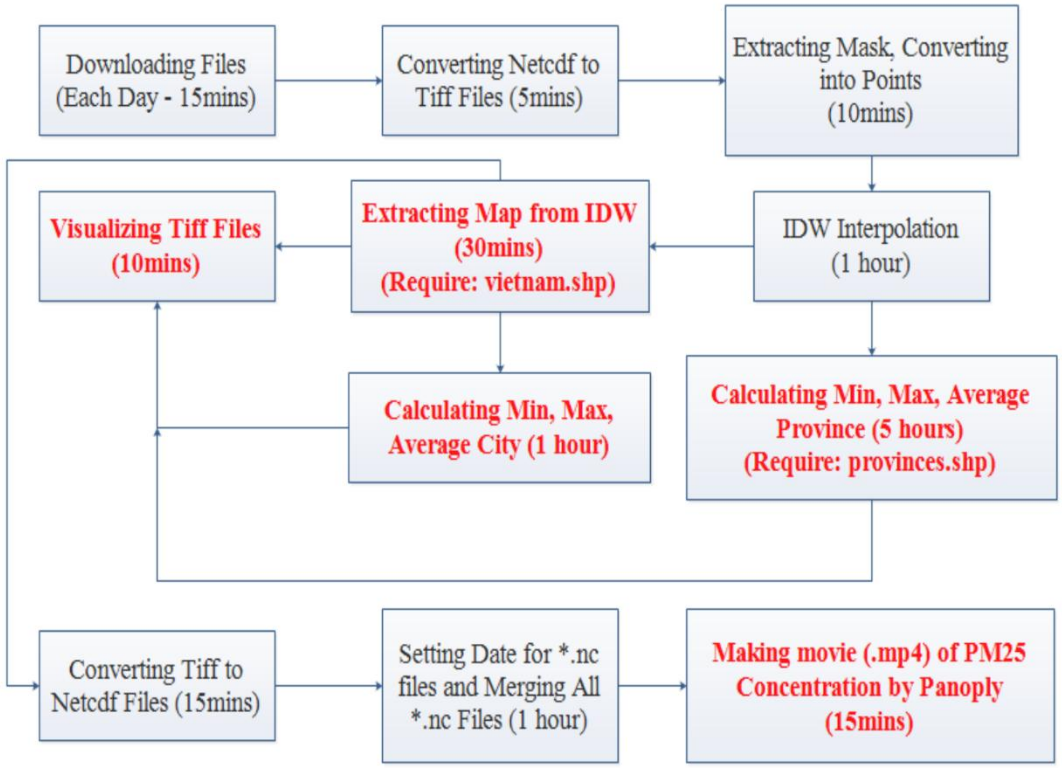
In this study, the authors focused on creating a tool to visualize and interpret global air quality data specifically for the entire territory of Vietnam.

Objective: The goal was to exploit the "GEOS-CF" (Goddard Earth Observing System Composition Forecast) model provided by NASA to build hourly $PM_{2.5}$ pollution maps for Vietnam. The authors aimed to provide near real-time information to help forecast air quality and manage environmental disasters.

htf_inst_15mn_g1440x721_x1: High Temporal Frequency Chemistry and Meteorology Diagnostics

Frequency: *15-minute from 00:00 UTC (instantaneous)*
Spatial Grid: *2D, single-level, full horizontal resolution*
Dimensions: *longitude=1440, latitude=721, level=1, time=1*
vertical level: *[72.] (layer)*
Granule Size: *~17 MB*
Shortname: CF01Rhtf_15mnI_g1440x721_X1
Mode: Replay and Forecast

Name	*Dim*	*Description*	*Units*
CO	tzyx	Carbon monoxide (CO, MW = 28.00 g mol-1) volume mixing ratio dry air	mol mol-1
NO2	tzyx	Nitrogen dioxide (NO2, MW = 46.00 g mol-1) volume mixing ratio dry air	mol mol-1
O3	tzyx	Ozone (O3, MW = 48.00 g mol-1) volume mixing ratio dry air	mol mol-1
PM25_RH35_GCC	tzyx	Particulate matter with diameter below 2.5 um RH 35	ug m-3
PM25_RH35_GOCART	tyx	Total reconstructed PM2.5 RH 35	kg m-3
Q	tzyx	specific humidity	kg kg-1
RH	tzyx	relative humidity after moist	1
SLP	tyx	sea level pressure	Pa
SO2	tzyx	Sulfur dioxide (SO2, MW = 64.00 g mol-1) volume mixing ratio dry air	mol mol-1
T	tzyx	air temperature	K
U	tzyx	eastward wind	m s-1
V	tzyx	northward wind	m s-1



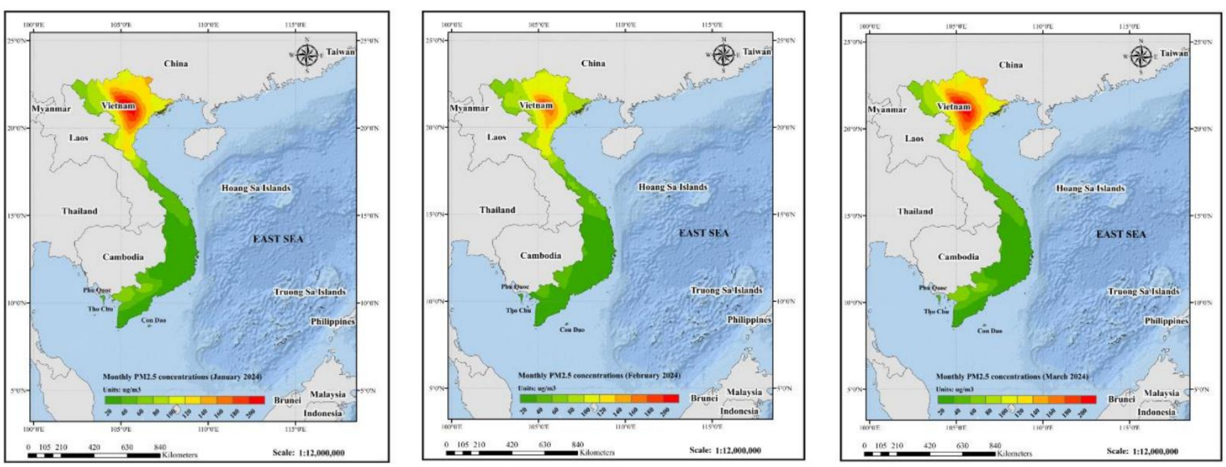


Figure 8. Average monthly $PM_{2.5}$ pollution distribution map from January to March 2024.

Table 1. Concentration by seven regions, Hoang Sa and Truong Sa Islands in January 2024.

Region	Concentration (μg/m³)	
	Min	Max
Northwest	40.69	174.27
East Northern	87.37	187.29
Red river delta	72.56	186.07

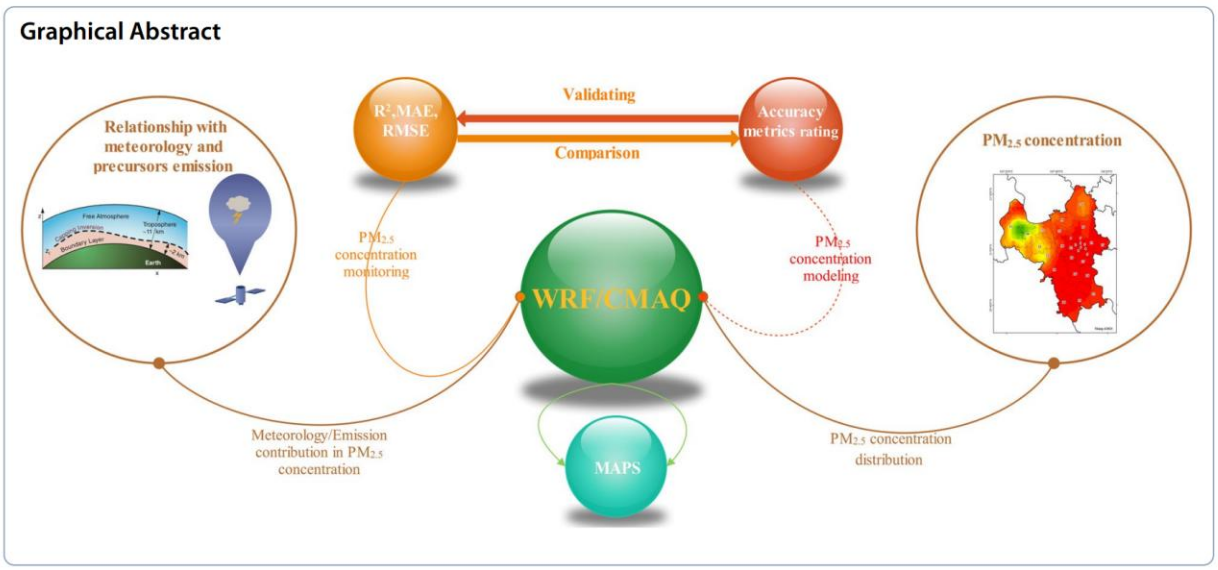
Journal 2: Exploiting the Results of Running the GEOS-CF Model to Evaluate PM2.5 Concentration in Near Real-Time in Vietnam. **Published:** *Journal of Hydro-Meteorology, 2024*

Title: Developing PM2.5 mitigation solutions based on the analysis of the relationships between PM2.5 concentrations and precursor factors: a case study of Hanoi, Vietnam

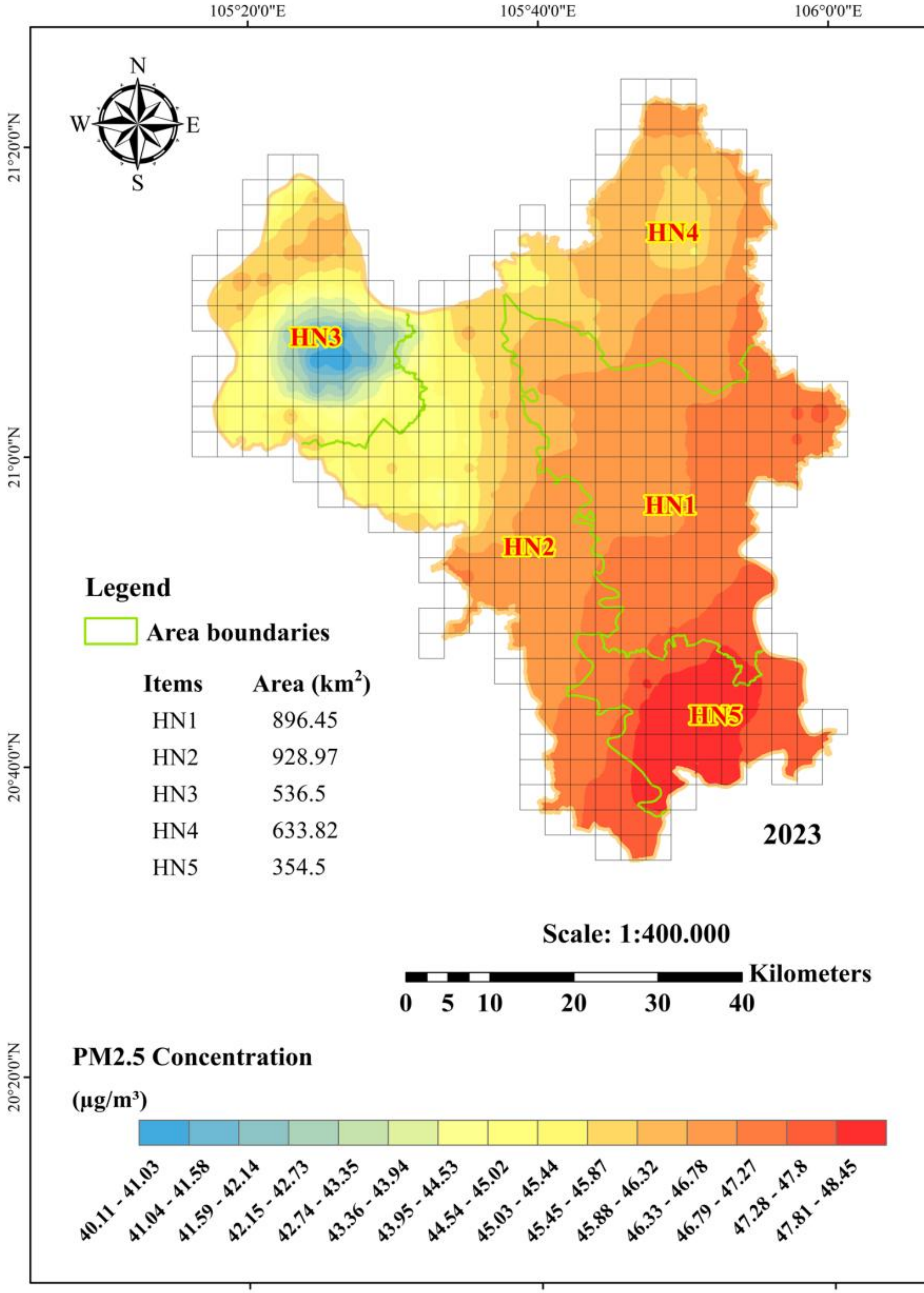
Published: Asian Journal of Atmospheric Environment, 2025

This study narrowed the scope to Hanoi to understand *why* pollution occurs by analyzing the chemical precursors and meteorological factors contributing to $PM_{2.5}$.

Objective: The study aimed to identify $PM_{2.5}$ cluster regions based on precursor emissions and propose mitigation solutions by clarifying the relationship between pollution concentrations, emissions, and meteorology.



	BC	CO	NH_3	NO_x	OC	SO_2	BENZEN	NMVOCs	VOCs
HN1	8611	22,809	670	4041	19,127	1617	95	556	99
HN2	26,779	63,891	1747	12,955	49,558	5342	280	1622	365
HN3	19,865	49,130	1244	9887	35,716	4136	214	1229	279
HN4	6098	15,831	480	2603	12,805	1042	61	388	63
HN5	3326	8283	312	1272	8288	537	41	203	38
Total	64,679	159,944	4453	30,757	125,494	12,673	691	3999	844


105°20'0"E
105°40'0"E
106°0'0"E
21°20'0"N
21°0'0"N
20°40'0"N
20°20'0"N
N
S
W
E
HN1
HN2
HN3
HN4
HN5
Legend
Area boundaries
Items
Area (km²)
HN1 896.45
HN2 928.97
HN3 536.5
HN4 633.82
HN5 354.5
2023
Scale: 1:400.000
Kilometers
0 5 10 20 30 40
PM2.5 Concentration
(µg/m³)
40.11 - 41.03
41.04 - 41.58
41.59 - 42.14
42.15 - 42.73
42.74 - 43.35
43.36 - 43.94
43.95 - 44.53
44.54 - 45.02
45.03 - 45.44
45.45 - 45.87
45.88 - 46.32
46.33 - 46.78
46.79 - 47.27
47.28 - 47.8
47.81 - 48.45

2. Motivation for a Physics-Informed Reconstruction Framework

PM2.5 is governed by physical processes including advection, turbulent diffusion, emission injection, and atmospheric dilution. Ignoring these mechanisms limits the reliability and interpretability of PM2.5 estimation.

To address the limitations of both statistical and purely data-driven approaches, this study proposes a diffusion-based reconstruction framework that integrates:

- Physical transport mechanisms
- Emission source information
- Meteorological drivers
- Ground-based observations
- Global chemical background constraints

The goal is not merely to predict PM2.5, but to reconstruct a physically consistent surface PM2.5 field over the entire study region.

2.1. State Variable: Let $C(x, y, z)$ = surface $PM_{2.5}$ concentration

2.2. Governing Equation (Core Diffusion Model)

Lagrangian Particle Transport – Advection and Diffusion

$$\frac{\partial C}{\partial t} = -\nabla \cdot (\vec{u}C) + \nabla \cdot (K\nabla C) + E(x, y, t) - D(x, y, t)$$

Where:

- $\vec{u}(x, y, t)$: wind field (from MCIP)
- $K(x, y, t)$: turbulent diffusivity

$$K = \alpha \cdot u_* \cdot h_{PBL}$$

- E: emissions (ANT + AIR + BIO)
- D: deposition (optional parameterization)

2.3. Initial Condition:

$$C(x, y, t_0) = C_{\text{GEOS-CF}}(x, y, t_0)$$

This ensures **chemical realism at large scale**.

2.4. Emission Modulation (Critical)

$$E_{\text{eff}} = E \cdot f(u, h_{PBL}, \text{stability})$$

$$f = \frac{1}{1 + \beta u} \cdot \frac{1}{h_{PBL}}$$

2.5. Sensor Assimilation

At sensor locations:

$$C^a = C^f + K_g(C^{obs} - HC^f)$$

Where:

- C^f: forecast (physics-only)
- C^{obs}: sensor PM2.5
- H: observation operator
- K_g: gain matrix

Gain weighting:

$$K_g \propto \exp\left(-\frac{d^2}{L^2}\right) \cdot \exp\left(-\frac{\theta^2}{\sigma_\theta^2}\right)$$

(distance + wind direction)