AeroChem: Closed-loop Physics-Informed State Space Modeling for Long-term Chemically-Reactive Air Quality Forecasting

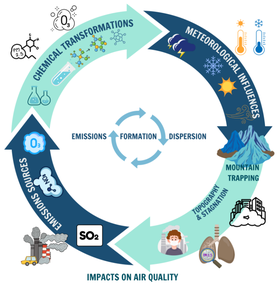
Kien Le[1*] Tung Kieu[2*] Nha Nguyen[1] Cuong Do[1] Bin Yang[3] Long Nguyen[4]

[1]VinUniversity, Vietnam [2]Aalborg University, Denmark [3]East China Normal University, China [4]Hanoi University of Industry, Vietnam [*]Equal contribution

1 Motivation

Long-term forecasting of **chemically reactive** pollutants is hard:

- **Transport & weather:** long-range transport and meteorological variability
- **Nonlinear chemistry:** secondary O_3 and $PM_{2.5}$ form from NO_x and VOCs under sunlight, heat and humidity
- **Drift:** data-driven models accumulate error over long rollouts; chemical transport models rely on simplified, costly solvers

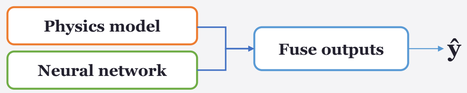


Coupled drivers of air quality

2 Key Idea: Close the Loop

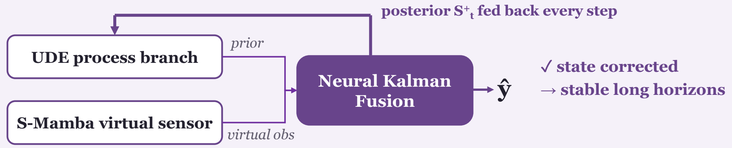
Prior physics which neural hybrids fuse the two branches **only at the output**: the latent physical state is never corrected, so errors compound over the horizon.

Prior hybrids — open loop



✗ state never corrected → long-horizon drift

AeroChem — closed loop



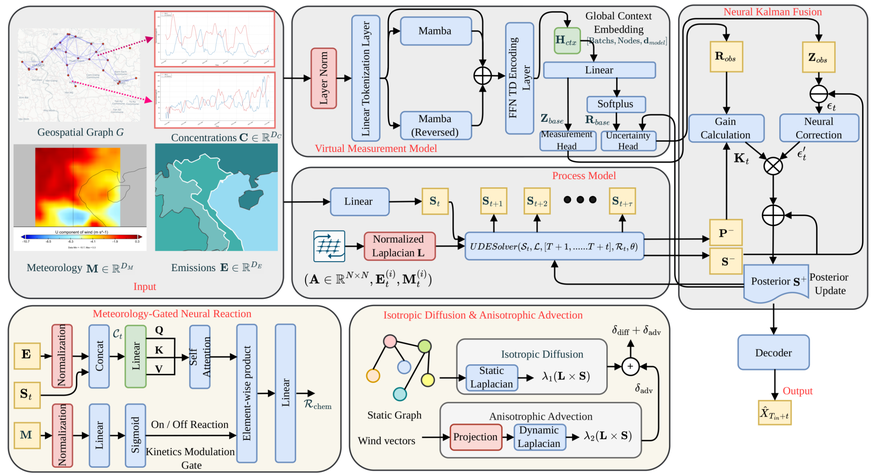
✓ state corrected → stable long horizons

3 HanoiAir: A New Regional Dataset

27	8 yrs	≈70K	3
monitoring stations	2018 – 2025	hourly time steps	data modalities

- **Pollutants:** $PM_{2.5}$, PM_{10}, O_3 and precursors NO_x, SO_2, CO, NO, NO_2
- **Meteorology:** temperature, pressure, solar radiation, humidity, wind
- **Emissions:** hourly-downscaled inventories (NO_x, VOC, SO_2, NH_4, primary $PM_{2.5}$)
- Northern Vietnam, from the Northern Centre for Environmental Monitoring; also evaluated on **KnowAir** (355 stations, 2016–2023)

4 AeroChem: Closed-loop Physics-Informed SSM



A Virtual Measurement Branch · S-Mamba

Encodes long-range history into probabilistic virtual observations that adapt to the rollout.

$$\mathbf{Z}_{\text{obs},t} = \mathbf{Z}_{\text{base}} + \alpha \tanh\left(\mathbf{S}^{+}_{t-1}\right), \qquad \mathbf{R}_{\text{obs},t} = \text{Softplus}(\mathbf{R}_{\text{base}})\left(1 + \beta \tfrac{t}{T_{\text{out}}}\right)$$

B Process Branch · Universal Differential Equation

Differentiable simulator: graph diffusion + wind-driven advection + meteorology-gated reaction.

$$d\mathbf{S}_t = \Big[\underbrace{\delta_{\text{diff}}}_{\text{diffusion}} + \underbrace{\delta_{\text{adv}}}_{\text{advection}} + \underbrace{\mathcal{R}_{\text{chem}}}_{\text{reaction}}\Big] dt + \mathbf{G}(\mathbf{S}_t)\, d\mathbf{W}_t \qquad \mathcal{R}_{\text{chem}} = \text{Attn}(\mathbf{Q},\mathbf{K},\mathbf{V}) \odot \underbrace{\sigma\big(\text{Linear}(\mathbf{M}^{\text{sc}}_t)\big)}_{\text{meteorology gate}}$$

C Neural Kalman Fusion · closed-loop correction

Fuses the UDE prior with virtual observations; the posterior state is fed back to the UDE.

$$\mathbf{K}_t = \mathbf{P}^{-}_t \oslash \left(\mathbf{P}^{-}_t + \mathbf{R}_{\text{obs},t}\right), \qquad \boldsymbol{\epsilon}'_t = \left(\mathbf{Z}_{\text{obs},t} - \mathbf{S}^{-}_t\right) + \text{Linear}\left([\mathbf{S}^{-}_t, \mathbf{Z}_{\text{obs},t}]\right)$$
$$\mathbf{S}^{+}_t = \mathbf{S}^{-}_t + \mathbf{K}_t \odot \boldsymbol{\epsilon}'_t, \qquad \mathbf{P}^{+}_t = (\mathbf{1} - \mathbf{K}_t) \odot \mathbf{P}^{-}_t$$

5 Theory: Bounded vs. Exploding Error

✓ Closed loop (AeroChem)

$$\sqrt{\mathbb{E}\|\mathbf{e}_t\|_2^2} \;\le\; \alpha^t \sqrt{\mathbb{E}\|\mathbf{e}_0\|_2^2} + \frac{V}{1-\alpha}$$

uniformly bounded in t when the contraction factor α < 1

✗ Open loop (no feedback)

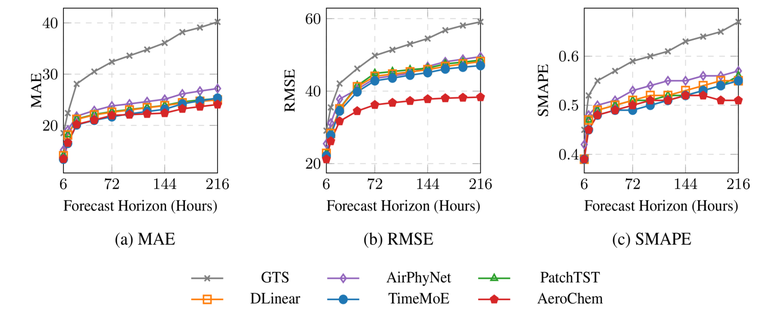
$$\mathbb{E}\|\mathbf{e}_t^{\text{open}}\|_2^2 \;\ge\; (1 + L_F \Delta t)^{2t}\, \mathbb{E}\|\mathbf{e}_0\|_2^2$$

grows exponentially in t whenever $L_F > 0$

6 Results: $PM_{2.5}$ Forecasting

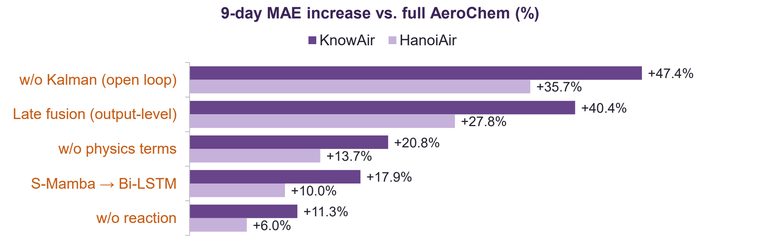
MAE ↓	KnowAir			HanoiAir		
	3 days	9 days	Sudden	3 days	9 days	Sudden
AirPhyNet	23.80	27.20	38.50	43.80	49.20	71.50
Air-DualODE	23.04	25.20	39.24	42.04	48.20	72.24
DLinear	22.57	24.50	38.45	42.57	47.10	73.45
PatchTST	22.69	25.25	39.12	42.10	47.80	74.12
S-Mamba	23.12	25.15	39.85	43.12	47.15	75.85
Fredformer	**21.45**	26.48	38.25	**41.60**	47.48	73.25
TimeMoE	21.68	25.32	36.54	41.85	47.50	71.54
AeroChem (ours)	21.95	**24.08**	**34.50**	42.74	**46.80**	**68.80**

Mean of 5 runs. **Bold** / underline = best / second best.



(a) MAE (b) RMSE (c) SMAPE

7 Ablation: Feedback Matters Most



8 Acknowledgments

This work was supported by the Center for Environmental Intelligence, VinUniversity.

kien.lt@vinuni.edu.vn