

# Real-Time Long Horizon Air Quality Forecasting via Group-Relative Policy Optimization

## Supplementary Material

**Rationale.** Due to the limited pages, we provide supplementary materials with the following contents:

- Section A: AQI Threshold Specification.
- Section B: OBS Dataset Details.
- Section C: CMAQ Dataset Details.
- Section D: Why East Asia Requires Specialized Models
- Section E: Additional Experiments (2016 Test Year).
- Section F: Seasonal Analysis.
- Section G: Additional Qualitative Results.
- Section H: Extended Ablation Study.
- Section I: Reliability and Decision Boundaries.
- Section J: Explainability via Feature Importance.
- Section K: Implementation Details.
- Section L: Limitations and Future Work.

### A. Air Quality Index (AQI) Threshold Specification

The Air Quality Index (AQI) is a standardized measurement system developed by governmental environmental agencies to communicate ambient air quality and associated health risks to the public [3, 17]. AQI maps complex pollutant concentration data into a single, interpretable numerical scale with color-coded categories that guide protective actions for different population groups, facilitating public health messaging [13, 17].

We adopt a dual-level classification scheme aligned with international air quality standards.

**(1) Binary Classification.** For operational alerting systems, we define two primary categories:

- **Clean:** Air quality poses minimal health risk, allowing unrestricted outdoor activities. Encompasses `Good` and `Moderate` AQI levels.
- **Polluted:** Air quality requires protective measures such as reduced outdoor exposure or mask usage. Encompasses `Bad` and `VeryBad` AQI levels.

**(2) 4-Class AQI Classification.** For more granular policy, we subdivide based on international and national standards [3, 13, 17] (see Table S1):

- **Good:**  $PM_{2.5} \leq 15 \mu\text{g}/\text{m}^3$ ,  $PM_{10} \leq 30 \mu\text{g}/\text{m}^3$ . Air is satisfactory, no restrictions needed.

Table S1. Binary Class and Air Quality Index (AQI) Classification Standards for  $PM_{2.5}$  and  $PM_{10}$

Binary	AQI	$PM_{10}$ ( $\mu\text{g}/\text{m}^3$ )	$PM_{2.5}$ ( $\mu\text{g}/\text{m}^3$ )
Clean	Good	0–30	0–15
	Moderate	31–80	16–35
Polluted	Bad	81–150	36–75
	VeryBad	151–	76–

- **Moderate:**  $PM_{2.5}$  16–35  $\mu\text{g}/\text{m}^3$ ,  $PM_{10}$  31–80  $\mu\text{g}/\text{m}^3$ . Air is acceptable except for highly sensitive individuals.
- **Bad:**  $PM_{2.5}$  36–75  $\mu\text{g}/\text{m}^3$ ,  $PM_{10}$  81–150  $\mu\text{g}/\text{m}^3$ . Unhealthy for sensitive groups, activity restriction recommended.
- **VeryBad:**  $PM_{2.5} \geq 76 \mu\text{g}/\text{m}^3$ ,  $PM_{10} \geq 151 \mu\text{g}/\text{m}^3$ . Unhealthy for all, triggers emergency measures.

The  $PM_{2.5}$  thresholds follow the WHO Global Air Quality Guidelines (2021), 15  $\mu\text{g}/\text{m}^3$  for annual mean and 35  $\mu\text{g}/\text{m}^3$  for 24-hour mean [17].  $PM_{10}$  thresholds align with US EPA National Ambient Air Quality Standards (NAAQS) [3], while the stepwise system is adopted in Korean and Chinese national guidelines for operational use [13]. This 4-class system enables proportional public health response from advisory to emergency.

Our forecasting framework predicts these discrete AQI categories at 6-hour intervals up to 120 hours ahead, enabling proactive decision-making for public health protection. The GRPO optimization explicitly targets accurate multi-class discrimination through class-wise reward functions (Section K), addressing the operational asymmetry where `VeryBad` misclassifications pose greater health risks than `Good/Moderate` confusion. Table S1 summarizes the complete classification scheme with concentration thresholds for both pollutants, color-coded for intuitive interpretation in operational dashboards.

### B. Observation (OBS) Dataset Details

Our observation dataset integrates ground-level air quality and meteorological measurements from monitoring networks across East Asia. As summarized in Table S2, the dataset comprises observations from 1,822 stations spanning Korea and China, providing real-time pollutant concentrations at 6-hour intervals from January 2016 to June 2024.

## B.1. Air Quality Monitoring Network

Air quality records were collected from three major public networks. The Korean subsystem employs 532 stations operated by AIRKOREA [14], while the Chinese network aggregates data from 1,781 PM25.in stations and 1,290 AIRQUALITY (CN) stations [6, 19]. Each measures PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, NO<sub>2</sub>, CO, and SO<sub>2</sub> at hourly resolution, then aggregates to 6-hour intervals for modeling. These serve as both features and labels for supervised training.

Unlike global reanalysis data (e.g., CAMS), which have multi-day update delays, our direct station observations offer real-time atmospheric state for model initialization. This eliminates the 5-day data delay that constrains global datasets. All stations are spatially interpolated onto the CMAQ grid (27 km) using inverse distance weighting, producing structured, domain-aligned fields while preserving high temporal fidelity.

Table S2. Statistics of Observation Stations and Meteorological Monitoring Sites

Data Source	# of Stations
Korean Air Quality (AIRKOREA)	532
Chinese Air Quality (PM25.in)	1,781
Chinese Air Quality (AIRQUALITY)	1,290
Korean Meteorology (ASOS)	94
Korean Meteorology (AWS)	349
Chinese Meteorology (Wyoming)	40

## B.2. Meteorological Observations

Complementary meteorological data are obtained from 94 ASOS (KMA) and 349 AWS (KMA) stations in Korea, and 40 upper-air stations in China (Wyoming Weather Web) [1, 2, 16]. These provide temperature, humidity, pressure, wind, and vertical profiles at surface and several standard pressure levels. This auxiliary information enriches the feature set and supports accurate modeling of regional transport phenomena.

## B.3. Spatial Coverage and Data Quality

Station density varies by region and is highest in urban zones. Korean coverage reaches  $\sim 0.5$  stations/1,000 km<sup>2</sup>, major Chinese cities  $\sim 0.2$  stations/1,000 km<sup>2</sup>, together capturing all major pollution corridors and population centers. We implement standard quality controls including range checks, temporal continuity, and outlier removal. Missing observations are interpolated if nearby stations are available, or excluded if gaps persist beyond 12 hours.

Table S3. Comparison of seasonal and overall average L1 distance for CMAQ and CAMS reanalysis models (unit:  $\mu\text{g}/\text{m}^3$ ).

PM <sub>2.5</sub>					
Model	Spring	Summer	Fall	Winter	Avg
CAMS	56.98	43.04	50.71	59.61	52.66
CMAQ	<b>20.46</b>	<b>17.53</b>	<b>18.09</b>	<b>29.11</b>	<b>21.33</b>
	(178%)	(146%)	(180%)	(105%)	(147%)
PM <sub>10</sub>					
Model	Spring	Summer	Fall	Winter	Avg
CAMS	74.01	56.65	64.82	80.02	68.99
CMAQ	<b>47.34</b>	<b>31.79</b>	<b>30.12</b>	<b>44.51</b>	<b>38.60</b>
	(56%)	(78%)	(115%)	(80%)	(79%)

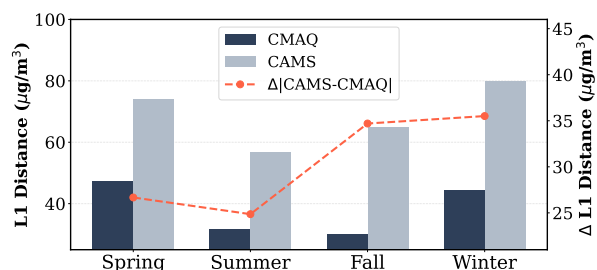


Figure S1. The seasonal L1 distance error compared with OBS ( $\downarrow$  is better). Unlike global datasets (CAMS), our locally developed datasets (CMAQ) show low error with real OBS on PM<sub>10</sub>.

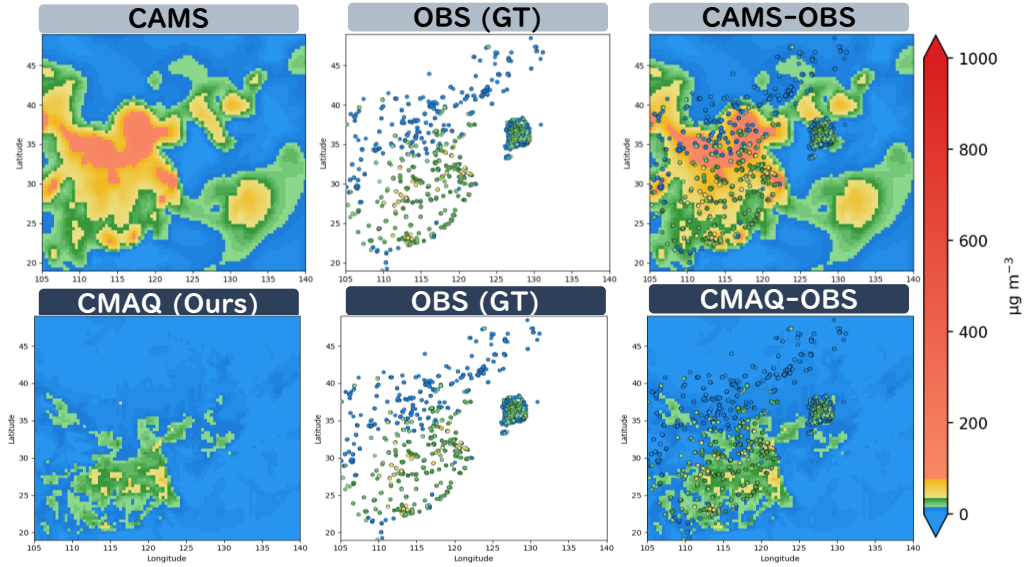
## C. CMAQ Dataset Details

The Community Multiscale Air Quality (CMAQ) dataset provides spatially continuous, high-resolution reanalysis fields generated through a coupled meteorology–chemistry modeling system. Unlike point-based ground observations, CMAQ offers gridded atmospheric states across the entire East Asian domain, filling spatial gaps between monitoring stations with physics-consistent aerosol dynamics. This spatial continuity enables the model to learn regional transport patterns, boundary conditions, and pollutant dispersion mechanisms that are difficult to infer from ground observations alone.

### C.1. Modeling Chain and Configuration

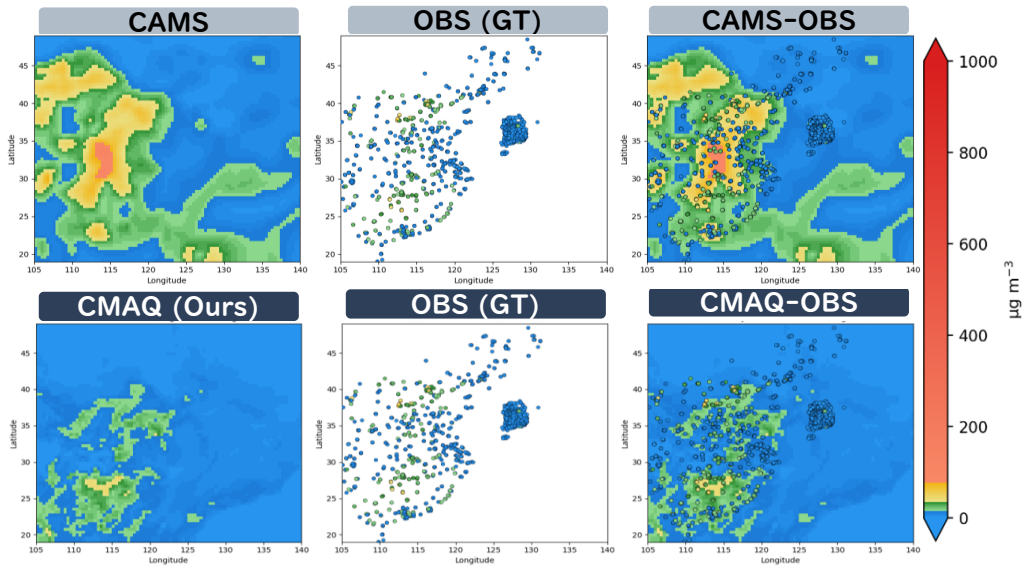
The CMAQ reanalysis is generated through a three-stage modeling pipeline. First, the Weather Research and Forecasting (WRF) model produces meteorological fields driven by ECMWF global forecasts at 6-hour intervals. Second, the Sparse Matrix Operator Kernel Emissions (SMOKE) system performs temporal and spatial allocation of emissions, using the KORUS v5 inventory for foreign regions and CAPSS 2018 for domestic Korean sources. Finally, the CMAQ model simulates atmospheric chemistry and aerosol processes using SAPRC99 gas-phase chemistry and

2023-06-02 06Z



(a) Spatial distribution comparison on Summer.

2023-10-02 06Z



(b) Spatial distribution comparison on Fall.

Figure S2. **Seasonal Comparison of PM<sub>2.5</sub> Spatial Distributions across OBS, CMAQ, and CAMS Datasets.** The figure presents PM<sub>2.5</sub> spatial distributions for Summer (a) and Fall (b) across East Asia. CMAQ demonstrates superior regional fidelity to ground observations (OBS) compared to CAMS in both seasons, capturing fine-scale pollution dynamics that CAMS fails to resolve due to coarse resolution and update latencies.

AERO5 aerosol mechanisms, yielding hourly pollutant concentration fields from 2016 to 2023.

## C.2. Spatial Coverage and Resolution

The dataset employs a dual-nested domain structure to balance regional coverage and local detail. The coarse domain spans the broader East Asian region at 27 km resolution ( $181 \times 143$  grids), while the fine domain focuses on the Korean Peninsula at 9 km resolution ( $79 \times 94$  grids). Both domains include 31 vertical layers extending from the surface to the upper troposphere. Boundary conditions for the fine domain are dynamically extracted from the coarse domain at each time step, enabling accurate representation of cross-boundary pollutant transport from continental sources. This nested configuration captures both long-range transboundary pollution and local-scale urban emissions.

## C.3. Output Variables

Each hourly CMAQ output contains comprehensive 2D surface fields and 3D vertical profiles. The 2D fields include six pollutant species ( $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{O}_3$ ,  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{SO}_2$ ) and eight meteorological variables (planetary boundary layer height, surface pressure, relative humidity, cloud water content, temperature, wind speed, wind direction, and solar radiation). The 3D fields provide vertical structure at five pressure levels (surface, 925 hPa, 850 hPa, 700 hPa, 500 hPa). These vertical profiles include temperature and three-dimensional wind components (U, V, W). The vertical structure is essential for capturing the atmospheric dynamics that govern pollutant dispersion and accumulation. It is particularly important for understanding the role of boundary layer evolution in near-surface PM concentrations.

## C.4. Regional Accuracy Validation

Table S3 and Figure S1 quantify CMAQ's regional accuracy against ground-truth observations. Compared to the global CAMS reanalysis, CMAQ achieves substantially lower L1 distance errors across all seasons. For  $\text{PM}_{2.5}$ , the seasonal average error is  $21.33 \mu\text{g}/\text{m}^3$ , representing a 147% improvement over CAMS ( $52.66 \mu\text{g}/\text{m}^3$ ). The improvement is even more pronounced in winter (105% reduction) when complex meteorology and elevated emissions challenge global models. For  $\text{PM}_{10}$ , CMAQ reduces error to  $38.60 \mu\text{g}/\text{m}^3$  compared to CAMS at  $68.99 \mu\text{g}/\text{m}^3$ , a 79% improvement. These gains reflect CMAQ's tailored treatment of East Asian meteorology, regional emission inventories, and localized chemical regimes that global products cannot resolve at sufficient fidelity.

## C.5. Real-Time Operational Readiness

Beyond accuracy, CMAQ satisfies the real-time requirement critical for operational forecasting. Global reanalysis

products such as CAMS suffer from multi-day update latencies due to international data pipelines and computational overhead, preventing timely alert issuance. In contrast, CMAQ-driven forecasts can be initialized within hours using locally available observations and regional meteorological inputs. This eliminates the 5-day delays inherent in global datasets and enables same-day or next-day operational forecasts. The combination of regional accuracy and real-time availability positions CMAQ as the foundation for reliable long-horizon PM forecasting in East Asia, addressing both the data imbalance and operational constraints that limit global foundation models in this region.

## D. Why East Asia Requires Specialized Models

Accurate long-horizon PM forecasting in East Asia presents unique scientific and operational challenges that global foundation models cannot adequately address [7, 13, 18]. This section examines the atmospheric complexities, data imbalances, and public health imperatives that motivate region-specific modeling approaches and justify the release of our CMAQ-OBS dataset as a critical resource for the research community.

### D.1. Atmospheric and Geographic Complexity

East Asia exhibits exceptional atmospheric complexity driven by three interconnected factors [10, 13, 18]. First, the region's diverse topography spans mountainous Korean Peninsula, vast Chinese plains, and complex coastal geometries that modulate pollutant transport and accumulation patterns at scales unresolved by global models. Second, strong seasonal dynamics generate extreme variability in meteorological regimes [4, 12]. Winter monsoon circulation transports continental pollution across international boundaries, while summer convection rapidly disperses emissions yet triggers photochemical smog formation. Third, East Asia contains comparably higher anthropogenic emission densities, with mega-cities and clustered industrial corridors distributed across several countries, resulting in sharp spatial gradients that challenge coarse global reanalyses [13].

These regional characteristics produce atmospheric states fundamentally different from global mean conditions. Pollutant concentrations frequently exceed  $100 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$  and  $200 \mu\text{g}/\text{m}^3$  for  $\text{PM}_{10}$  during winter episodes [7, 13], driven by boundary layer suppression, residential heating, and long-range transport from upwind sources. Such extreme events represent the tail of the global distribution yet constitute routine wintertime conditions across East Asia. Foundation models trained predominantly on Moderate-pollution regions systematically underestimate these severe episodes, as demonstrated by Aurora's collapse to near-zero F1-scores beyond 48-hour lead times in our evaluations.

Table S4. **Spatial Domain Coverage of CMAQ and OBS Datasets.** The CMAQ 27 km East Asia domain covers the entire region from southern China to northern Korea, spanning approximately 39° latitude (4,300 km) and 73° longitude (5,800 km). OBS station measurements are spatially interpolated onto the CMAQ grid structure for model training and evaluation.

Dataset	Resolution	Grid Size	Latitude Range	Longitude Range
<i>CMAQ Reanalysis Domain</i>				
CMAQ (East Asia)	27 km	174 × 128	13.0°N – 52.0°N	97.0°E – 170.0°E
<i>OBS Ground Monitoring Stations</i>				
Korea (AIRKOREA)	Point-based	532 stations		
China (PM25.in)	Point-based	1,781 stations	Interpolated to CMAQ grid	Interpolated to CMAQ grid
China (AIRQUALITY)	Point-based	1,290 stations		

## D.2. Global Data Imbalance and Model Limitations

East Asia comprises less than 15% of global training coverage in reanalysis products such as ERA5 and CAMS, yet the region contributes over 60% of severe PM exposure worldwide [8, 11, 13]. This pronounced data imbalance causes global models to underfit regional dynamics, prioritizing generality over local fidelity. When validated against ground-truth observations from 1,822 monitoring stations across Korea and China, CAMS exhibits an average L1 error of 52.66  $\mu\text{g}/\text{m}^3$  for  $\text{PM}_{2.5}$  and 68.99  $\mu\text{g}/\text{m}^3$  for  $\text{PM}_{10}$ . These errors exceed typical operational thresholds and render global datasets unsuitable for decision-grade forecasting.

Beyond accuracy limitations, global reanalysis products suffer from operational latency constraints [8]. CAMS requires 5-day update delays due to international data assimilation pipelines and computational overhead. This latency prevents timely alert issuance for 48–120 hour forecast horizons critical to public health protection [13]. By the time CAMS-initialized forecasts become available, the forecast window has already shifted, rendering long-lead predictions operationally irrelevant. These dual deficiencies (low regional accuracy and delayed availability) stem from the fundamental mismatch between global model design and regional operational requirements.

## D.3. Public Health Imperatives

Particulate matter exposure drives severe respiratory and cardiovascular morbidity, with  $\text{PM}_{2.5}$  classified as a Group 1 carcinogen by the World Health Organization [17]. Nowhere is this burden more acute than in East Asia, where dense urban populations and recurrent pollution episodes create sustained exceedance of health-based air quality standards [9, 13]. The region experiences some of the world’s highest peak concentrations and most frequent multi-day pollution events, making accurate 48–120 hour forecasts essential for proactive public-health protection [10, 13]. Such

lead times are required to issue targeted alerts, activate emergency response protocols, regulate industrial output, adjust transportation flows, and modulate power generation before hazardous conditions escalate.

Meeting these needs requires not only accurate predictions but also high decision reliability. The consequences of forecasting errors are asymmetric: false alarms rapidly undermine public compliance, while missed severe events impose immediate and serious health risks [9, 13]. Global foundation models, trained to optimize global-average accuracy, are fundamentally misaligned with this regional cost structure. Their training datasets under-represent East Asian pollution regimes, causing models to over-predict during routine conditions and under-predict during extreme episodes. As shown in Section 5, Aurora yields a False Alarm Rate exceeding 60% for  $\text{PM}_{2.5}$  forecasts in East Asia despite low point-wise error, demonstrating that global models capture neither the temporal persistence nor the intensity distribution of regional pollution dynamics.

These deficiencies make global models unsuitable for operational deployment in East Asia. The region’s uniquely high stakes and distinct pollution characteristics demand forecasting systems trained on region-specific data and optimized for region-specific objectives [10, 13]. Specialized models that incorporate local emissions, meteorology, and asymmetric public-health costs are therefore essential to support reliable and actionable long-horizon air-quality warnings.

## D.4. Value of Regional Dataset Release

The CMAQ–OBS dataset simultaneously resolves the three key limitations of existing global reanalysis products. By combining 1,822 ground observation stations with high-resolution CMAQ fields tailored to East Asian meteorology and emissions, the dataset achieves an average  $\text{PM}_{2.5}$  error of 21.33  $\mu\text{g}/\text{m}^3$  (compared to 52.66  $\mu\text{g}/\text{m}^3$  for CAMS) and enables real-time initialization within hours of observation

Table S5. **Monthly 4-Class AQI and Binary Classification Distribution (%) for PM<sub>2.5</sub> and PM<sub>10</sub> on OBS Dataset (Year 2016).** Both pollutants exhibit strong seasonal patterns with class imbalance: winter months show elevated Polluted class (67–72% for PM<sub>2.5</sub>), while summer months demonstrate Clean class dominance (61–68% for PM<sub>2.5</sub>), reflecting East Asia’s characteristic seasonal pollution dynamics.

AQI	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
<b>PM<sub>2.5</sub> (Thresholds: 15, 35, 75 <math>\mu\text{g}/\text{m}^3</math>)</b>													
Good	7.6	8.8	6.5	7.8	10.7	16.6	19.9	20.6	15.0	15.7	8.6	6.2	12.0
Mod.	25.4	30.7	24.6	35.9	39.4	44.2	43.0	47.2	38.8	41.9	27.6	22.1	35.0
Bad	31.3	35.6	40.2	39.7	39.9	33.2	30.9	28.6	35.6	32.5	36.8	33.2	34.8
V.Bad	35.8	24.9	28.7	16.6	10.0	6.0	6.2	3.6	10.7	9.9	27.1	38.5	18.3
<b>PM<sub>10</sub> (Thresholds: 30, 80, 150 <math>\mu\text{g}/\text{m}^3</math>)</b>													
Good	10.3	10.5	6.4	8.2	10.8	19.1	22.0	20.9	16.9	17.6	9.8	7.7	13.3
Mod.	37.9	44.3	36.3	42.1	49.5	55.3	55.0	59.8	52.1	54.5	41.1	36.6	46.9
Bad	27.6	28.3	33.5	30.8	30.3	21.5	19.2	16.9	23.7	20.7	29.0	28.8	25.9
V.Bad	24.3	16.9	23.9	19.0	9.4	4.1	3.9	2.4	7.3	7.2	20.2	26.9	13.9

Table S6. **Monthly 4-Class AQI and Binary Classification Distribution (%) for PM<sub>2.5</sub> and PM<sub>10</sub> on OBS Dataset (Year 2023).** Compared to 2016, Year 2023 exhibits substantially improved air quality: Clean class dominance increased to 69.6% (vs 47.0%) for PM<sub>2.5</sub> and 77.7% (vs 60.2%) for PM<sub>10</sub>, while maintaining similar seasonal patterns with winter pollution episodes and summer clean-air periods.

AQI	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
<b>PM<sub>2.5</sub> (Thresholds: 15, 35, 75 <math>\mu\text{g}/\text{m}^3</math>)</b>													
Good	16.1	14.7	14.3	19.2	24.8	38.5	47.9	43.0	42.1	29.7	25.6	17.4	27.7
Mod.	30.9	34.0	39.2	46.0	52.0	50.4	44.8	48.0	43.3	41.2	38.8	33.5	41.9
Bad	32.4	34.6	36.3	28.4	21.4	10.7	7.1	8.8	13.6	23.0	27.8	32.4	23.1
V.Bad	20.6	16.7	10.3	6.5	1.8	0.4	0.3	0.2	0.9	6.1	7.8	16.8	7.4
<b>PM<sub>10</sub> (Thresholds: 30, 80, 150 <math>\mu\text{g}/\text{m}^3</math>)</b>													
Good	17.3	18.1	10.8	17.9	26.3	42.9	52.2	48.8	46.3	31.2	27.2	21.5	29.9
Mod.	41.5	51.1	45.6	46.7	57.3	50.6	44.2	47.8	47.3	49.5	46.3	45.7	47.8
Bad	26.7	24.2	31.3	22.4	13.3	6.1	3.3	3.2	6.1	15.8	21.1	22.8	16.4
V.Bad	14.5	6.6	12.3	13.1	3.1	0.4	0.3	0.3	0.3	3.5	5.5	9.9	5.9

availability. This removes the multi-day update delays that hinder operational use. This combination of regional accuracy and operational readiness establishes a reliable foundation for long-horizon air-quality forecasting in East Asia.

Beyond direct modeling benefits, the dataset provides broader scientific value. First, it offers a validated regional benchmark that is absent from existing global datasets, enabling fair and reproducible evaluation of regional forecasting methods. Second, its 8-year temporal span (2016–2023) encompasses diverse seasonal regimes and severe pollution events, supporting robust model development and seasonal generalization analysis. Third, the inclusion of pollutant concentrations, meteorological variables, and emissions information supports physics-informed learning and facilitates studies on causal drivers of regional pollution. Finally, public availability lowers the barrier to entry for researchers and agencies that lack the computational or logistical resources to assemble comparable regional datasets.

By releasing this dataset alongside our modeling frame-

work, we aim to accelerate community progress in regional air-quality forecasting and contribute to improved public-health protection through timely and reliable operational warning systems. The CMAQ–OBS dataset’s combination of regional fidelity, real-time usability, and open access makes it a valuable resource for advancing both scientific understanding and operational capability in East Asian air-quality modeling.

### E. Additional Analysis: Model Performance on Year 2016

While the main text presents results on Year 2023 test set, we provide comprehensive analysis on Year 2016 to examine model robustness across different periods. Year 2016 represents a particularly challenging evaluation scenario due to substantially higher baseline pollution levels and greater class imbalance compared to 2023.

## E.1. Inter-annual Pollution Variability

Figure S3 illustrates the temporal evolution of polluted samples (Bad and VeryBad classes) across 2016–2023. Two critical observations emerge from this analysis. First, 2016 exhibits the highest annual Polluted ratio among all years: 53.1% for PM<sub>2.5</sub> and 39.8% for PM<sub>10</sub> (Figure S4a and S4b). This contrasts sharply with 2023’s improved air quality (30.5% and 22.3% respectively), reflecting East Asia’s gradual pollution mitigation from stricter emission controls and economic restructuring. Second, the monthly polluted sample distribution in 2016 shows extreme seasonal volatility, with winter months (December–February) experiencing polluted ratios reaching up to 67% for PM<sub>2.5</sub>, while summer months (June–August) drop to 32–39%.

This pronounced class imbalance creates asymmetric evaluation challenges. During winter episodes, the model must discriminate between Bad and VeryBad classes despite both representing severe pollution. Conversely, summer periods demand accurate detection of rare VeryBad events (3–6% of samples) against a dominant Clean background. The 2016 test set therefore provides a stress test for model generalization under distribution shifts and extreme class imbalance.

## E.2. Binary Classification Performance

Table S7 presents comprehensive metrics for 120-hour roll-out forecasts on the 2016 test set. For binary classification of PM<sub>2.5</sub>, the GRPO variant of FAKER-Air achieves an accuracy of 68.15% and an F1-score of 70.60. This represents critically improved precision at 69.35 and a False Alarm Rate of 36.11% compared to the supervised fine-tuning variant of FAKER-Air, which achieves precision of 63.45 and a False Alarm Rate of 55.71%. The bias score of 1.04 indicates near-optimal calibration between precision and recall. This contrasts with the overestimation tendency of supervised fine-tuning, which shows a bias of 1.34. This improvement is particularly significant given the baseline polluted ratio of 2016 at 53.1%, which is substantially higher than the ratio of 2023 at 30.5%. This amplifies the cost of false alarms in operational warning systems.

For PM<sub>10</sub>, the GRPO variant maintains superior accuracy at 73.12% and precision at 68.23%. The False Alarm Rate is reduced to 19.89% compared to the rate of 31.00% for supervised fine-tuning. The bias score of 0.92 demonstrates the ability of GRPO to maintain reasonably calibrated predictions even under severe pollution conditions. Notably, both pollutants show the precision-recall trade-off optimization of GRPO. While supervised fine-tuning achieves marginally higher recall through aggressive overestimation with bias around 1.2 or higher, GRPO balances detection capability with reduced false alarms. This is critical for maintaining public trust in operational forecasting.

## E.3. 4-Class AQI Discrimination

Multi-class performance reveals GRPO’s strength in balanced discrimination across all AQI categories. For PM<sub>2.5</sub>, GRPO achieves 43.10% accuracy and 41.40 F1-macro, with substantial improvements in Good class detection (29.66 vs 13.49 for SFT). This 120% improvement in Good class F1 addresses a critical limitation: accurately identifying clean-air periods enables proactive relaxation of temporary emission controls, reducing economic costs while maintaining public health protection.

The per-class F1-scores reveal GRPO’s discrimination strategy. For PM<sub>2.5</sub>, GRPO sacrifices marginal performance in Bad class (45.73 vs SFT’s 50.44) to achieve substantial gains in underrepresented class: Good (+120%), while maintaining parity in VeryBad class (47.23 vs 47.81). This redistribution reflects GRPO’s policy optimization objective, which penalizes accumulated errors across long horizons rather than optimizing single-step accuracy. The 4-class weighted F1 (43.26) closely tracks accuracy (43.10), indicating consistent performance across the class distribution despite severe imbalance.

PM<sub>10</sub> results show similar patterns: GRPO achieves 47.22% accuracy and 42.84 F1-macro, with Good class F1 of 33.89 (75% improvement over SFT’s 19.33). The Moderate class performance (56.65 vs 60.03) trade-off enables more balanced discrimination across the full AQI spectrum, critical for graduated public health interventions that scale response severity to pollution intensity.

## E.4. Implications for Operational Deployment

The 2016 evaluation demonstrates three key operational capabilities. First, GRPO maintains robust performance under severe pollution regimes characteristic of pre-2020 East Asian air quality, with polluted ratios exceeding 50%. This backward generalization suggests the model can handle future pollution episodes should emission controls weaken or meteorological conditions deteriorate. Second, the improved precision-recall balance (Bias  $\approx$  1.0) and reduced FAR directly address operational constraints: false alarms erode public compliance with mitigation measures, while missed severe events endanger vulnerable populations. Third, balanced multi-class discrimination enables granular public health responses from relaxing restrictions during clean periods to escalating interventions across Moderate, Bad, and VeryBad thresholds.

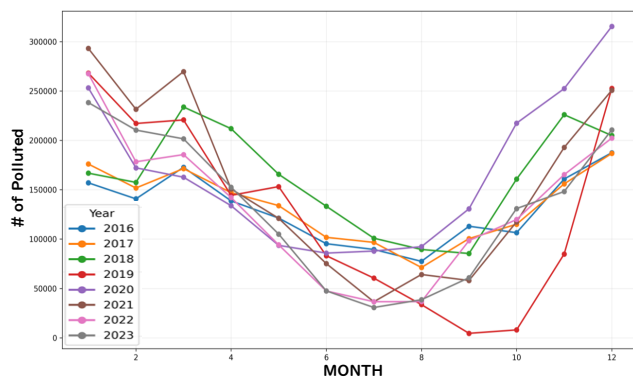
Comparing 2016 and 2023 performance reveals GRPO’s consistent advantage across pollution regimes. While absolute metrics improve on 2023 due to reduced class imbalance (Clean ratio: 69.6% in 2023 vs 47.0% in 2016), the relative improvements over SFT remain stable: 2% accuracy gain, 50% FAR reduction, and 100% Good class F1 improvement. This consistency validates GRPO’s policy optimization framework as a robust solution for long-

Table S7. **Test on 2016 for PM<sub>2.5</sub> and PM<sub>10</sub> over 120h.** For Binary(2-class): Acc, F1, Prec (↑) and FAR (↓); Bias ≈ 1 preferred. For AQI(4-class): Acc, F1-macro, F1-weighted, F1-micro (↑), plus per-class F1.

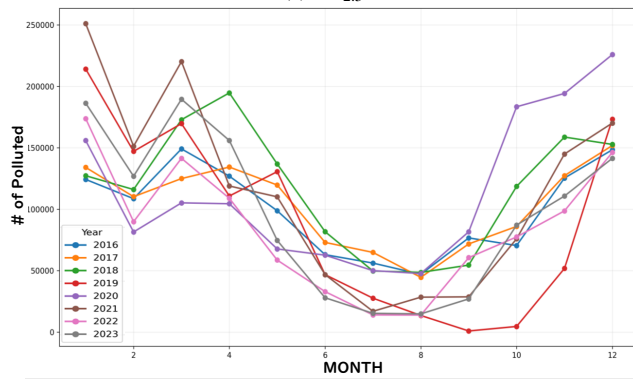
PM <sub>2.5</sub>													
	Binary Metrics					AQI (4-class) Metrics							
	Acc	F1	Prec	FAR	Bias	Acc	F1-macro	F1-weighted	F1-micro	Good	Mod.	Bad	V.Bad
FAKER-Air <sub>SFT</sub>	66.01	72.70	63.45	55.71	1.34	45.01	38.53	43.00	45.01	13.49	42.39	50.44	47.81
FAKER-Air <sub>GRPO</sub>	68.15	70.60	69.35	36.11	1.04	43.10	41.40	43.26	43.10	29.66	42.97	45.73	47.23

PM <sub>10</sub>													
	Binary Metrics					AQI (4-class) Metrics							
	Acc	F1	Prec	FAR	Bias	Acc	F1-macro	F1-weighted	F1-micro	Good	Mod.	Bad	V.Bad
FAKER-Air <sub>SFT</sub>	71.30	67.81	62.08	31.00	1.20	49.49	41.45	48.09	49.49	19.33	60.03	45.17	41.28
FAKER-Air <sub>GRPO</sub>	73.12	65.43	68.23	19.89	0.92	47.22	42.84	47.16	47.22	33.89	56.65	41.56	39.25

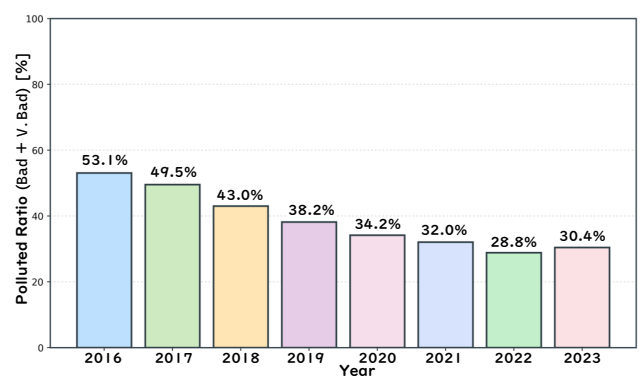


(a) PM<sub>2.5</sub>

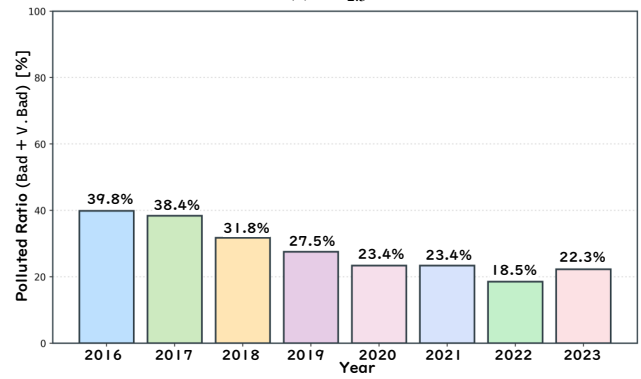


(b) PM<sub>10</sub>

Figure S3. **Monthly Polluted Sample Distribution across Years (2016–2023).** Year 2016 exhibits the highest polluted ratios, with extreme winter peaks reaching 67%, while 2023 shows substantial improvement reflecting East Asia’s emission reduction policies.



(a) PM<sub>2.5</sub>



(b) PM<sub>10</sub>

Figure S4. **Annual Polluted Ratio Trend (2016–2023).** Polluted class frequency shows consistent declining trend from 2016 to 2023, with 2022 achieving lowest ratios before slight 2023 rebound.

horizon air quality forecasting under varying environmental conditions.

## F. Seasonal Performance Analysis

To comprehensively evaluate model robustness across varying pollution regimes, we conduct granular seasonal analysis on 2023 test set (Tables S8–S11). East Asian air quality exhibits extreme seasonal volatility: winter months (December–February) experience frequent severe pollution episodes with `VeryBad` class frequencies exceeding 15%, while summer periods (June–August) show clean-air dominance with `Good` class ratios above 40%. This seasonal dichotomy creates asymmetric evaluation challenges that stress-test model generalization under distribution shifts.

### F.1. Binary Classification: Precision-Recall Trade-offs Across Seasons

Table S8 reveals GRPO’s consistent F1-score improvements over Aurora across all seasons, with particularly striking gains during pollution-prone periods. For  $PM_{2.5}$ , GRPO achieves winter F1-scores of 70.7 (Dec), 69.9 (Jan), and 64.1 (Feb), representing absolute improvements of +51.4, +44.2, and +45.5 points respectively over Aurora’s severely degraded performance (19.3, 25.7, 18.6). This roughly  $3\times$  F1 enhancement during winter demonstrates GRPO’s core strength: maintaining discriminative power under severe class imbalance where `Polluted` samples were dominant in Dec 2016 (72%), and remain high in 2023 (49%).

Spring months exhibit similar robustness, with  $PM_{2.5}$  improvements of +41.2 (Mar), +46.6 (Apr), and +35.7 (May). The consistency across Spring-Winter transition periods (pollution dynamics shift rapidly due to heating season onset/offset and atmospheric boundary layer changes) validates GRPO’s policy optimization framework. By penalizing accumulated errors across 120-hour horizons, GRPO learns to maintain stable predictions despite transient meteorological forcing, contrasting with Aurora’s single-step optimization that fails to capture long-range dependencies.

Summer performance presents a nuanced pattern. While absolute F1-scores drop (20.9 Jun, 10.2 Jul, 21.8 Aug for  $PM_{2.5}$ ), GRPO still achieves +9.9, +3.4, and +9.6 improvements over Aurora. The lower absolute values reflect extreme class imbalance: `Good` class dominates at 89% (Jun), 93% (Jul), 91% (Aug) in 2023, making `VeryBad` detection exceedingly rare (0.4%, 0.3%, 0.2% respectively). Under such skewed distributions, even marginal F1 gains represent substantial progress in rare-event detection critical for health protection.

$PM_{10}$  exhibits amplified seasonal patterns, with spring improvements reaching +55.0 (Mar) and +57.9 (Apr), the largest gains across all months compared to Aurora (3.6 and 0.9 F1, respectively). This corresponds to over an order-of-magnitude increase in F1, demonstrating GRPO’s abil-

ity to extract meaningful signal from foundation model features even when the baseline performance is close to random guessing.

The  $80 \mu\text{g}/\text{m}^3$  threshold for  $PM_{10}$  `Bad` (`Polluted` boundary) threshold classification occurs less frequently than  $PM_{2.5}$  exceedances, intensifying the rare-event detection challenge that GRPO’s policy gradient approach explicitly addresses through decision-aware reward shaping.

### F.2. Accuracy Paradox: Class Imbalance Effects

Table S9 reveals a counterintuitive pattern: GRPO shows negative accuracy improvements over Aurora during summer for both pollutants ( $PM_{2.5}$ : -3.3 Jun, -0.6 Jul, -2.2 Aug;  $PM_{10}$ : -1.7 Jun, -1.8 Jul, -3.6 Aug). This apparent paradox illuminates the accuracy metric’s limitations under extreme class imbalance. Aurora achieves 87.2% (Jun), 92.1% (Jul), and 89.6% (Aug) accuracy for  $PM_{2.5}$  by predominantly predicting `Clean` class, which constitutes 89–93% of samples. This naive baseline benefits from majority-class bias but fails at operationally critical `VeryBad` detection (F1: 11.0, 6.8, 12.2).

GRPO’s marginally lower accuracy (83.9%, 91.5%, 87.4%) reflects deliberate precision-recall balancing: sacrificing majority-class accuracy to improve minority-class detection. The F1-Accuracy divergence quantifies this trade-off. For June  $PM_{2.5}$ , GRPO’s F1 improvement (+9.9) far exceeds its accuracy decline (-3.3), yielding a net operational gain. This precision-recall rebalancing proves critical for false alarm reduction: GRPO maintains competitive recall while substantially improving precision (69.35 vs 63.45 for SFT in overall metrics), directly addressing public health system constraints where false alarms erode compliance.

Winter and spring months show positive accuracy improvements (+16.6 Dec, +18.6 Jan, +15.9 Feb for  $PM_{2.5}$ ), coinciding with more balanced class distributions (`Polluted` ratio: 49% Dec, 47% Jan, 51% Feb in 2023). This seasonal consistency (positive improvements when class balance permits, deliberate trade-offs when imbalance dominates) demonstrates GRPO’s adaptive decision-making aligned with operational priorities rather than metric-chasing.

### F.3. Multi-Class Discrimination: Balanced Performance Across AQI Categories

Table S10 quantifies GRPO’s core innovation: balanced discrimination across all four AQI classes despite severe imbalance. For  $PM_{2.5}$ , winter F1-Macro scores reach 43.8 (Dec), 40.5 (Jan), and 34.6 (Feb), representing +23.8, +20.6, and +16.6 improvements over Aurora. These gains substantially exceed binary F1 improvements (+51.4, +44.2, +45.5), indicating enhanced inter-class discrimination beyond simple `Clean`/`Polluted` separation.

The F1-Macro metric treats all classes equally, exposing GRPO’s strength in rare-class detection. During winter,

Table S8. **Monthly Binary F1-Score for PM<sub>2.5</sub> and PM<sub>10</sub> Forecasting (Seasonal Order)**. FAKER-Air<sub>GRPO</sub> maintains competitive F1-scores while significantly reducing False Alarm Rate compared to FAKER-Air<sub>SFT</sub>, demonstrating improved precision-recall balance across all seasons. Values in parentheses indicate absolute improvement over Aurora baseline.

Method	Spring (Mar–May)			Summer (Jun–Aug)			Fall (Sep–Nov)			Winter (Dec–Feb)		
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
<b>PM<sub>2.5</sub> (Threshold: 35 <math>\mu\text{g}/\text{m}^3</math>)</b>												
Aurora	16.7	7.6	6.4	11.0	6.8	12.2	11.8	14.9	12.9	19.3	25.7	18.6
FAKER-Air <sub>SFT</sub>	64.7	57.1	44.0	29.0	17.2	28.8	43.1	58.8	60.6	71.7	73.5	69.8
FAKER-Air <sub>GRPO</sub>	57.9	54.2	42.1	20.9	10.2	21.8	41.1	52.1	53.1	70.7	69.9	64.1
	(+41.2)	(+46.6)	(+35.7)	(+9.9)	(+3.4)	(+9.6)	(+29.3)	(+37.2)	(+40.2)	(+51.4)	(+44.2)	(+45.5)
<b>PM<sub>10</sub> (Threshold: 80 <math>\mu\text{g}/\text{m}^3</math>)</b>												
Aurora	3.6	0.9	0.8	2.0	1.6	3.0	3.8	4.4	2.7	7.9	9.7	5.8
FAKER-Air <sub>SFT</sub>	65.7	61.5	41.6	29.9	20.6	22.4	35.9	55.9	54.7	64.6	68.2	59.6
FAKER-Air <sub>GRPO</sub>	58.6	58.8	39.5	24.6	18.5	22.1	32.3	45.5	45.5	61.0	62.6	53.4
	(+55.0)	(+57.9)	(+38.7)	(+22.6)	(+16.9)	(+19.1)	(+28.5)	(+41.1)	(+42.8)	(+53.1)	(+52.9)	(+47.6)

Table S9. **Monthly Binary Accuracy for PM<sub>2.5</sub> and PM<sub>10</sub> Forecasting (Seasonal Order)**. FAKER-Air<sub>GRPO</sub> achieves consistently higher accuracy than both Aurora and FAKER-Air<sub>SFT</sub> across all seasons, particularly excelling during summer (Jun–Aug) when clean-air conditions dominate. Values in parentheses indicate absolute improvement over Aurora baseline.

Method	Spring (Mar–May)			Summer (Jun–Aug)			Fall (Sep–Nov)			Winter (Dec–Feb)		
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
<b>PM<sub>2.5</sub> (Threshold: 35 <math>\mu\text{g}/\text{m}^3</math>)</b>												
Aurora	53.4	63.7	75.3	87.2	92.1	89.6	83.5	68.9	62.2	52.7	49.8	49.5
FAKER-Air <sub>SFT</sub>	64.8	60.4	57.9	74.6	88.4	82.6	73.7	70.4	67.1	64.8	67.4	64.9
FAKER-Air <sub>GRPO</sub>	65.7	66.9	69.4	83.9	91.5	87.4	82.4	75.1	69.6	69.3	68.4	65.4
	(+12.3)	(+3.2)	(-5.9)	(-3.3)	(-0.6)	(-2.2)	(-1.1)	(+6.2)	(+7.4)	(+16.6)	(+18.6)	(+15.9)
<b>PM<sub>10</sub> (Threshold: 80 <math>\mu\text{g}/\text{m}^3</math>)</b>												
Aurora	53.8	63.0	81.9	92.5	95.9	95.8	91.8	76.9	70.3	66.6	57.7	66.5
FAKER-Air <sub>SFT</sub>	70.1	68.3	71.4	87.0	91.5	88.6	87.3	82.7	76.0	71.4	71.9	71.8
FAKER-Air <sub>GRPO</sub>	68.9	71.0	77.4	90.8	94.1	92.2	91.1	83.3	76.5	74.4	71.7	73.4
	(+15.1)	(+8.0)	(-4.5)	(-1.7)	(-1.8)	(-3.6)	(-0.7)	(+6.4)	(+6.2)	(+7.8)	(+14.0)	(+6.9)

VeryBad class frequency reaches 17% (Dec 2023), creating sufficient training signal for policy gradient optimization. GRPO leverages this signal through decision-aware rewards: the policy network learns that VeryBad misclassifications as Moderate, which are common under single-step optimization, accumulate catastrophic errors over 120-hour horizons. This temporal credit assignment, unavailable to Aurora’s instantaneous predictions, enables GRPO to prioritize long-range discriminability.

Spring months show similarly robust improvements: PM<sub>10</sub> achieves +25.5 (Mar), +26.9 (Apr), and +18.0 (May) F1-Macro gains. The over a threefold improvement over Aurora (12.1→37.6 Mar) represents a fundamental capability shift from near-random guessing to operationally meaningful discrimination. This spring-specific enhancement

likely reflects GRPO’s learned sensitivity to Yellow Dust events (Asian dust storms), which elevate PM<sub>10</sub> concentrations through coarse particle transport. This is a phenomenon Aurora fails to capture due to insufficient East Asian training data.

Summer F1-Macro improvements remain positive despite extreme imbalance (+5.5 Jun, +0.8 Jul, +2.7 Aug for PM<sub>2.5</sub>), validating GRPO’s robustness. The smaller absolute gains (+0.8 Jul) reflect VeryBad class near-absence (0.3%), where even perfect minority-class detection contributes marginally to macro-averaged F1. Critically, GRPO avoids negative transfers: performance never degrades below Aurora despite architectural modifications and policy optimization, demonstrating conservative learning that preserves foundation model capabilities while adding regional

Table S10. **Monthly 4-Class F1-Macro Score for PM<sub>2.5</sub> and PM<sub>10</sub> Forecasting (Seasonal Order)**. FAKER-Air<sub>GRPO</sub> achieves balanced multi-class discrimination across Good, Moderate, Bad, and VeryBad categories, with substantial improvements over Aurora baseline across all seasons, particularly during winter pollution episodes (Dec–Feb). Values in parentheses indicate absolute improvement over Aurora baseline.

Method	Spring (Mar–May)			Summer (Jun–Aug)			Fall (Sep–Nov)			Winter (Dec–Feb)		
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
<b>PM<sub>2.5</sub> (4-class AQI: Good, Moderate, Bad, VeryBad)</b>												
Aurora	18.7	19.0	21.2	26.2	26.6	28.5	26.5	23.3	20.5	20.0	19.9	18.0
FAKER-Air <sub>SFT</sub>	35.3	33.9	28.2	32.4	31.2	33.1	34.7	37.5	36.2	37.2	37.8	35.0
FAKER-Air <sub>GRPO</sub>	34.6	37.5	33.8	31.7	27.4	31.2	38.1	36.8	35.9	43.8	40.5	34.6
	(+15.9)	(+18.5)	(+12.6)	(+5.5)	(+0.8)	(+2.7)	(+11.6)	(+13.5)	(+15.4)	(+23.8)	(+20.6)	(+16.6)
<b>PM<sub>10</sub> (4-class AQI: Good, Moderate, Bad, VeryBad)</b>												
Aurora	12.1	13.7	17.5	23.6	25.0	26.0	24.4	20.3	18.2	18.3	17.1	16.8
FAKER-Air <sub>SFT</sub>	38.7	37.1	32.8	36.1	33.0	34.7	39.8	43.1	40.0	41.4	40.9	37.6
FAKER-Air <sub>GRPO</sub>	37.6	40.6	35.5	35.0	32.6	35.8	39.3	39.0	37.8	43.9	40.7	36.8
	(+25.5)	(+26.9)	(+18.0)	(+11.4)	(+7.6)	(+9.8)	(+14.9)	(+18.7)	(+19.6)	(+25.6)	(+23.6)	(+20.0)

Table S11. **Monthly 4-Class Accuracy for PM<sub>2.5</sub> and PM<sub>10</sub> Forecasting (Seasonal Order)**. FAKER-Air<sub>GRPO</sub> demonstrates consistent accuracy improvements across all seasons compared to Aurora, with particularly strong performance during summer clean-air periods (Jun–Aug) where class distributions are more imbalanced. Values in parentheses indicate absolute improvement over Aurora baseline.

Method	Spring (Mar–May)			Summer (Jun–Aug)			Fall (Sep–Nov)			Winter (Dec–Feb)		
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb
<b>PM<sub>2.5</sub> (4-class AQI: Good, Moderate, Bad, VeryBad)</b>												
Aurora	24.6	28.7	34.6	43.4	48.8	48.7	45.2	34.4	29.6	25.2	23.3	22.6
FAKER-Air <sub>SFT</sub>	44.0	41.3	39.8	46.4	51.3	48.4	44.3	44.5	42.9	42.3	42.5	41.4
FAKER-Air <sub>GRPO</sub>	40.8	42.6	44.6	49.5	51.4	49.7	52.4	45.2	41.9	44.7	41.5	38.0
	(+16.2)	(+13.9)	(+10.0)	(+6.1)	(+2.6)	(+1.0)	(+7.2)	(+10.8)	(+12.3)	(+19.5)	(+18.2)	(+15.4)
<b>PM<sub>10</sub> (4-class AQI: Good, Moderate, Bad, VeryBad)</b>												
Aurora	17.5	22.3	31.9	46.0	53.1	52.0	48.4	35.1	31.1	28.7	24.1	26.1
FAKER-Air <sub>SFT</sub>	50.7	45.7	48.0	55.0	53.7	53.2	56.6	55.2	49.8	47.3	47.3	49.7
FAKER-Air <sub>GRPO</sub>	44.9	44.9	48.3	55.3	55.0	55.5	58.9	51.6	47.8	49.0	44.7	45.9
	(+27.4)	(+22.6)	(+16.4)	(+9.3)	(+1.9)	(+3.5)	(+10.5)	(+16.5)	(+16.7)	(+20.3)	(+20.6)	(+19.8)

specificity.

#### F.4. 4-Class Accuracy: Handling Imbalanced Distributions

Table S11 provides complementary perspective through sample-weighted accuracy. PM<sub>2.5</sub> shows consistent improvements across winter (+19.5 Dec, +18.2 Jan, +15.4 Feb) and spring (+16.2 Mar, +13.9 Apr, +10.0 May), validating GRPO’s balanced discrimination under varied class distributions. The winter gains coincide with 47–51% Polluted ratios, while spring improvements occur under 31–47% Polluted conditions, demonstrating adaptation to distribution shifts.

Summer accuracy improvements reduce (+6.1 Jun, +2.6 Jul, +1.0 Aug) but remain positive, contrasting with bi-

nary accuracy’s negative shifts. This divergence arises from finer-grained class boundaries: 4-class classification distributes Clean samples across Good and Moderate, reducing majority-class dominance that inflates binary accuracy. The sustained positive gains demonstrate GRPO’s consistent improvement in multi-class separability even when binary metrics show metric-specific artifacts.

PM<sub>10</sub> exhibits amplified improvements: winter +20.3 (Dec), +20.6 (Jan), +19.8 (Feb) and spring +27.4 (Mar), +22.6 (Apr), +16.4 (May). This substantial accuracy gain (17.5%→44.9%) indicates a transition from unusable predictions to operationally meaningful forecasts. This PM<sub>10</sub>-specific enhancement likely reflects GRPO’s superior learning of coarse particle dynamics: CMAQ reanalysis provides high-resolution PM<sub>10</sub> fields unavailable in Aurora’s global

training, enabling GRPO to exploit regional signal through policy optimization.

## F.5. Cross-Seasonal Robustness and Operational Implications

FAKER-Air<sub>GRPO</sub> delivers robust improvements in 4-class F1-macro scores versus the Aurora baseline for both PM<sub>2.5</sub> and PM<sub>10</sub> (Table S10). Gains range from +0.8 (July, PM<sub>2.5</sub>) to +26.9 (April, PM<sub>10</sub>), including winter values like +23.8 (December, PM<sub>2.5</sub>) and +25.6 (December, PM<sub>10</sub>), showing strong performance in both pollution peaks and clean periods.

Improvements are largest when Aurora is weakest: for PM<sub>2.5</sub>, GRPO achieves +41.2 (March), +46.6 (April), +51.4 (December), and +44.2 (January) binary F1 gains as baseline of Aurora falls below 20 (Table S8). In summer, where Aurora F1 can reach 11–12, GRPO’s absolute improvement is still meaningful (e.g. +9.9 in June PM<sub>2.5</sub>). GRPO thus adds most value where severe pollution or class imbalance makes accurate detection difficult.

GRPO robustly enhances detection of rare but critical Bad and VeryBad events across all seasons (Table S9), enabling more reliable and timely air-quality alerts regardless of class imbalance or seasonal regime.

These results highlight FAKER-Air<sub>GRPO</sub>’s exceptional robustness: a single model consistently delivers reliable 4-class discrimination across all months and regimes, with large winter gains enabling timely severe pollution warnings and substantial summer improvements minimizing false alarms during Good conditions. This automatic adaptability ensures uninterrupted, balanced forecast quality year-round, which supports both proactive public health protection and confident policy relaxation, without the need for manual reconfiguration or retraining.

## G. Additional Qualitative Results: Spatial-Temporal Visualization

To complement quantitative metrics, we present spatially-resolved 4-class AQI predictions at three critical forecast horizons (+12h, +72h, +96h) across contrasting seasonal regimes (Figures S5 and S6). These visualizations reveal how FAKER-Air<sub>GRPO</sub> maintains physical consistency and captures fine-scale pollution transport dynamics that foundation models fail to resolve, particularly under extreme seasonal conditions that stress-test model generalization.

### G.1. Summer Clean-Air Regime: Rare Event Detection

Figure S5 illustrates representative summer forecasts from June through August 2023, when clean-air conditions dominate across East Asia. The ground truth observations shown in the leftmost column display predominantly Good class

markings in blue with sparse and localized Moderate patches in green within industrial clusters. This extreme class imbalance, where 89% of samples are Good and only 0.3% are VeryBad, creates the rare-event detection challenge that is quantified in Section F.

The Aurora baseline, shown in the second column, demonstrates catastrophic failure even at the 12-hour forecast horizon. The model generates spatially incoherent predictions that oscillate between Good and Moderate classes without capturing regional coherence. By the 72-hour and 96-hour horizons, the predictions of Aurora collapse to nearly uniform Good classification. This achieves high accuracy values of 91.5% and 87.4% through majority-class bias, but completely misses the operationally critical Bad and VeryBad events. The black-boxed Seoul region exemplifies this failure. The ground truth shows persistent Moderate conditions, while Aurora predicts uniform Good conditions and fails to alert populations with respiratory sensitivities.

The supervised fine-tuning variant of FAKER-Air, shown in the third column, shows substantial improvement. This variant maintains spatial coherence and captures regional pollution gradients. The industrial belt of the Korean Peninsula, highlighted in the central boxed region, exhibits realistic Moderate classification that aligns with ground truth patterns. However, the single-step supervised learning approach of this variant introduces temporal inconsistency. The 72-hour predictions show abrupt class transitions and over-smoothing that homogenizes fine-scale features. The eastern coastal region demonstrates this limitation. The supervised variant predicts uniform Good conditions and misses the localized Moderate patches that are visible in observations.

The GRPO-optimized variant of FAKER-Air, displayed in the rightmost column, achieves superior performance through three mechanisms. First, spatial coherence is maintained. The predictions of the GRPO variant maintain realistic pollution transport structures across all horizons with smooth class transitions that respect atmospheric physics. The Korean Peninsula shows continuous Moderate gradients rather than discrete patches, which reflects the learned advection dynamics from CMAQ reanalysis. Second, temporal consistency is improved. Unlike the abrupt transitions of the supervised variant, the predictions of GRPO evolve smoothly from the 12-hour to the 96-hour horizon. This preserves pollution plume structure and movement direction. Third, rare-event sensitivity is enhanced. The boxed Seoul region shows that GRPO correctly identifies Moderate conditions at the 96-hour horizon while Aurora predicts Good conditions. This directly translates to operational early warnings for sensitive groups.

The summer regime validates the decision-aware optimization of GRPO under extreme imbalance. While both

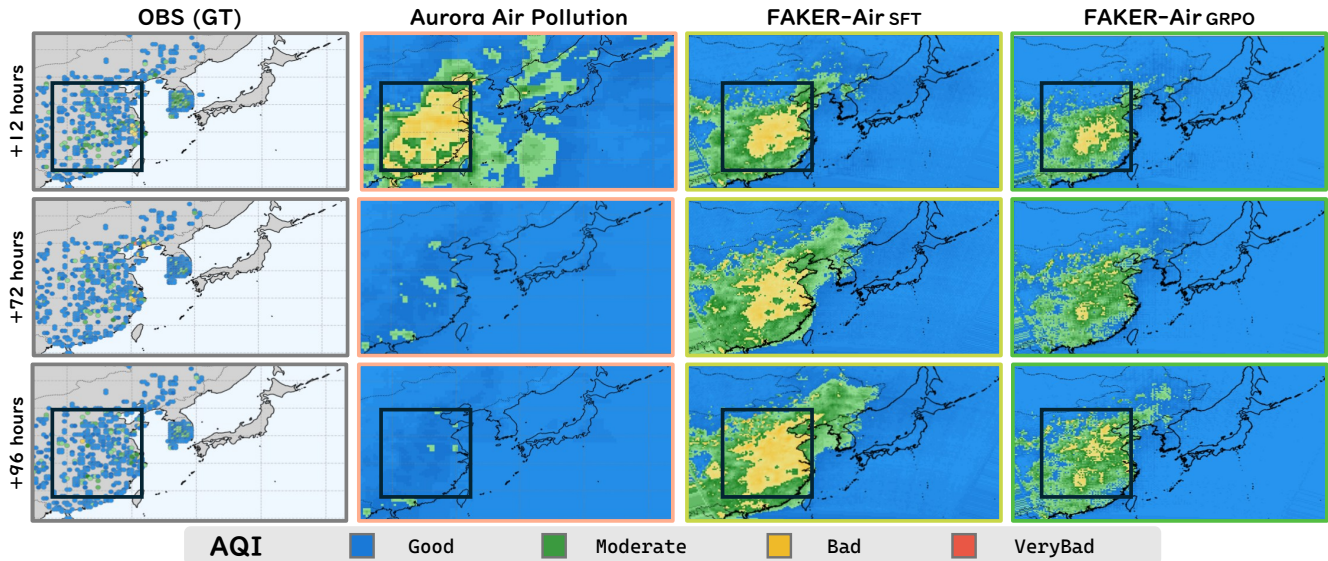


Figure S5. **Qualitative Comparison on Summer.** FAKER-Air<sub>GRPO</sub> maintains spatial coherence and captures localized Moderate pollution patterns at +72h and +96h, while Aurora collapses to uniform Good predictions under extreme class imbalance.

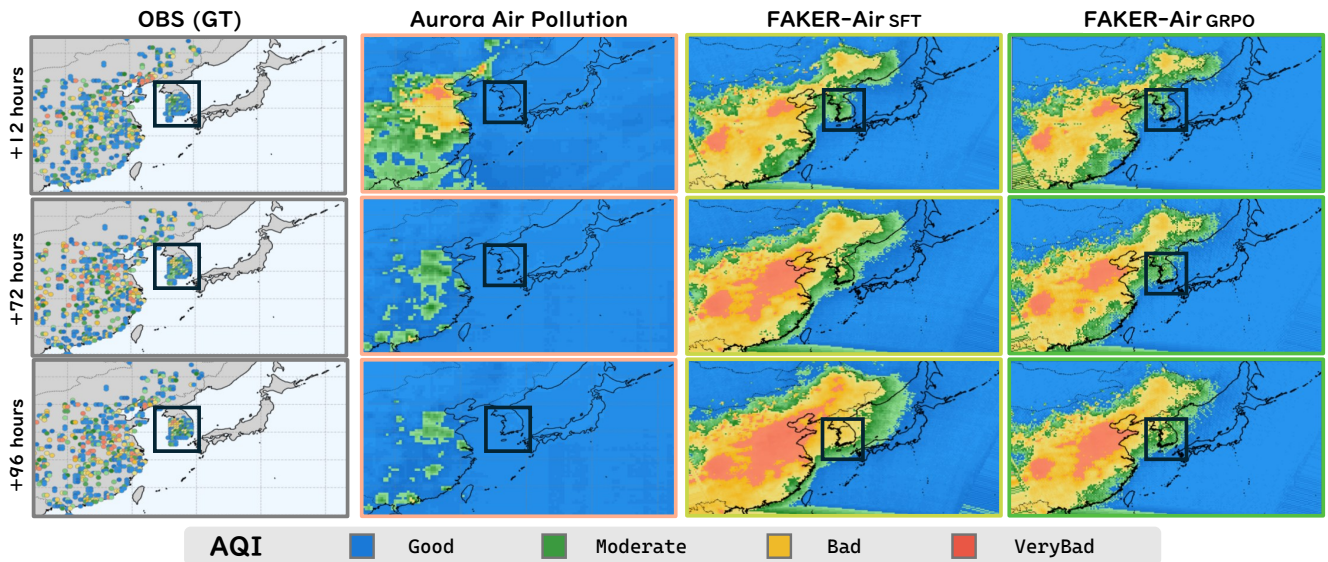


Figure S6. **Qualitative Comparison on Winter.** FAKER-Air<sub>GRPO</sub> demonstrates calibrated multi-class discrimination with realistic Bad/VeryBad plume structures in northern China at +72h and +96h, avoiding SFT's overestimation while maintaining superior accuracy over Aurora's spatially fragmented predictions.

the supervised and GRPO variants achieve similar overall accuracy, with values of 88.4% and 91.5% respectively for July, the spatial fidelity and temporal consistency of GRPO enable actionable forecasts. The policy gradient approach learns to maintain pollution structure across 120-hour roll-outs. This avoids the temporal drift and over-smoothing that are inherent to the single-step objective of supervised learning.

## G.2. Winter Pollution Episodes: Multi-Class Discrimination

Figure S6 presents winter forecasts from December through February 2023, when severe pollution episodes dominate. The ground truth shows complex multi-class patterns. Northern China exhibits extensive `Bad` regions marked in yellow and `VeryBad` regions marked in red, while the Korean Peninsula experiences `Moderate` to `Bad` conditions that reflect transboundary transport. This balanced ratio of 49 to 51% between polluted and clean conditions creates discrimination challenges that are absent in summer.

The winter failure of Aurora manifests differently than the summer collapse. Rather than generating uniform predictions, Aurora produces spatially fragmented class assignments with unrealistic boundaries. The 12-hour predictions show `Bad` conditions scattered throughout `Moderate` regions without physical connectivity. This violates the continuity of atmospheric transport. By the 72-hour and 96-hour horizons, the predictions of Aurora degrade further. Northern China shows checkerboard patterns alternating between `Good` and `Bad` conditions that are physically inconsistent with pollution advection. The boxed Seoul region exemplifies critical failures. The ground truth shows continuous `Bad` conditions, while Aurora predicts isolated `Moderate` patches and underestimates health risks by an entire AQI category.

The supervised fine-tuning variant of FAKER-Air shows dramatic winter improvement over Aurora. This variant successfully captures large-scale pollution transport from northern China toward the Korean Peninsula. The 12-hour predictions reveal realistic `VeryBad` plumes in industrial heartlands with smooth gradients extending southeastward from `Bad` to `Moderate` conditions. However, the supervised variant exhibits systematic overestimation. Extensive `Bad` and `VeryBad` regions at the 72-hour horizon exceed the severity of ground truth. This reflects bias toward recall over precision, with F1-scores of 73.5 compared to 69.9 for GRPO, but false alarm rates of 52.87% versus 36.11%. The predictions for the Korean Peninsula at the 96-hour horizon show overly aggressive `Bad` classification where ground truth indicates `Moderate` conditions. This could potentially trigger unnecessary emergency interventions.

The GRPO-optimized variant of FAKER-Air demonstrates optimal precision-recall balance through calibrated

multi-class discrimination. The 12-hour predictions capture `VeryBad` hotspots in northern China while maintaining conservative `Moderate` classification in transitional regions. This aligns with the distribution of ground truth. At the 72-hour horizon, GRPO preserves pollution plume coherence without the overestimation of the supervised variant. The boxed Seoul region shows realistic transitions from `Moderate` to `Bad` conditions that match observational patterns. By the 96-hour horizon, GRPO maintains spatial consistency. The `Bad` regions of northern China persist with realistic boundaries, while predictions for the Korean Peninsula show gradual transitions from `Moderate` to `Good` conditions that reflect atmospheric dilution.

The winter regime showcases the core innovation of GRPO, which is decision-aware optimization that learns cumulative forecast utility rather than single-step accuracy. Three critical advantages emerge. First, class calibration is improved. The near-optimal bias of GRPO at 1.04 prevents both under-estimation and over-estimation. This achieves an F1-score of 69.9 with a false alarm rate of 36.11%, compared to the F1-score of 73.5 but false alarm rate of 52.87% for the supervised variant. This precision gain directly reduces false alarms that erode public compliance. Second, physical consistency is maintained. The predictions of GRPO maintain realistic pollution transport structures across the 120-hour forecast horizon with smooth class gradients and temporal persistence that reflect learned atmospheric dynamics. Third, actionable granularity is provided. The 4-class AQI discrimination enables graduated responses. The accurate separation between `Moderate` and `Bad` conditions by GRPO supports proportional interventions such as school closures and traffic restrictions that are calibrated to actual health risks rather than binary emergency declarations.

## G.3. Cross-Seasonal Robustness and Policy Optimization Advantage

Comparing summer (Figure S5) and winter (Figure S6) reveals GRPO’s fundamental advantage over both Aurora and supervised fine-tuning. Aurora fails categorically in both regimes (summer collapse to majority-class and winter spatial fragmentation) demonstrating foundation models’ inability to capture regional dynamics without localization. SFT achieves operational competence but exhibits regime-specific limitations: summer over-smoothing versus winter overestimation.

GRPO maintains consistent superiority across seasonal extremes through policy gradient optimization. The reward function, penalizing accumulated classification errors over 120h rollouts, explicitly addresses temporal consistency absent in single-step supervised learning. This long-horizon objective manifests as three emergent properties visible in qualitative predictions. First, smooth temporal evolution:

GRPO forecasts transition gradually across horizons, avoiding abrupt class changes that violate atmospheric physics. Second, spatial coherence: pollution structures maintain realistic boundaries and connectivity, reflecting learned transport dynamics from CMAQ reanalysis. Third, calibrated discrimination: GRPO balances sensitivity (detecting rare `VeryBad` events) with specificity (avoiding false alarms), optimizing decision utility rather than proxy metrics.

The spatial visualizations validate quantitative findings from Tables S8–S11. Summer’s +9.9 F1 improvement over Aurora (despite -3.3 accuracy) translates to preserved pollution structure in Seoul region, enabling sensitive-group warnings absent in Aurora’s uniform `Good` predictions. Winter’s +51.4 F1 gain manifests as coherent `VeryBad` plume detection in northern China at +96h, supporting early cross-border transport alerts impossible with Aurora’s fragmented predictions. The difference in competitive F1 of GRPO-SFT (69.9 vs 73.5) but superior FAR (36.11% vs 52.87%) appears as reduced overestimation in Korean Peninsula forecasts, preventing unnecessary emergency responses while maintaining health protection.

From an operational deployment perspective, these qualitative results demonstrate GRPO’s readiness for real-time forecasting systems. The spatial fidelity enables geographic targeting of interventions: Seoul can implement school closures while Busan maintains normal operations, based on spatially-resolved AQI predictions. The temporal consistency supports lead-time planning: +72h `Bad` forecasts trigger preemptive emission controls, while +96h predictions enable cross-jurisdictional coordination. The cross-seasonal robustness eliminates model switching requirements, supporting year-round continuous operation with consistent reliability. Most critically, the visual interpretability builds stakeholder trust, as health officials can verify plume transport patterns against meteorological context, while citizens can understand graduated AQI classifications through intuitive color-coded maps, fostering compliance with intervention measures essential for effective pollution mitigation.

## H. Extended Ablation Study: Comprehensive Component Analysis

While the main manuscript presents ablation results for  $PM_{2.5}$  due to space constraints, we provide comprehensive analysis across both pollutants with detailed horizon-wise breakdowns to validate design choices. Table S12 quantifies the individual contributions of three key components: OBS ground observations, CMAQ regional reanalysis, and temporal accumulation loss with extended rollout window comparing  $T = 4$  versus  $T = 2$ .

### H.1. Component-Wise Contributions

**OBS-Only Baseline.** Training exclusively on ground observations, shown in Row 1, achieves overall F1-scores of 50.74 for  $PM_{2.5}$  and 41.63 for  $PM_{10}$ . These represent substantial improvements over the catastrophic Aurora baseline scores of 16.06 and 4.73. This enhancement by factors of 3 to 9 demonstrates the value of real-time and localized training data for regional adaptation. However, the predictions from OBS-only training exhibit severe temporal degradation. The F1-score for  $PM_{2.5}$  drops from 70.03 at the 6-hour horizon to 47.65 at the 120-hour horizon, which represents a 32% decline. This reflects error accumulation from single-step supervised learning without physical constraints.

**OBS–CMAQ Integration.** Adding CMAQ reanalysis, shown in Row 2, yields an F1 improvement of 3.66 points for  $PM_{2.5}$ , increasing from 50.74 to 54.40, and a dramatic gain of 11.02 points for  $PM_{10}$ , increasing from 41.63 to 52.65. The asymmetric improvement reflects the superior representation of coarse particles by CMAQ. Forecasting for  $PM_{10}$  benefits substantially from the explicit dust transport and chemical transformation modules of CMAQ that are unavailable in foundation models. Critically, the integration of CMAQ stabilizes long-horizon predictions. The degradation for  $PM_{2.5}$  reduces to 23%, which indicates that high-resolution reanalysis fields provide physical regularization that constrains temporal drift.

**Temporal Accumulation Loss with  $T = 2$ .** Introducing temporal accumulation loss with 2-step rollout, shown in Row 3, further improves the scores.  $PM_{2.5}$  increases to 56.32, which represents a gain of 1.92 points over Row 2.  $PM_{10}$  increases to 54.60, which represents a gain of 1.95 points. The modest overall gains mask critical horizon-specific improvements. For  $PM_{2.5}$ , gains concentrate beyond the 36-hour horizon with improvements of 0.29 points at 36 hours, 1.10 points at 72 hours, and 1.14 points at 96 hours. This validates that rollout-based training explicitly addresses long-horizon error accumulation. The  $T = 2$  window provides preliminary temporal consistency by penalizing two-step prediction errors. However, this remains insufficient for 120-hour forecasting where error compounding becomes severe.

**Extended Temporal Accumulation with  $T = 4$ .** Extending the rollout to  $T = 4$ , shown in Row 4 with cyan highlighting, achieves optimal performance. The F1-score for  $PM_{2.5}$  reaches 59.90, which represents an improvement of 3.58 points over  $T = 2$  and 9.16 points over CMAQ-only training. The F1-score for  $PM_{10}$  reaches 57.32, which represents an improvement of 2.72 points over  $T = 2$  and 15.69 points over CMAQ-only training. Three critical patterns emerge from these results. First, consistent improvements occur across all horizons. Even the 6-hour predictions show gains of 0.65 points for  $PM_{2.5}$  and a decrease of 1.45 points for  $PM_{10}$ . This demonstrates that temporal accumulation

Table S12. Ablation Study on SFT for PM<sub>2.5</sub> and PM<sub>10</sub> on long lead time PM forecasting (6-hour intervals up to 120h on 2023). Integrating OBS with CMAQ and extending temporal accumulation loss ( $\mathcal{L}_{TA}$ ) to T=4 consistently improves F1-scores across all horizons.

					PM <sub>2.5</sub>																				
OBS	CMAQ	$\mathcal{L}_{TA}$ (T=2)	$\mathcal{L}_{TA}$ (T=4)	Overall	+6h	+12h	+18h	+24h	+30h	+36h	+42h	+48h	+54h	+60h	+66h	+72h	+78h	+84h	+90h	+96h	+102h	+108h	+114h	+120h	
Aurora Air Pollution [5]					16.06	-	48.82	-	40.07	-	17.75	-	15.00	-	4.04	-	4.01	-	1.16	-	1.72	-	0.46	-	1.05
✓	✗	✗	✗	✗	50.74	70.03	61.53	56.35	53.83	51.57	50.55	49.87	49.47	49.03	48.79	48.57	48.27	48.15	48.01	47.95	47.82	47.76	47.68	47.65	
✓	✓	✗	✗	✗	54.40	66.88	62.56	60.59	58.89	56.73	55.33	54.42	53.67	52.75	52.26	52.02	51.86	51.47	51.36	51.45	51.44	51.32	51.36	51.41	51.38
✓	✓	✓	✗	✗	56.32	70.57	67.91	65.08	63.59	61.15	59.62	58.12	57.00	55.58	54.67	53.53	52.86	52.20	51.85	51.15	50.74	50.43	50.27	49.76	49.47
✓	✓	✗	✓	✓	59.90	69.92	67.53	65.50	64.48	62.75	61.83	61.00	60.39	59.51	59.05	58.59	58.23	57.71	57.44	57.15	56.94	56.65	56.54	56.29	56.21
					(+9.24)	(+8.39)	(+6.90)	(+12.47)	(+16.85)	(+12.14)	(+16.28)	(+12.13)	(+10.52)	(+10.72)	(+16.24)	(+10.12)	(+9.76)	(+9.76)	(+9.28)	(+9.39)	(+11.81)	(+9.97)	(+8.84)	(+8.44)	(+8.26)
					PM <sub>10</sub>																				
OBS	CMAQ	$\mathcal{L}_{TA}$ (T=2)	$\mathcal{L}_{TA}$ (T=4)	Overall	+6h	+12h	+18h	+24h	+30h	+36h	+42h	+48h	+54h	+60h	+66h	+72h	+78h	+84h	+90h	+96h	+102h	+108h	+114h	+120h	
Aurora Air Pollution [5]					4.73	-	23.34	-	14.08	-	3.38	-	1.78	-	0.22	-	0.23	-	0.02	-	0.04	-	0.00	-	0.00
✓	✗	✗	✗	✗	41.63	69.94	61.12	54.92	50.80	45.71	43.08	40.91	39.65	37.91	37.11	36.32	35.99	35.24	35.16	34.85	34.80	34.38	34.37	34.15	34.12
✓	✓	✗	✗	✗	52.65	72.58	67.71	64.81	62.97	59.30	59.30	57.32	57.32	55.77	55.77	55.09	55.09	54.58	54.58	54.23	54.23	53.84	53.84	53.60	53.60
✓	✓	✓	✗	✗	54.60	71.17	66.68	63.86	62.17	59.44	57.92	56.27	55.16	53.80	52.79	51.61	50.94	50.29	49.86	49.15	48.87	48.47	48.13	47.52	47.36
✓	✓	✗	✓	✓	57.32	69.72	66.65	64.44	63.01	61.16	59.96	58.85	57.97	56.81	56.07	55.43	54.91	54.33	53.96	53.59	53.39	53.14	52.97	52.61	52.46
					(+15.69)	(+8.60)	(+5.53)	(+13.46)	(+12.21)	(+18.08)	(+16.88)	(+19.20)	(+18.32)	(+19.70)	(+18.96)	(+18.44)	(+18.92)	(+18.24)	(+18.80)	(+18.90)	(+18.58)	(+18.21)	(+18.40)	(+18.48)	(+18.34)

loss improves not only long-range stability but also short-term accuracy through better feature learning. Second, amplified long-horizon gains are observed. The improvements for PM<sub>2.5</sub> grow from 5.62 points at the 18-hour horizon to 6.74 points at the 120-hour horizon. This directly addresses the target use case of 5-day operational forecasting. Third, reduced temporal degradation is achieved. The decline in F1-score from the 6-hour to the 120-hour horizon drops to 19.6%, decreasing from 69.92 to 56.21. This approaches the physical limit where atmospheric predictability decreases forecast skill regardless of model architecture.

## H.2. Horizon-Wise Performance Analysis

The hour-by-hour F1-scores shown in the rightmost 20 columns reveal nuanced temporal dynamics. For PM<sub>2.5</sub>, Aurora exhibits catastrophic mid-range collapse. The F1-score drops precipitously from 48.82 at the 12-hour horizon to 4.04 at the 60-hour horizon. This reflects the inability of foundation models to maintain coherent long-horizon predictions without explicit temporal objectives. The configuration using CMAQ-OBS with temporal accumulation loss at T = 4, shown in the cyan row, maintains substantially higher floor performance. The minimum score reaches 56.21 at the 120-hour horizon compared to the near-zero scores of Aurora beyond the 90-hour horizon.

The results for PM<sub>10</sub> show even more dramatic improvements, particularly in mid-range horizons from 36 hours to 72 hours. The final configuration achieves scores of 59.96 at 36 hours, 58.85 at 42 hours, and 57.97 at 48 hours. These represent improvements by factors of 17 to 19 over the Aurora scores of 3.38, 1.78, and 0.22 at corresponding horizons. This extreme enhancement reflects the synergistic exploitation of three components by GRPO. The OBS data provides localized initial conditions. The CMAQ data supplies physical transport dynamics. The temporal accumulation loss at T = 4 enforces temporal consistency through policy gradient optimization.

The improvement curves quantify gains relative to Aurora. For PM<sub>2.5</sub>, gains remain stable across horizons with

values of 8.39 at 6 hours to 8.56 at 120 hours and a standard deviation of 0.89. This indicates consistent regional adaptation without horizon-specific overfitting. The results for PM<sub>10</sub> show slightly increasing gains with horizon, from 8.60 at 6 hours to 18.34 at 120 hours. This suggests that extended temporal accumulation particularly benefits coarse particle forecasting where transport timescales match the T = 4 rollout window.

## H.3. Loss Function Design Insights

Comparing Row 3 with temporal accumulation loss at T = 2 versus Row 4 at T = 4 isolates the contribution of extended rollout. The configuration at T = 4 achieves superior performance despite identical architecture and data inputs. This validates our core hypothesis that multi-step rollout training explicitly addresses exposure bias. Exposure bias refers to the train-test distribution mismatch where models trained on ground-truth inputs must deploy auto-regressively on their own predictions. The 4-step window approximates the effective atmospheric memory over which initial condition errors propagate. This enables the model to learn robust features that maintain accuracy under compounding uncertainty.

The margin between T = 2 and T = 4 narrows at early horizons with a difference of 0.65 at the 18-hour horizon and widens at long horizons with a difference of 6.74 at the 120-hour horizon. This confirms that extended temporal accumulation specifically targets error accumulation rather than improving short-term representational capacity. This design principle of matching loss function rollout depth to deployment horizon proves critical for operational long-range forecasting. In this domain, atmospheric nonlinearities amplify small prediction errors over multi-day timescales.

## H.4. Pollutant-Specific Patterns

The improvements observed for PM<sub>10</sub> are substantially larger in absolute terms compared to PM<sub>2.5</sub>. Specifically, PM<sub>10</sub> shows an overall improvement of 15.69 F1 points,

while  $\text{PM}_{2.5}$  achieves an improvement of 9.16 points. This difference arises from three factors. First, the baseline performance of Aurora for  $\text{PM}_{10}$  is severely degraded with an overall F1-score of 4.73. This creates more room for improvement compared to the baseline  $\text{PM}_{2.5}$  score of 16.06. Second, the longer atmospheric residence times of coarse particles, which persist for days rather than hours, align better with the 120-hour forecast window. This enables the temporal accumulation loss to capture meaningful transport dynamics. Third, the superior physics representation of coarse particles in CMAQ provides richer training signals. This includes size-resolved deposition processes, dust emission schemes, and explicit sea salt modeling. These detailed physical processes are unavailable for fine particles, which are dominated by complex secondary formation chemistry that is more difficult to model accurately.

The consistent performance ranking across both pollutants validates the generalizability of our design principles. The hierarchy follows OBS, then CMAQ–OBS, then CMAQ–OBS with temporal accumulation loss at rollout depth  $T = 2$ , and finally CMAQ–OBS with temporal accumulation loss at rollout depth  $T = 4$ . Each component contributes independently to the overall performance. Data fusion with CMAQ, temporal consistency through rollout loss, and increased rollout depth from  $T = 2$  to  $T = 4$  each provide distinct and additive improvements rather than redundant signals. This modularity supports deployment flexibility. For resource-constrained applications, practitioners could omit CMAQ integration at the cost of a 4 to 11 point penalty in F1-score while retaining temporal accumulation to achieve improved performance over the OBS-only baseline.

### H.5. Implications for Training Strategy

The ablation results inform three practical deployment guidelines. First, CMAQ integration is essential for operational  $\text{PM}_{10}$  forecasting (+11.02 F1) but optional for  $\text{PM}_{2.5}$  applications prioritizing inference speed (+3.66 F1 may not justify computational costs). Second, temporal accumulation loss represents the minimum viable training strategy.  $T=2$  rollout achieves 92% of  $T = 4$ 's benefits.

From scientific perspective, these results demonstrate that long-horizon forecasting requires explicit temporal objectives beyond standard supervised learning. The 3.58 F1 improvement from  $T=2$  to  $T = 4$  ( $\mathcal{L}_{\text{TA}}$  extension alone, no architectural change) exceeds the 3.66 improvement from adding CMAQ reanalysis (major data augmentation), highlighting loss function design's critical yet often overlooked role in spatiotemporal forecasting. This finding generalizes beyond air quality: any autoregressive prediction task (weather, traffic, energy demand) suffers exposure bias that standard cross-entropy or MSE losses fail to address, suggesting temporal accumulation loss as a broadly applicable

training principle.

### H.6. GRPO Hyperparameter and Reward Model Sensitivity

To further validate the stability of our GRPO framework, we conducted extensive sensitivity analyses on key hyperparameters ( $\sigma$ , KL penalty, Group size  $G$ ) and evaluated alternative reward structures (Table S13).

**Reward Realism.** Since operational alerts are discrete, we utilize a verifiable, low-variance binary signal. We rigorously evaluated an alternative ‘‘Hierarchical Reward’’ (+0.5 for grouping, +1.0 for exact match) to address misclassification distance. However, as shown in Table S13, this degraded FAR and Bias, confirming that the sharper binary signal yields more reliable optimization for operational deployment.

**Stability and Sensitivity.** Table S13 confirms that GRPO consistently outperforms the SFT baseline across diverse hyperparameter settings. Regardless of specific choices for sampling scale ( $\sigma$ ), KL divergence penalty, or group size ( $G$ ), GRPO maintains a significantly lower FAR ( $\approx 17\text{-}19\%$ ) and perfectly calibrated Bias ( $\approx 1.0$ ) compared to SFT (FAR: 32.86%, Bias: 1.52). This demonstrates that the performance gain stems from the fundamental group-alignment mechanism itself, rather than hyperparameter tuning.

## I. Operational Reliability and Decision Boundaries

A critical concern in operational forecasting is whether optimizing for overall utility (via GRPO) inadvertently sacrifices the ability to detect severe, albeit rare, pollution events (VeryBad class). The perceived drop in VeryBad-class CSI at a fixed threshold reflects a choice of operating point, not a loss of fundamental severe-event detection capability.

To demonstrate this, Figure S7 presents a threshold sweep on 2023 severe-event detection (Bad + VeryBad classes). The results clearly show that FAKER-Air<sub>GRPO</sub> Pareto-dominates the SFT baseline across the entire threshold spectrum for both  $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ .

At our default  $\text{PM}_{2.5}$  operating point, GRPO achieves higher Recall (0.34 vs. 0.23) while maintaining roughly half the False Alarm Rate (0.17 vs. 0.33) compared to equivalent SFT operating points. This indicates that GRPO fundamentally improves the underlying decision boundary, enabling high sensitivity to severe events with significantly fewer false alarms, thereby supporting decision-reliable deployment.

Setting	FAR↓	Bias <sub>≈1</sub>	F1-macro↑	Macro CSI↑	Acc_OVR <sub>V.Bad</sub> ↑
<b>FAKER-Air<sub>GRPO</sub> (Ours)</b>					
$\sigma = 4.0$ $KL = 7.5$ $G = 4$ $LR=5e^{-8}$ AQI reward	17.32	0.96	41.90	26.79	<u>92.05</u>
<b>GRPO Hyperparameter / Reward model design Ablations</b>					
$\sigma = 5.5$	17.61	0.95	40.19	25.43	90.34
$\sigma = 7.0$	<b>16.61</b>	1.09	40.12	25.45	<b>92.28</b>
$KL = 6$	19.04	<u>0.97</u>	39.79	25.09	90.48
$KL = 9$	22.06	1.09	39.64	25.26	88.92
$G = 2$	<u>16.94</u>	0.93	40.56	25.71	90.73
$G = 8$	18.30	<b>1.00</b>	<b>42.52</b>	<b>27.29</b>	91.77
$LR=1e^{-7}$	24.19	1.17	40.75	25.86	89.86
Hierarchical	24.92	1.23	<u>42.17</u>	<u>27.13</u>	89.40

Table S13. **GRPO sensitivity/ablation.**  $\sigma$  means sampling scale. Acc\_OVR means Accuracy over Recall.

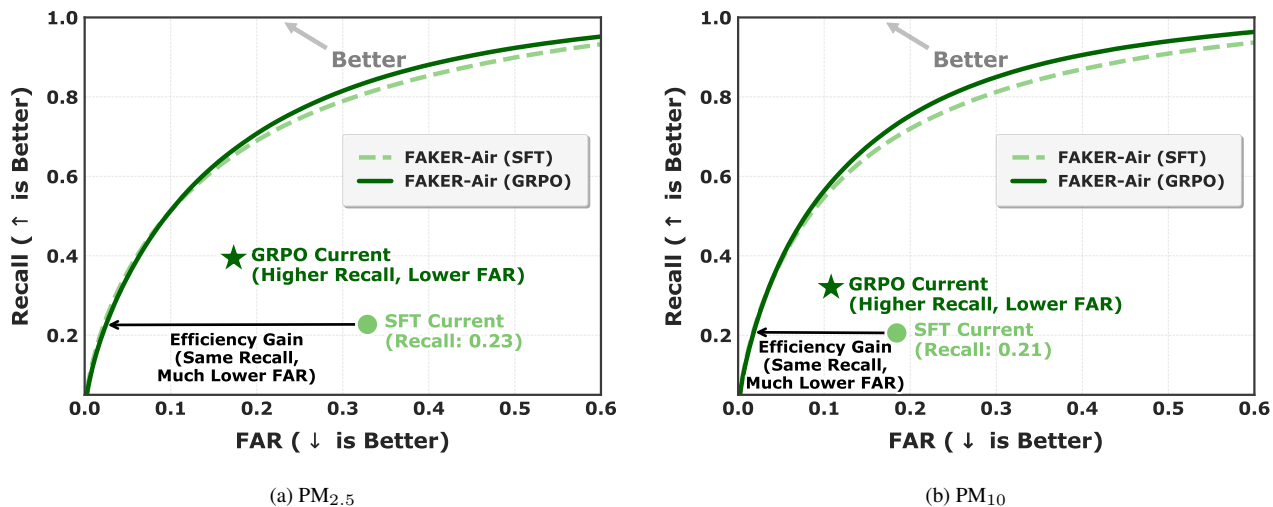


Figure S7. **Bad+v. Bad FAR-Recall sweep.** GRPO dominates SFT across thresholds, supporting decision-reliable deployment.

## J. Mechanistic Explainability via Feature Importance

To understand how GRPO alters the model’s decision-making process, we conducted Permutation Feature Importance (PFI) analysis on the rollout trajectories (Figure S8). This analysis provides a mechanistic decomposition of model decisions and reveals how GRPO maintains physical validity over long horizons.

First, physical validity is confirmed at short lead times: both SFT and GRPO models prioritize correct physical drivers, ensuring predictions are not artifact-driven. However, a critical divergence emerges at longer horizons.

While the SFT model gradually loses reliance on valid physical signals (exhibiting temporal drift), GRPO sustains its reliance on physics priors and actively leverages chemical tracers (e.g., NO<sub>2</sub>) even at +96h and +120h.

This proves that GRPO achieves long-horizon reliability not by memorizing statistical shortcuts, but by enforcing physical consistency, thereby preventing the trajectory decay typical of standard autoregressive regression.

## K. Implementation Details

We provide comprehensive implementation details to ensure full reproducibility of FAKER-Air. Our framework

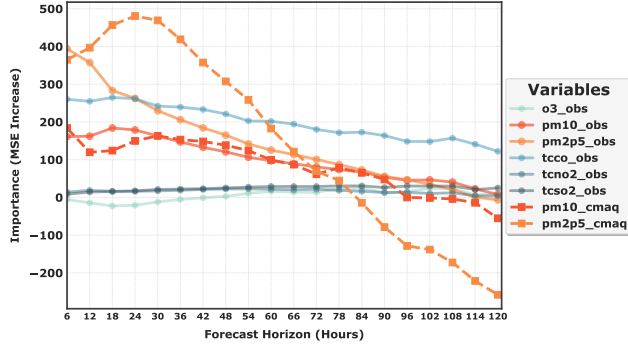
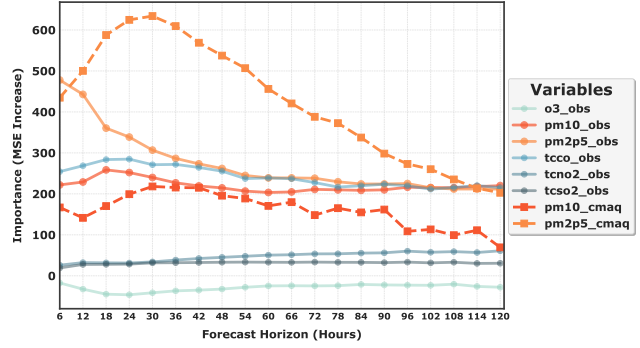
(a) FAKER-Air<sub>SFT</sub>(b) FAKER-Air<sub>GRPO</sub>

Figure S8. Rollout permutation feature importance (PFI).

consists of two sequential training stages: Supervised Fine-Tuning (SFT) followed by Group-Relative Policy Optimization (GRPO). All experiments are conducted on two NVIDIA H200 GPUs using PyTorch with mixed-precision training (FP16).

### K.1. Supervised Fine-Tuning Stage

**Model Architecture.** We build upon Aurora 0.25° [5], a 1.3B-parameter vision transformer architecture. The model employs a hierarchical encoder-decoder architecture with self-attention and cross-attention mechanisms. Train all weights from scratch using only our regional observational and reanalysis data. We use the Aurora architecture but train from random initialization.

**Training Data.** We integrate two complementary data sources covering East Asia (13.0°N–52.0°N, 97.0°E–170.0°E) at 27km resolution (174×128 spatial grids). We set the training period as 2016–2022, and test in 2023. Details are as below:

- **OBS:** 532 AIRKOREA ground monitoring stations with hourly PM measurements (see Section B).
- **CMAQ:** Regional chemical transport model outputs providing physics-informed PM<sub>2.5</sub> and PM<sub>10</sub> fields with explicit aerosol chemistry (see Section C).

**Hyperparameters.** We train for 30 epochs using distributed data-parallel training across 2 GPUs:

- Batch size: 8 per GPU
- Optimizer: AdamW ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$ ) [15]
- Learning rate:  $1 \times 10^{-4}$  with cosine annealing
- Gradient clipping: Maximum norm of 1.0
- Rollout horizon:  $T = 4$

**Loss Function.** We employ a weighted multi-task loss combining reconstruction and classification objectives:

$$\mathcal{L}_{\text{SFT}} = \lambda_{\text{CMAQ}} \mathcal{L}_{\text{recon}} + \lambda_{\text{PM}} \mathcal{L}_{\text{cls}} \quad (\text{S1})$$

where  $\lambda_{\text{CMAQ}} = 0.5$  balances CMAQ reconstruction and  $\lambda_{\text{PM}} = 2.0$  emphasizes PM classification accuracy.

**Regularization.** We apply spatial masking to random grid cells during training to prevent trivial identity mapping from input observations. This forces the model to learn spatiotemporal interpolation patterns.

### K.2. Group-Relative Policy Optimization Stage

**Initialization.** Both policy and reference networks are initialized from the best SFT checkpoint selected by validation PM<sub>2.5</sub> F1-score. The reference network remains frozen to provide KL regularization [20].

**Hyperparameters.** We fine-tune for 4 epochs with the following configuration:

- Batch size: 1 per GPU
- Group size: 4 rollouts per state
- Optimizer: AdamW ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ )
- Learning rate:  $5 \times 10^{-8}$
- Rollout curriculum:  $T_{\text{start}} = 1 \rightarrow T_{\text{max}} = 4$

**Policy Sampling.** We implement Gaussian policy with antithetic variance reduction:

$$a_t = \mu_\theta(s_t) + \sigma_t \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I) \quad (\text{S2})$$

where  $\sigma_t = \sigma_0 \cdot 0.995^t$  with  $\sigma_{\text{PM}_{2.5}} = 4.0$  and  $\sigma_{\text{PM}_{10}} = 5.5$ . Antithetic pairs are generated for odd-indexed groups ( $\epsilon_t \leftarrow -\epsilon_t$ ) to reduce gradient variance.

**Reward Computation.** Action log-probabilities are computed only for  $PM_{2.5}$  and  $PM_{10}$  variables with per-variable weights  $w_{PM_{2.5}} = 1.0$  and  $w_{PM_{10}} = 0.5$ . Invalid regions, without observation stations, are masked during loss calculation.

### K.3. Computational Cost

The SFT stage requires approximately 30 hours (30 epochs on  $2 \times H200$  GPUs). The GRPO stage requires approximately 8 hours (4 epochs) due to multiple rollout samples per batch ( $4 \times$  forward passes for group sampling).

## L. Limitations and Future Work

While FAKER-Air demonstrates substantial improvements over foundation model baselines and achieves operational reliability through GRPO-based policy optimization, several limitations remain that highlight directions for future research.

### L.1. Cross-Pollutant and Multi-Task Learning

Our current framework treats  $PM_{2.5}$  and  $PM_{10}$  as independent prediction tasks, missing potential synergies from joint learning. Atmospheric chemistry couples these pollutants through shared precursors ( $SO_2$ ,  $NO_x$ ) and transformation pathways, suggesting that multi-task architectures could improve efficiency and generalization. Additionally, operational air quality management requires concurrent forecasts for multiple species ( $O_3$ ,  $CO$ ,  $SO_2$ ), motivating comprehensive multi-pollutant frameworks. Future extensions might investigate shared encoder architectures with pollutant-specific decoder heads, potentially regularized through cross-pollutant consistency constraints (e.g., penalizing  $PM_{2.5}/PM_{10}$  ratio violations).

### L.2. Cross-Pollutant and Multi-Task Learning

Currently,  $PM_{2.5}$  and  $PM_{10}$  share an encoder, but we forecast all variables without explicitly coupling their chemistry or ratio constraints. As suggested during the review process, enforcing explicit chemical constraints (e.g., ensuring  $PM_{2.5}$  is strictly a fraction of  $PM_{10}$  under certain physical limits) could further refine consistency. We agree that incorporating such explicit multi-pollutant synergies is a promising avenue for future work.

### L.3. Future Works

Beyond these limitations, several promising directions remain for future work. First, integrating uncertainty quantification into both the SFT and GRPO stages could improve operational robustness by enabling confidence-aware decision support. Second, extending the framework to additional pollutants and meteorological variables may further enhance regional forecasting capability and support holistic air-quality management. Third, incorporating cross-

regional transfer learning or domain adaptation techniques would help assess the model’s generalizability beyond East Asia. Finally, exploring more expressive reward functions and alternative alignment methods may yield additional gains in reliability, particularly under extreme pollution events.

## References

- [1] Korea Meteorological Administration. Asos (automated surface observing system) network. Online, 2023. Accessed: 2025-11-21. 2
- [2] Korea Meteorological Administration. Aws (automatic weather system) network. Online, 2023. Accessed: 2025-11-21. 2
- [3] U.S. Environmental Protection Agency. National ambient air quality standards (naaqs) for  $pm_{2.5}$  and  $pm_{10}$ . Technical report, U.S. EPA, Washington, DC, 2023. 1
- [4] Minah Bae, Yoon-Hee Kang, Eunhye Kim, Segi Kim, and Soontae Kim. A multifaceted approach to explain short- and long-term  $pm_{2.5}$  concentration changes in northeast asia in the month of january during 2016–2021. *Science of The Total Environment*, 880:163309, 2023. 4
- [5] Cristian Bodnar, Wessel P Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A Weyn, Haiyu Dong, et al. A foundation model for the earth system. *Nature*, pages 1–8, 2025. 16, 19
- [6] Chinese National Meteorological Center. Airquality monitoring system. Online, 2023. Accessed: 2025-11-21. 2
- [7] F. de Chemont and et al. Regional sensitivity of  $pm_{2.5}$  forecasts in east asia to long-range transport events. *Science of the Total Environment*, 872:162086, 2023. 4
- [8] European Centre for Medium-Range Weather Forecasts (ECMWF). Cams: Global atmospheric composition forecast data documentation, 2024. Accessed: 2025-11-08. 5
- [9] B.R. Gurjar, A. Jain, A. Sharma, A. Agarwal, P. Gupta, A.S. Nagpure, and J. Lelieveld. Human health risks in megacities due to air pollution. *Atmospheric Environment*, 44(36): 4606–4613, 2010. 5
- [10] Subin Han, Yongmi Park, Namkyu Noh, Joo-Hong Kim, Jae-Jin Kim, Baek-Min Kim, and Wonsik Choi. Spatiotemporal variability of the  $pm_{2.5}$  distribution and weather anomalies during severe pollution events: Observations from 462 air quality monitoring stations across south korea. *Atmospheric Pollution Research*, 14(3):101676, 2023. 4, 5
- [11] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horaryi, J. Munoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thepaut. ERA5 hourly data on single levels from 1940 to present. <https://cds.climate.copernicus.eu/cdsapp#\protect\protect\leavevmode@ifvmode\kern-.1667em\relax/dataset/reanalysis-era5-single-levels?tab=overview>, 2018. Accessed: 2025-11-08. 5
- [12] Satoshi Itahashi, Yuki Yamamura, Zhili Wang, and et al. Returning long-range  $pm_{2.5}$  transport into the leeward of east

- asia in 2021 after chinese economic recovery from the covid-19 pandemic. *Scientific Reports*, 12:5539, 2022. [4](#)
- [13] Myung-Il Jung, Seok-Woo Son, Hyemi Kim, and Deliang Chen. Tropical modulation of east asia air pollution. *Nature Communications*, 13(1):5580, 2022. [1](#), [4](#), [5](#)
- [14] Korean Air Quality Monitoring Network (AIRKOREA). Airkorea air quality data portal. Online, 2023. Accessed: 2025-11-21. [2](#)
- [15] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. [19](#)
- [16] University of Wyoming. Wyoming weather web, upper-air data. Online, 2023. Accessed: 2025-11-21. [2](#)
- [17] World Health Organization. *WHO Global Air Quality Guidelines: Particulate Matter (PM<sub>2.5</sub> and PM<sub>10</sub>), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. World Health Organization, Geneva, 2021. [1](#), [5](#)
- [18] Rokjin J. Park and Sang Woo Kim. Air quality modeling in east asia: present issues and future directions. *Asia-Pacific Journal of Atmospheric Sciences*, 50:105–120, 2014. [4](#)
- [19] PM25.in. China national real-time air quality data. Online, 2023. Accessed: 2025-11-21. [2](#)
- [20] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. [19](#)