Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Data integrity auditing of model-derived $PM_{2.5}$ series and reanalysis-referenced evaluation of lagged dependence networks in Luzon, Philippines

Jerald B. Bongalos

jbongalos.phdinds2028@aim.edu

Asian Institute of Management

Christopher P. Monterola

Asian Institute of Management

Article

Keywords: PM2.5, data integrity, lagged networks, Granger causality, monsoon, Philippines

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11617673/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

City-labelled air-quality series from global models are increasingly used to infer how fine particulate matter ($PM_{2.5}$) moves between cities. We show that such panels need integrity checks and test what lagged networks from them support. In a 40-city 2025 panel for Luzon, Philippines (OpenWeatherMap via Kaggle), the publisher's query coordinates placed ten cities 72–356 km from their stated locations, and 15 others formed five groups with identical daily means on 344–360 of 365 days (95–100% of 2026 hours), leaving 20 distinct series. Uncorrected, they produced spurious long-range links and, in our earlier draft, an 8.9% prediction gain. After correction, long-shift surrogates yield few edges; short-shift surrogates, which largely preserve seasonal structure and may preserve shared episodes, yield a third to half the observed counts. In exploratory ERA5 wind tests, we detected downwind alignment under both monsoons for edges found without co-pollutant conditioning, not for edges added by it. Sender–receiver roles were unstable; network neighbors improved forecasts and warnings little more than nearest locations. Against reference stations, the product tracked daily variation poorly (median r = 0.15 at 13 stations) and read about half the measured level; CAMS had lower error at 10 of them.

Introduction

Fine particulate matter ($PM_{2.5}$) moves between cities on timescales of hours to days, carried by regional winds that change with season[1,2]. Knowing which places send pollution to which others matters for source attribution, exposure assessment and the design of regional controls[3]. Chemical transport models simulate this movement directly, but they need emission inventories, meteorological fields and chemical mechanisms that are uncertain or unavailable in many regions[4–6]. Statistical and machine-learning approaches instead learn spatial structure from concentration data[7,8]. Lagged-dependence methods go a step further, asking whether one location's past improves prediction of another's beyond its own history[9–13]. Such predictive links are not transport by themselves: they separate transport from shared drivers only to the extent that those drivers are measured and conditioned on[14,15].

Two developments make such networks newly attractive and newly risky. First, model-based air-quality services report co-pollutants alongside $PM_{2.5}$. Nitrogen dioxide and carbon monoxide track combustion, sulfur dioxide tracks industrial sources and ozone tracks photochemistry[16–20]. Conditioning a receptor's $PM_{2.5}$ on its own co-pollutants has been proposed as a way to absorb shared emission and chemistry signals. These co-pollutants are not meteorological controls, however, and they respond to the same weather that moves $PM_{2.5}$. Second, global model-based services now publish series for thousands of named cities at no cost[21], and curated copies circulate widely[22]. The provider used here describes its users as enterprises, public bodies, research institutions and developers[23]; its pollutant fields are named in several hundred indexed works, most of them forecasting and monitoring applications[24]; and the Philippine panel analyzed here had been downloaded over 1,200 times and feeds public $PM_{2.5}$ forecasting notebooks[22]. We found no published comparison of the product with reference monitors. These series make network analysis possible where ground monitors are sparse, which is where chemical transport models are hardest to run.

City-labelled data carry two problems that are invisible from names alone. A provider resolves each city name to coordinates and returns the series at those coordinates, so a common place name can return the series of a different place with the same name. Neighbouring cities can also receive identical series, presumably because they fall in one cell of the provider's underlying model, although providers rarely document that grid. Both problems change the number of distinct signals in a panel, and both can manufacture network structure: a mislocated series looks like long-range influence, and a duplicated one looks like perfect local coupling. Neither is addressed by better network inference.

Here we diagnose and correct both problems in a 40-city panel for Luzon, Philippines, then build lagged $PM_{2.5}$ networks on the corrected panel. We treat each directed edge as a statistical lagged association supported by conditional and unadjusted tests, and ask which links are consistent with an independent record of the circulation. Luzon's northeast (Amihan) and southwest (Habagat) monsoons reverse the prevailing low-level flow between seasons[25], so edges that reflect transport should point downwind in both. We make five contributions:

1. Two integrity diagnostics, one from the provider's own query coordinates and one from exact value matches, reduce 40 named cities to 20 distinct series. Uncorrected, the panel produces spurious long-range links and a spurious predictive advantage.
2. We calibrate the whole discovery pipeline against surrogate nulls that keep each series' time dependence, separate the effect of co-pollutant conditioning from that of adding a third test, and report how many edges depend on the treatment of serial dependence.
3. We test edge direction against ERA5 reanalysis winds with inference that respects the network structure. In these exploratory tests, we detect downwind alignment under both monsoons for edges found without co-pollutant conditioning but not for edges added by it; compared directly, the two sets differ detectably only in December–February, and with a matched test rule we detect no effect of conditioning.
4. We show the limits of the annual network: sender–receiver roles and partitions are not stable; in forward-time forecasts network neighbors add little beyond each location's own history and its nearest neighbors; and neighbors improve next-day warning of high-$PM_{2.5}$ onsets almost as much when chosen by proximity as by the network.
5. We evaluate the product against reference stations: in Metro Manila it tracks daily variation poorly and reads about half the measured level, its seasonal agreement cannot be determined from the available stations, and a second global model has lower error at most of the same stations, though not uniformly better agreement, so the networks describe the product's field rather than measured exposure.

Results

Source metadata and value checks reduce 40 named cities to 20 distinct series

The panel holds daily means of eight pollutants for 40 Luzon cities over the 365 days of 2025, computed from hourly OpenWeatherMap Air Pollution values[21] as distributed on Kaggle (version 429)[22]; no city-day is missing (Methods). The publisher's city table records, for every city, the coordinates returned with its query. For 30

cities these lie within 5 km of the stated location. For ten they lie 72–356 km away (Table 1, Fig. 1a). Nine of the ten have names shared with other Philippine places; their coordinates had been corrected in our panel, but their series were still those of the other places. The tenth, San Fernando, was queried at San Fernando, La Union (16.62°N), not San Fernando, Pampanga. These offsets come from the source's metadata, not from the concentration data.

Table 1

Cities queried at another location. Query point: coordinates in the publisher's city table. r, mean Pearson correlation of daily $PM_{2.5}$ (2025) with the three panel cities nearest the stated location and nearest the query point.

City (stated location)	Query point	Offset (km)	r, 3 nearest stated	r, 3 nearest query point	Most-correlated cities
San Carlos (Pangasinan)	15.54°N, 120.89°E	72	0.73	0.80	Santa Rosa, Tarlac City, Antipolo
Antipolo (Rizal)	15.62°N, 121.19°E	115	0.26	0.70	Santa Rosa, San Carlos, Tarlac City
Santa Rosa (Laguna)	15.42°N, 120.94°E	125	0.55	0.78	San Carlos, Tarlac City, Antipolo
San Pablo (Laguna)	14.97°N, 120.62°E	126	0.68	0.79	Malolos, Olongapo, Valenzuela
Santiago (Isabela)	17.29°N, 120.44°E	135	0.31	0.89	San Juan, Vigan, Laoag
Malabon (Metro Manila)	15.64°N, 119.94°E	154	0.55	0.57	Tarlac City, Olongapo, Urdaneta
San Fernando (Pampanga)	16.62°N, 120.32°E	181	0.44	0.68	Baguio, Vigan, Santiago
Dagupan (Pangasinan)	17.71°N, 121.50°E	223	0.38	0.85	Tuguegarao, Cauayan, Ilagan
San Juan (Metro Manila)	17.74°N, 120.46°E	355	0.12	0.78	Vigan, Laoag, Santiago
Lucena (Quezon)	10.88°N, 122.60°E	356	0.50	0.33	Urdaneta, Tarlac City, San Carlos

Query points are identical in all eleven monthly copies of the city table in the 2026 edition[26]; the other 30 panel cities lie 0.0–4.9 km from their stated locations (Supplementary Table S2). Lucena's query point lies in Iloilo (Visayas), outside the panel, so its nearest panel cities are distant.

Each mislocated series behaves like its query location rather than its stated one. For nine of the ten, the mean correlation of daily $PM_{2.5}$ with the three panel cities nearest the query point exceeds that with the three nearest the stated location; for San Juan it is 0.78 against 0.12 (Table 1). The exception, Lucena, was queried in Iloilo, outside the panel. A rule based on concentrations alone does not separate the ten cleanly. It also flags seven correctly located cities whose correlations are distorted by the mislocated or duplicated series themselves (Supplementary Table S2), so we use it only as corroboration. We excluded the ten cities.

Fifteen of the remaining 30 cities form five groups whose series are copies of one another (Fig. 1b; Supplementary Table S16). In the hourly 2026 records from the same service[26], every within-group pair reports identical values in 95.2–100% of 6,158–7,283 matched hours, for each of the eight pollutants. Six neighboring pairs outside the groups, 11–52 km apart, match in at most 1.0% of hours for $PM_{2.5}$ and at most 5.1% for every other pollutant except NO, whose near-zero values coincide in up to 61% of hours. In the 2025 daily panel, group members have identical daily means on 344–360 of 365 days for every pollutant, and on the remaining days their $PM_{2.5}$ means differ by at most 1.7 μg m^{-3}. The members were queried at points 3–17 km apart, consistent with a shared cell of the provider's underlying model. The provider does not document that model or its grid (Methods), so we describe the groups as duplicated series. Merging each group leaves 20 distinct series, which we use as network nodes from here on. They are distinct, not statistically independent: neighboring nodes remain strongly correlated.

Both problems change downstream results. On the uncorrected 40-city panel, the lagged-correlation effective distance of our earlier draft (the |r| version in Methods) predicted next-day $PM_{2.5}$ 8.9% more accurately than geographic neighbors, in sample. The gain comes entirely from the ten mislocated cities (− 28.0% error relative to geography) and is absent for the other 30 (− 0.4%). In the uncorrected network, 18 of the 20 strongest edges involve a mislocated city; the strongest run from Antipolo, Santa Rosa and San Carlos, whose series come from Central Luzon, into Pasig, Taguig, Quezon City and Cavite. Duplicates act in the opposite direction: 15.5% of each city's five geographic nearest neighbors are near-copies of its own series ($r > 0.99$), which flatters any geographic baseline.

Lagged networks on the distinct series, and how many edges a null produces

For every ordered pair of the 20 nodes and lag orders of one to three days, the discovery pipeline (Methods) combines a Granger test[9], a lagged cross-correlation and a partial regression; it applies Benjamini–Hochberg control[27] within each lag, requires agreement of tests and a positive effect, keeps the lag order with the largest Granger F statistic, and keeps pairs active in at least three of four seasons. An edge in the co-pollutant networks is therefore a composite lagged association supported by at least two of three tests: a conditional Granger test, an unadjusted lagged cross-correlation and a conditional partial regression. An edge in the unconditioned network must pass both the Granger test, conditional on the target's own past only, and the cross-correlation. Neither kind of edge is evidence of out-of-sample improvement, which we test separately below, nor of transport or an emission source. Because two of three tests suffice, not every edge passes the Granger test: 120 of the 121 annual co-pollutant edges do, but 31% (December–February) and 40% (June–August) of the edges in the season-specific networks used for the direction tests pass only the cross-correlation and the partial regression (Supplementary Table S15). With the target's co-pollutants (NO_2, CO, SO_2, O_3) as regressors and two of three tests required, the pipeline retains 121 edges; without conditioning, requiring both the Granger test and the cross-correlation, it retains 63 (Table 2). Most retained models have lag order one (99 of 121 and 54 of 63). An order-L Granger model contains source lags 1 to L, so the order is a model choice, not a transport time.

Table 2

Network construction and calibration on the 20 distinct series. Four consensus variants separate co-pollutant conditioning from the number of tests. Counts are pair–lag tests (rows 1–4) or directed node pairs (rows 5–6).

Step	Co-pollutants, 2 of 3 tests (main)	Co-pollutants, 2 of 2 tests	No conditioning, 2 of 3 tests	No conditioning, 2 of 2 tests (main)
Granger significant after FDR (q = 0.10 within lag)	792	792	516	516
Lagged cross-correlation significant	729	729	729	729
Partial regression significant	433	not used	218	not used
Consensus with positive effect (pair–lags)	654	550	374	355
Directed pairs (lag order with largest F)	266	237	157	144
Active in ≥ 3 of 4 seasons = final edges	**121**	**114**	**67**	**63**
Circular-shift null: mean edges (95th pct.)	6.9 (12)	4.7 (9)	4.0 (8)	3.5 (7)
Mean null edges / observed (sensitivity)	6%	4%	6%	6%
Local-shift null: mean edges (95th pct.)	38.1 (51)	32.8 (45)	29.5 (39)	28.2 (38)
Edges reproducible by shared seasonal variation	32%	29%	44%	45%
HAC (Newey–West) chain: final edges (shared with classical)	115 (86)	113 (81)	—	56 (38)

Each variant tests 1,140 pair–lags (20 × 19 ordered pairs × 3 lags). Nulls: 500 surrogate panels each; every node's $PM_{2.5}$ and co-pollutants are shifted together, by 30 days or more (circular) or by 7–21 days in either direction (local). Proportions are mean null edges divided by observed edges, reported as a sensitivity result, not as false-discovery proportions. The top-ten rule (an edge is kept if it is among its source's ten strongest outgoing or its target's ten strongest incoming edges) removed no edge from any network; it does not cap node degree (Supplementary Table S1). HAC: Newey–West covariance with seven lags, used in place of the classical F-test at every step.

The two main networks differ both in conditioning and in the number of tests. Holding the consensus rule fixed separates the two effects (Table 2). Adding a third test changes little: 112 of the 114 two-test edges with co-pollutants remain, with 9 additions, and all 63 unconditioned edges remain, with 4 additions.

Conditioning changes much more. At the same two-test rule, the conditioned and unconditioned networks share 35 edges (Jaccard index 0.25); 79 are found only with conditioning and 28 only without.

To calibrate the full pipeline, including every filter, we reran it on 500 surrogate panels per null (Methods, Fig. 2). Each surrogate shifts every node's $PM_{2.5}$ and co-pollutants together in time, keeping each node's autocorrelation and cross-pollutant structure and breaking cross-node alignment at the scale of the shift[28]. With circular shifts of 30 days or more, which also remove the shared seasonal cycle, the pipeline returns on average 6.9 edges with conditioning (95th percentile 12) and 3.5 without (7), against 121 and 63 observed (4–6% across the four variants). Shifts of 7–21 days largely preserve seasonal structure and may preserve some shared episodes, and the pipeline then returns on average 38.1 and 28.2 edges, counts equal to a third and nearly half of the observed counts (counts, not the same edges). Every observed count exceeds every null run ($p = 0.002$, the smallest value attainable with 500 runs; evidence relative to each surrogate construction, not a general guarantee). These are sensitivity results, not estimates of the share of false edges: neither null removes all cross-node structure, because episodes lasting one to three weeks can stay partly aligned after a short shift, and seasonal variation and such episodes could account for a substantial part of the observed structure.

The series are strongly autocorrelated (median lag-one autocorrelation 0.69 after detrending), and stationarity tests disagree for many of them: an augmented Dickey–Fuller test[29] rejects a unit root for 62% of the 100 detrended series, and a KPSS test[30] rejects stationarity for 56%. Regression residuals remain autocorrelated for 98% of conditioned edges and 32% of unconditioned edges (Ljung–Box[31], seven lags, $p < 0.05$). Rerunning the whole pipeline with Newey–West[32] (HAC) Wald tests in place of the classical F-test gives 115 conditioned and 56 unconditioned edges, of which 86 and 38 are in the main networks (Table 2). Between 29% and 40% of the edges therefore depend on how serial dependence is treated.

Downwind alignment is detected for edges found without co-pollutant conditioning

ERA5 reanalysis[33] winds at 100 m over the node locations confirm the seasonal reversal in 2025 (Fig. 3a,b; Supplementary Table S7). The December–February mean wind blows from 57° (northeast) at 3.6 m s^{-1}, with a northerly component on 98% of days; the June–August mean blows from 218° (southwest) at 2.0 m s^{-1}, with a northerly component on 16% of days. In the transition seasons the mean wind blows from the east (83° in March–May, 74° in September–November), across rather than along Luzon's north–south axis, and in September–November it is weak (1.1 m s^{-1}). This agrees with the climatology of Philippine monsoon circulation[25].

Latitude is a coarse proxy for flow direction, but it gives a first check. In season-specific networks, a larger share of north–south edges points southward under the northeast monsoon than under the southwest monsoon in every full network, though not among edges found only with conditioning (Table 3). The full networks differ. Without conditioning, 69% of north–south edges point southward in December–February and 12% in June–August. With co-pollutant conditioning the shares are 60% and 44%, and neither differs from 50% ($p = 0.06$ and 0.21). Edges found by both networks reverse (71% and 8%); edges found only with

conditioning do not (55% and 61%). Because edges share nodes and node pairs recur across seasons, these tests use a node-latitude permutation null and node-cluster bootstrap intervals, which answer different questions (Methods): for the full conditioned network the seasonal difference (0.16) exceeds the permutation null (p = 0.027) but its interval includes zero (− 0.07 to 0.35), so we treat it as weak evidence.

Table 3

Edge direction in season-specific networks, December–February (DJF) and June–August (JJA) 2025.

Southward share of north–south edges (n) and mean edge–wind alignment (node-permutation p).

Edge set	DJF southward	JJA southward	Difference [95% CI]	p, difference	p, JJA vs 50%	Wind alignment, DJF	Wind alignment, JJA
Co-pollutant conditioning, 2 of 3 (main)	60% (147)	44% (184)	+ 0.16 [− 0.07, 0.35]	0.027	0.207	0.08 (0.072)	0.11 (0.002)
Co-pollutant conditioning, 2 of 2	68% (85)	33% (100)	+ 0.35 [0.03, 0.66]	0.003	0.037	0.17 (0.034)	0.19 (0.002)
No conditioning (main)	69% (64)	12% (66)	+ 0.57 [0.29, 0.79]	< 0.001	< 0.001	0.27 (0.004)	0.21 (0.014)
Found with and without co-pollutants	71% (48)	8% (60)	+ 0.62 [0.37, 0.86]	< 0.001	< 0.001	0.30 (0.002)	0.26 (0.004)
Found only with co-pollutants	55% (99)	61% (124)	−0.07 [− 0.28, 0.17]	0.769	0.113	−0.03 (0.667)	0.04 (0.224)
ERA5 meteorology conditioning	58% (114)	29% (106)	+ 0.29 [0.01, 0.53]	< 0.001	< 0.001	0.18 (0.002)	0.19 (0.002)
Meteorology and co-pollutants	58% (130)	43% (179)	+ 0.15 [− 0.06, 0.34]	0.036	0.157	0.08 (0.050)	0.11 (0.002)

Intervals: node-cluster bootstrap (2,000 resamples). p: node-permutation nulls (2,000 permutations of latitudes for shares; 500 permutations of node locations for alignment). Edges within 0.1° of latitude are excluded from shares. ERA5 meteorology: the target's daily wind components, boundary-layer height and precipitation.

Table 3

(continued). One contiguous northeast-monsoon season, December 2025 to February 2026.

Edge set	Edges	Southward (n)	Wind alignment (p)
Co-pollutant conditioning	210	68% (198)	0.19 (0.002)
No conditioning	168	62% (157)	0.18 (0.006)
Found with and without	133	67% (123)	—

The wind gives a sharper test. For each edge we averaged, over the season's days, the cosine between the ERA5 wind at its endpoints and its bearing; positive values mean the edge points downwind (Methods). The null permutes node identities among locations, keeping each network's degrees, reciprocity and dyads. Under the northeast monsoon, unconditioned edges align at 0.27 (72% downwind; p = 0.004) and edges found by both networks at 0.30 (p = 0.002); under the southwest monsoon, at 0.21 (p = 0.014, though its node-bootstrap interval, − 0.04 to 0.39, includes zero) and 0.26 (p = 0.004). We did not detect alignment for edges found only with co-pollutant conditioning (− 0.03 and 0.04; p = 0.67 and 0.22; Fig. 3c). Compared directly, they differ from unconditioned edges in December–February (difference 0.30, 95% interval 0.04 to 0.59; two-sided node-permutation p = 0.016) but not detectably in June–August (0.17; −0.19 to 0.41; p = 0.21), and with the same two tests required, we detect no effect of conditioning (p = 0.28 and 0.77; Supplementary Table S15), which does not show that the two procedures are equivalent. These comparisons are exploratory and not adjusted for the several edge sets and seasons compared. The full conditioned network aligns in June–August (0.11, p = 0.002) but not in December–February (0.08, p = 0.07). No edge set aligns in the easterly transition seasons, and the annual networks do not align with the annual-mean wind (Supplementary Table S7).

Conditioning on meteorology instead of co-pollutants keeps a wind-consistent signal, but this is a same-source check, because these networks use ERA5 wind as a regressor and are evaluated against it. Conditioned on the target's ERA5 wind components, boundary-layer height and precipitation, the network keeps 56 annual edges; its seasonal edges point southward in 58% of cases in December–February (not different from 50%, p = 0.24) and 29% in June–August, and align with the wind in both seasons (0.18 and 0.19; p = 0.002). Adding the co-pollutants to these controls brings back the co-pollutant pattern: 43% southward in June–August, not different from 50% (Table 3).

The 2025 December–February season joins January and February with December of the same year; lagged values never cross the gap between February and December (Methods). On one contiguous northeast-monsoon season, December 2025 to February 2026 (90 days; 2026 values from the same service[26]), the December–February result holds. With and without conditioning, 68% and 62% of north–south edges point southward, and both edge sets align with the wind (0.19 and 0.18; p ≤ 0.006; Table 3).

The annual network's roles and partitions are not stable

In the full-year conditioned network, a node's net out-degree (outgoing minus incoming edges) tends to rise with latitude (Spearman ρ = 0.40, p = 0.08): the Pasig, Cavite, Dasmariñas and Laguna nodes are net network receivers. This weak pattern does not survive re-estimation (Supplementary Fig. S1, Table S3). Without one season, ρ ranges from − 0.20 to 0.48 (significant in one of four rebuilds); without one month, from − 0.22 to 0.49 (3 of 12). Rebuilt edge sets overlap the full-year network with mean Jaccard indices of 0.36 and 0.68, the sign of net out-degree agrees for 61% and 74% of nodes, and effective distances correlate at 0.33 and 0.62 with their full-year values. We do not interpret sender and receiver roles, which in any case describe the lagged-association network, not physical sources or receptors.

Partitions are equally fragile. Ilagan has no outgoing edge, so its distances rest on the rule for unreachable pairs, and every method separates it alone at two groups (silhouette[34] 0.60; node-permutation p = 0.10). At

three groups, Ward clustering[35] separates 15 nodes from four and one (silhouette 0.25; p = 0.005), but the distance matrix is not Euclidean (classical scaling[36]: 5 negative eigenvalues of 20, 8% of the eigenvalue mass), and average linkage[37] and k-medoids[38] do not reproduce the split (k-medoids: 11–8–1, silhouette 0.04, p = 0.27; Supplementary Table S4). We report no partition.

Re-testing the 121 conditioned edges month by month shows episodic activity, most in November (53 edges); with about 30 days per month the pattern is descriptive (Supplementary Fig. S2).

Networks add little to forward-time prediction or warning beyond proximity

We converted the networks into effective distances[39] and asked whether a node's $PM_{2.5}$ is predicted better from its nearest nodes in such a distance than from its nearest geographic neighbors[40]. We used two protocols: a next-day forecast, the exp(− distance)-weighted mean of the five nearest nodes on the previous day, and a contemporaneous spatial estimate, the mean of the five nearest nodes on the same day (Methods). Every distance, including detrending and standardization, was estimated on the training days only. Intervals come from a moving-block bootstrap over days[41].

Within 2025, holding out one month at a time, plain correlation between series improves next-day error by 11.7% relative to geography (95% CI 6.3–16.4%), and a lagged-correlation distance that keeps only positive links by 8.6% (0.8–14.9%) (Fig. 4; Supplementary Table S17); the filtered network's distance is worse than geography (+ 14.0%). These are blocked, within-year evaluations, with training days on both sides of each held-out block. In a forward-time test[42], trained on 2025 and evaluated on January–October 2026, the gains disappear: plain correlation − 3.1% (− 7.6% to + 0.5%), the lagged-correlation distance + 2.5% (− 3.5% to + 7.7%) and the network distance + 15.7%; a monthly rolling origin over 2026 gives the same picture. Persistence has at least 17% lower error than every spatial predictor in every scheme (17–19% lower than the best, plain correlation; 19–28% lower than geography). The contemporaneous spatial estimate behaves the same way (Supplementary Table S6).

Averaging neighbors is a weak use of a network, so we also tested forecasts that combine each location's own history with its network neighbors (Methods; Supplementary Table S8). A ridge regression on a location's own last three days, fitted on 2025 and tested on 2026, beats persistence by 7.6% two days ahead and 11.8% three days ahead (95% CIs 0.1–12.4% and 2.5–17.8%) and matches it one day ahead. Adding the lags of the location's network sources does not improve these forecasts (− 0.8% to + 0.8%; every interval includes zero). Refitted monthly, unconditioned network sources reduce one- and two-day error by a further 3.0–3.3% (intervals exclude zero), but beat the same number of random or nearest locations by only 1.0–1.1% (every interval includes or touches zero). The network carries a small, perishable forecasting signal that geography largely reproduces.

Exposure management acts on high days, so we also scored the forecasts as warnings of days above 15 μg m^{-3}, the World Health Organization 24-hour guideline[43], and above 25 and 35 μg m^{-3}, the upper limits of the 'Good' and 'Fair' categories of the Philippine $PM_{2.5}$ index[44] (Methods; Supplementary Tables S9 and S14).

Because high days come in spells, every model ranks them above other days one day ahead with an area under the ROC curve (AUC)[45] of 0.90–0.92. The harder task is an onset: a day above the threshold after a day at or below it. Fitted on 2025, the models rank 2026 next-day onsets with AUC 0.78–0.87 and catch 42–70% of them at a 10% false-alarm rate read from the test-period ROC, or 35–72% with the alarm level set on earlier months (realised false-alarm rates 8–17%). Network sources raise next-day onset AUC over own history by 0.015–0.023 (every interval excludes zero), but the same number of nearest nodes gives most of this gain (network minus nearest 0.001–0.005, no interval excluding zero). Two and three days ahead, whether a day begins a new spell depends on intervening days not known at forecast time, so we scored those warnings operationally: on days at or below the threshold at forecast time, against the target day. The models then reach AUC 0.71–0.76 and 0.63–0.73, against 0.70–0.71 and 0.64–0.65 for persistence; the unconditioned network beats own history in 1 of 12 cases and the nearest nodes in four of six when fitted once (none when refitted monthly), and the co-pollutant network never does (Supplementary Table S14). None of these intervals is adjusted for the many thresholds, leads, models and fits, so cases that exclude zero are exploratory.

Neighbors help more in the seven nodes of Metro Manila and CALABARZON, a secondary analysis chosen after the all-node results. One day ahead and fitted on 2025, network sources raise onset AUC there from 0.77–0.80 to 0.81–0.85 (+ 0.037 to + 0.046), but the nearest nodes do as well (network minus nearest − 0.008 to + 0.024, no interval excluding zero; Supplementary Table S9).

Against reference stations, the product tracks daily variation poorly and reads about half the measured level

We compared the distinct series with daily 24-hour $PM_{2.5}$ from DENR-EMB continuous stations (16 with 2025 data, 23 with January–March 2026 data), released under a freedom-of-information request[46], matching each station to the node whose query point is nearest (0.2–12 km in Metro Manila and CALABARZON; Methods; Fig. 5; Supplementary Table S10). Across 13 stations in Metro Manila and CALABARZON (2,855 station-days in 2025), the median daily correlation is 0.15 (range − 0.31 to 0.49), the correlation of day-to-day changes 0.09, and the model reads 0.48 times the station level; in January–March 2026 (17 stations) the correlation is 0.08 and the ratio 0.44. Skill at identifying station days above 15 or 25 μg m^{-3} is near chance (median Heidke skill score 0.07 in 2025). Agreement is barely better at a one-day shift (median r 0.21 against 0.18 for the 12 stations with at least 100 days) and no better for stations within 5 km of their node's query point (median r 0.15, against 0.18 beyond), or for stations with more data (Supplementary Table S13). Baguio does better (0.34; 0.63 in early 2026).

A grid-cell model is not expected to match a point measurement[47], so we also compared regional means; the answer depends on which stations are averaged and on their dates (Supplementary Table S13). The mean of the nine mutually coherent Metro Manila stations, on the 320 days of 2025 with at least four of them, correlates with the three Metro Manila nodes at 0.04 daily and − 0.60 for 30-day means, but the contributing stations change through the year. Fixed sets do not settle this, because they report in different months: three northern stations (Ateneo, Marikina, North Caloocan) on 85 days from 25 August to 31 December (r = 0.52 daily, 0.66 for 30-day means), four southern ones (Makati, Pateros, Pasig, Parañaque) on 167 days from 7 January to 28 September (− 0.19 and − 0.88). Sharing 6 days, the two sets mix location with season, so these data do not establish whether the product reproduces or reverses the Metro Manila seasonal cycle. In every

set the model reads 0.34–0.60 of the station level. A boundary-layer adjustment from an earlier draft, min(1000 m / ERA5 boundary-layer height, 3.17), restores the level (median ratio 1.08) but raises the median station correlation only to 0.19.

Stations at the stated locations of four mislocated cities are consistent with the diagnosis from the query coordinates, which remains the primary evidence: 'Antipolo' correlates at − 0.23 with the Antipolo station over 297 days, 'Santa Rosa' at − 0.10 (2026), 'Dagupan' at − 0.09 and 'San Carlos' at 0.14, and the 'San Fernando' series correlates − 0.28 with the La Union station 0.4 km from its query point.

Discussion

The most consequential result concerns the data rather than the method. A panel of 40 named cities held 20 distinct series. A quarter of its series had been queried at other places, and half of the rest were copies of a neighbour. Neither was visible from the names; both appeared once the provider's query coordinates were read and values compared record by record. The checks are cheap, need nothing beyond the source files, and should precede any network inference on city-labelled data. Uncorrected, these artefacts produced apparent long-range links and a 9% predictive advantage over geography that was 28% where series were mislocated and near zero elsewhere. A mislocated series is indistinguishable from strong long-range coupling, which makes a network analysis a natural place for such artefacts to look like discoveries.

On the distinct series, the networks are best read as lagged associations within one model product. The surrogate results are comparisons under two specified null constructions, not estimates of the observed networks' error rate. Shifts of 30 days or more, which keep each series' autocorrelation but remove cross-node alignment and the shared seasonal cycle, leave few edges (4–6% of the observed counts); shifts of 7–21 days, which largely preserve seasonal structure and may preserve some shared episodes, give edge counts equal to a third to a half of the observed counts. Neither construction removes every source of shared structure, residuals remain autocorrelated in 98% of conditioned-edge regressions, and inference robust to serial dependence changes 29–40% of the edges. Agreement among dependent tests on the same series is not itself a multiple-testing correction, and we do not treat it as one.

The ERA5 benchmark tests which edge sets align with the wind. We detected downwind alignment under both monsoons for edges found without co-pollutant conditioning and for edges found by both networks, and the December–February result holds on a contiguous December–February season; the meteorology-conditioned network does too, but that is a same-source check, because ERA5 wind is among its regressors. This is a consistency check, not proof of transport. The wind was not used to build the other networks, but the concentrations come from an atmospheric model that is itself driven by meteorology, so the check shows that the product carries wind-consistent lagged structure; it does not show that the same structure exists in measured concentrations.

The station comparison bounds what the product represents. It tracks daily variation poorly and reads about half the measured level, whatever the station's distance or coverage; its seasonal agreement cannot be judged, because complete station sets cover different months. Part of the gap is representativeness[47]. Instrument effects, such as beta-attenuation monitors reading high in humid air[48], are possible but cannot be

assessed without station-specific instrument and quality-assurance records; the daily result does not depend on them. The networks therefore describe structure within the product's field. A second global model differs on the same station-days, though not uniformly: Copernicus CAMS, taken at the product's locations because the product could not be re-queried at station coordinates (Supplementary Table S12), has a median daily correlation of 0.28 against 0.15, paired intervals that favor it at 5 of 13 stations and the product at 2, and lower mean absolute error at 10. This shows that the product's weak agreement is not shared by every global model here, not that model data in general track the stations; CAMS did worse at five provincial stations (median r = 0.02, against 0.18). The CAMS series is the archived day-0 forecast run, issue time uncontrolled, so we use it as a concentration estimate, not to judge forecasting skill.

Co-pollutant conditioning did not behave as its rationale predicts, but the evidence on why is limited. If conditioning removed emission-driven associations and exposed masked transport, the edges it adds should follow the wind at least as well as the rest. We did not detect alignment for them, yet they differ detectably from unconditioned edges only in December–February; with the same two tests required, we detect no effect of conditioning, and in June–August the conditioned edges that fail the Granger test align worse than those that pass it (Supplementary Table S15). Part of the pattern may therefore reflect the two-of-three rule rather than conditioning itself. That receptor co-pollutants respond to the meteorology that moves $PM_{2.5}$, so that conditioning on them removes wind-driven signal, remains a hypothesis. We treat co-pollutant conditioning as a sensitivity analysis, not a transport filter, and would report the edges robust to both conditioning and the test rule.

The annual network does not generalize. Its weak sender–receiver gradient, its partitions and its distances change when a season or month is left out, and its distances predict worse than geography. Correlation-based distances beat geography within 2025 but not forward in time. Network neighbors added about 3% to own-history forecasts only when refitted monthly, little more than random or nearest locations, and for warning of new high spells the nearest locations did as well a day ahead; scored operationally two and three days ahead, the unconditioned network beat the nearest locations in four of six cases when fitted once, mostly because adding the nearest locations lowered skill, and beat own history in only one; refitted monthly it beat neither. For exposure management the network's use is modest: it suggests which neighboring series to watch, a choice proximity makes as well a day ahead.

Several limitations apply. The data are model output; the provider does not document the model or its grid, and one announcement names SILAM[49,50]. The station comparison covers Metro Manila, CALABARZON, Baguio and a few other sites, with uneven station coverage, a regional mean that depends on the stations included and the months they cover and, for 2026, data that EMB had not yet quality-controlled; station coordinates were read from addresses, and the product was not re-queried at station coordinates. The analysis covers one year, with 2026 used for replication and forward-time tests; the contiguous December–February check covers a single northeast-monsoon season (June–August 2025 is already contiguous). The wind test uses 100 m ERA5 winds on a 0.25° grid, averaged over each edge's endpoints and over the season; trajectories over one to three days would sharpen it. Linear Granger tests assume stationarity that the series only partly satisfy. Twenty nodes form a small network. The ten excluded cities could be recovered by querying the provider at their intended coordinates.

City-labelled model data can support network analysis only after integrity checks, and the networks they support are exploratory, product-specific lagged associations. Where those associations are tested against an independent circulation record, the unconditioned links align with the wind as transport would; we did not detect such alignment for the links added by co-pollutant conditioning. Against reference stations, those associations describe the product's field, not measured concentrations.

Methods

Data and provenance

The 2025 panel comes from the Kaggle dataset Philippine Cities Air Quality Index Data 2025[22], version 429, downloaded on 29 January 2026 with kagglehub. Its monthly files hold hourly records from 1 January 2025 00:00 to 31 December 2025 23:04 Philippine time. The columns (coord.lat, coord.lon, main.aqi and components.co, no, no2, o3, so2, pm2_5, pm10, nh3) are the fields of the OpenWeatherMap Air Pollution API response[21]; the provider gives concentrations in μg m^{-3} and serves current, forecast and historical endpoints by latitude and longitude. The provider does not state the model, grid or resolution behind the service; a 2020 provider announcement names SILAM, the Finnish Meteorological Institute's atmospheric composition model[49]. The publisher's city table (Cities.csv) lists, for each city name, the coordinates returned with its query. Hourly records were converted to Philippine time and averaged to daily means per city. The balanced panel holds 40 Luzon cities with complete daily records (14,600 city-days). Stated locations follow PhilAtlas and the Philippine Standard Geographic Code.

For replication and forward-time tests we used the 2026 edition[26], version 331, downloaded on 3 November 2026: hourly records for 1 January to 31 October 2026, aggregated to the same 20 nodes with the same daily-mean rule, and its city table, which is identical in all eleven monthly copies for the 40 panel cities.

Integrity diagnostics

Query location. For each city we computed the great-circle distance between its stated location and the coordinates in the publisher's city table, and flagged distances above 25 km. Any threshold from 5 to 72 km gives the same ten cities (Supplementary Table S2). As corroboration we compared, for each flagged city, the mean correlation of daily $PM_{2.5}$ with the three panel cities nearest its stated location and with the three nearest its query point.

Concentration rule. For all 40 cities we also applied a rule using concentrations only: flag a city when its mean correlation with its three nearest cities is below 0.80 and its mean correlation with its three most-correlated cities exceeds that by more than 0.20. Results for all cities, and a sensitivity grid over both thresholds, are in Supplementary Table S2 and Supplementary Note 1.

Duplicates. Among the remaining 30 cities we linked pairs whose daily $PM_{2.5}$ correlation exceeded 0.99 and took connected components as candidate groups; the groups are the same at any threshold from 0.95 to 0.99. For every within-group pair and for six nearby pairs outside the groups we counted, for each pollutant, the days with identical 2025 daily means and the share of matched hourly 2026 records (same hour,

Philippine time) with identical values, and fitted a regression of one member on the other. Each group was merged into one node: the daily mean of its members for every pollutant, at the centroid of their coordinates.

Network construction

Each node's $PM_{2.5}$ and co-pollutant series were linearly detrended and standardized on the days analyzed. For every ordered pair (i, j) and lag L = 1, 2, 3 days, the conditioned Granger test compares

$$PM_j(t) = \alpha + \Sigma_{l=1..L}\, \beta_l\, PM_j(t - l) + \Sigma_{l=1..L}\, \gamma_l\, PM_i(t - l) + \Sigma_p\, \delta_p\, X_{p,j}(t) + \varepsilon(t)$$

with the same model without the γ terms, by an F-test. $X_{p,j}$ are the target's NO_2, CO, SO_2 and O_3; the unconditioned test omits them. We did not use NO or NH_3: NO is near zero in this product (median daily mean 0.08 μg m^{-3}; below 0.01 μg m^{-3} on 10% of city-days), and NH_3 is mainly agricultural rather than a tracer of the combustion, industrial and photochemical processes the conditioning is meant to absorb. Benjamini–Hochberg control at q = 0.10 is applied to the Granger p-values within each lag. The lagged cross-correlation $r(PM_i(t - L), PM_j(t))$, which is not adjusted for the target's past or the regressors, counts if $p < 0.10$ and $|r| > 0.15$. The partial-regression test is the coefficient of the source's lagged value $PM_i(t - L)$ in a regression of the target's $PM_j(t)$ on its own lags 1 to L, its co-pollutants (when conditioning) and $PM_i(t - L)$; it counts if $p < 0.10$ and the coefficient exceeds 0.1 in absolute value. These two auxiliary tests are not corrected for multiplicity; the consensus requirement is a filter, not a correction, and the calibration below assesses the whole pipeline. An edge also needs a positive effect, defined as a positive estimated sum of the γ coefficients or a positive cross-correlation, whether or not that test was significant; the rule does not require every significant test to be positive. The Granger F-test assesses the source coefficients jointly; it does not test the sign of their sum, so for that test we report only the estimated sign. In the annual co-pollutant network 100 of 121 edges have a positive estimate for every significant test; 18 have a significant joint Granger test with a negative estimated coefficient sum, and 11 a significant negative partial-regression coefficient (Supplementary Table S15). A strict subset of the season-specific co-pollutant networks, requiring a significant Granger test and a positive estimate for every significant test, shows the same direction pattern (Supplementary Table S15 note). For each pair the lag order with the largest Granger F is kept: at order L the Granger model contains source lags 1 to L, while the cross-correlation and the partial regression use the source at lag L only, so the selected order is a model order and not an estimate of travel time. Each candidate pair is then re-tested at its selected order on the days of each season (DJF, MAM, JJA, SON), re-standardized within season, and counted as active if $p < 0.15$; pairs active in at least three seasons are kept. These four seasons are the climatological three-month blocks, not temperature seasons: in Luzon, DJF falls within the northeast (Amihan) monsoon and JJA within the southwest (Habagat) monsoon[25], and MAM and SON span the transitions between them. Finally an edge is kept if it is among its source's ten strongest outgoing edges or its target's ten strongest incoming edges by F. Because the rule keeps the union of the two lists, a node can keep more than ten edges in either direction (Supplementary Table S1); on 20 nodes the rule removed no edge from any network. DJF joins January, February and December 2025. All lagged regressions and correlations are computed on the full date sequence with omitted days set to missing, and a row enters only if the target day and every lagged value lie on retained days. Observations separated by a held-out month or season, or by the gap between February and December within the 2025 December–February season, therefore never become adjacent lags; detrending and standardization use the retained days only. An earlier version of this analysis concatenated the retained

days, joining 28 February to 1 December within the December–February season; correcting it changed the annual networks from 113 to 121 and from 62 to 63 edges and weakened the December–February direction results.

Consensus variants. The main conditioned network requires two of three tests (conditioned Granger, cross-correlation, conditioned partial regression); the main unconditioned network requires both the unconditioned Granger test and the cross-correlation. To separate conditioning from the number of tests, we also built a conditioned network requiring both of two tests and an unconditioned network requiring two of three, with the partial regression run without co-pollutants.

Calibration and serial dependence

Surrogate nulls. For each surrogate, every node's $PM_{2.5}$ and its four co-pollutants were shifted together by the same offset, independently across nodes[28]: uniformly between 30 days and the series length minus 30 days with wrap-around (circular), or by 7 to 21 days in a random direction (local). The full pipeline, including false-discovery control, consensus, positive effect, lag selection, the seasonal rule and the top-ten rule, was rerun on 500 surrogates per null for each of the four variants. We report the mean number of null edges and its ratio to the observed count as a sensitivity result: they show how many edges the pipeline produces under each null construction, not the false-discovery proportion of the observed network, and the null edges need not be the observed ones. The p-value is the share of null runs with at least as many edges as observed, counting the observed network; it is valid only under the null's assumptions. Local shifts largely preserve seasonal structure and can leave multi-week episodes partly aligned across nodes, and circular shifts can leave some alignment at long periods, so neither is a complete null of no cross-node dependence.

Diagnostics. We applied augmented Dickey–Fuller[29] (lag order by AIC) and KPSS[30] tests to each detrended series, and a Ljung–Box test[31] at 7 and 14 lags to the residuals of each final edge's full regression. As a robustness check we reran the pipeline with every Granger F-test replaced by a Wald test using Newey–West[32] covariance with seven lags.

Direction tests

Season-specific networks. For each season and each variant we ran the discovery steps on that season's days, without the seasonal-stability step, which is defined across seasons. An edge is southward if its source lies more than 0.1° north of its target and northward if more than 0.1° south; others are excluded. Every southward pair has a northward mirror, so 50% is the no-bias level. The prediction that north–south edges would dominate in DJF and weaken in JJA was stated in an earlier draft of this work, before the test was run.

Network-aware inference. Edges are not independent observations: they share nodes, and node pairs recur across seasons and conditions. For southward shares we used a null that permutes latitudes among nodes while keeping every network fixed (2,000 permutations), and intervals from a node-cluster bootstrap that resamples nodes with replacement and weights each edge by the product of its endpoints' multiplicities (2,000 resamples). The two procedures test different things. The permutation p-value for the DJF–JJA difference is one-sided and conditional on the observed seasonal networks: the same random reassignment of latitudes to nodes is applied to both seasons, so the null is that edge direction is unrelated to node position

in either season, and p is the share of permutations with a difference at least as large as observed, counting the observed one. The bootstrap is a two-sided 95% percentile interval for the same difference, treating the 20 nodes as a sample of locations; it shows how much the difference depends on which nodes are included. Neither re-estimates the networks, so neither carries the uncertainty of edge discovery. A small permutation p-value with an interval that includes zero, as for the full conditioned network (0.16; −0.07 to 0.35; p = 0.027) and the network conditioned on both meteorology and co-pollutants (0.15; −0.06 to 0.34; p = 0.036), means that the difference exceeds what random placement produces but is sensitive to which nodes are included. Edge-level Fisher and binomial tests from the earlier draft are not used.

Wind benchmark. Hourly ERA5[33] 100 m wind speed and direction, boundary-layer height and precipitation at the 20 node coordinates (16 distinct 0.25° grid points) were obtained from the Open-Meteo archive API[51] for December 2024 to February 2026, and averaged to daily wind vectors in Philippine time. For an edge from s to t active in a season, its alignment is the mean, over the season's days, of the cosine between the day's mean wind vector at s and t (the direction the air moves toward) and the bearing from s to t. A set's statistic is the mean alignment of its edges. The null permutes node identities among the 20 locations, carrying each location's wind series, while keeping the edge list (500 permutations); intervals use the node-cluster bootstrap (1,000 resamples). Differences in mean alignment between two edge sets were tested with the same node permutation applied to both sets (two-sided, 2,000 permutations), with node-cluster bootstrap intervals that draw one set of nodes for both (2,000 resamples). These comparisons are not adjusted for multiplicity.

Contiguous December–February. We rebuilt the DJF networks on 1 December 2025 to 28 February 2026, taking December from the 2025 panel and January–February from the 2026 edition with the same daily-mean rule.

Meteorological conditioning. We replaced the co-pollutants $X_{p,j}(t)$ by the target's daily ERA5 eastward and northward wind, boundary-layer height and precipitation (meteorology), or added them to the co-pollutants (meteorology and co-pollutants), and reran the annual pipeline and the season-specific networks. Because ERA5 wind enters these models as a regressor, their alignment with ERA5 wind is a same-source consistency check, not an independent test.

Stability, roles and partitions

Net out-degree is outgoing minus incoming edges; we call nodes with positive values network senders and those with negative values network receivers, which describes the lagged-association network only. Stability rebuilds the annual network without each season (4 rebuilds) and without each month (12), and reports the Jaccard index of edge sets, the Spearman correlation of effective distances with the full-year values, agreement in the sign of net out-degree and the Spearman correlation of net out-degree with latitude[52]. Partitions were computed from the symmetrised effective distance with Ward[35] and average[37] linkage and with k-medoids[38] (50 random starts) for k = 2 and 3, scored by silhouette[34]. Whether the distance is Euclidean was checked by classical scaling[36]. The contrast between within- and between-group $PM_{2.5}$ correlation (difference of medians) was tested by permuting group labels among nodes with group sizes fixed (5,000 permutations), not by treating node pairs as independent.

Effective distance

For the filtered network, transition probabilities are $P_{ij} = F_{ij} / \Sigma_k F_{ik}$ over node i's outgoing edges; edge length is $1 - \log P_{ij}$ and effective distance is the shortest-path length[39]. F measures evidence against a regression null, not a transport probability, so we use this only as a graph distance on the lagged-association network and attach no physical meaning to path lengths. In the full-year conditioned network Ilagan has no outgoing edge, so the 19 ordered pairs starting from it are unreachable; every other pair is reachable and every node has incoming edges. Distances are symmetrized as $(D_{ij} + D_{ji})/2$. Where a pair is unreachable (Ilagan's 19 pairs in the full-year network, or any pair in a rebuilt network), clustering uses 1.5 times the largest finite distance; the contemporaneous estimate skips unreachable nodes when choosing the five nearest, and the next-day forecast gives them the 95th percentile of the target's finite distances. The lagged-correlation distance applies the same construction to each season's lagged cross-correlations ($p < 0.05$, $r > 0.20$, lags one to three days, weight r) and takes the day-weighted mean of the four seasonal distances. Unlike the earlier draft, which used |r|, it keeps only positive correlations, so that a strong negative correlation is not treated as a short path in a predictor that averages concentrations with positive weights; the |r| version is reported for comparison. The plain-correlation distance is 1 − r of same-day $PM_{2.5}$.

Prediction

The next-day forecast for node j on day t is the exp(− D)-weighted mean of the five nodes nearest to j on day t − 1; for geographic neighbors the mean is unweighted. The contemporaneous spatial estimate is the unweighted mean of the five nearest nodes on day t; it is not a forecast. Persistence is the node's own value on day t − 1. Within 2025 we held out one month (12 folds) or one season (4 folds), with training days on both sides of each held-out block. The forward-time test trains on 2025 and evaluates on 1 January to 31 October 2026; the rolling-origin test[42], for each month of 2026, trains on all earlier days of 2025–2026 and evaluates on that month. In every fold, detrending, standardization and every distance were estimated on the training days only. For each scheme, out-of-fold daily errors were pooled and the ratio of mean absolute errors was bootstrapped with 7-day moving blocks[41] (2,000 resamples).

Forecast models. For each node and lead h of one to three days, a direct ridge regression predicts $PM_{2.5}(t)$ from values known at t − h: the node's own values at t − h, t − h − 1 and t − h − 2 (own history), plus the same three lags of every node with an edge into it in the conditioned or the unconditioned network. Controls add the same lags of the same number of nearest nodes, of the same number of randomly chosen nodes (20 draws, averaged), or of all 19 other nodes. Inputs are standardized on the training days, and the ridge penalty is chosen on the training days (25 values from 10^{-3} to 10^{3}). Models were fitted on 2025 and tested on 1 January to 31 October 2026, and, in a rolling origin, refitted with both networks re-estimated on all earlier days before each month of 2026. Errors were pooled over nodes and days, and paired differences bootstrapped with 7-day moving blocks.

Early warning. Each model's forecast of $PM_{2.5}(t)$, issued at the origin t − h, served as a warning score for $PM_{2.5}(t) > c$, with c = 15, 25 or 35 μg m^{-3}. AUC over all node-days of the test period measures how well the score ranks exceedance days above other days. Onsets were scored in two ways. The retrospective, conditional onset AUC uses only node-days on which every value from $PM_{2.5}(t - h)$ to $PM_{2.5}(t - 1)$ was at or below c, so that a day counts only if it begins a new spell; for $h \geq 2$ this cohort depends on values after the

forecast origin and cannot be identified when the forecast is issued, so it measures discrimination of new spells after the fact. The operational evaluation uses only information available at the origin: a node-day is eligible if $PM_{2.5}(t - h) \le c$, the event is $PM_{2.5}(t) > c$, and each forecast is scored on its target day whether or not an intermediate day already exceeded c; we report how many events follow such an intermediate exceedance. For h = 1 the two definitions coincide. In either cohort, the probability of detection when the probability of false detection is 10% was read from the test-period ROC curve by linear interpolation, a retrospective quantity. A prospective version sets the alarm level at the 90th percentile of scores on eligible non-event node-days in the earlier months of the test period (from the third month on) and applies it to each later month, reporting detection and the realized false-alarm rate. Differences in onset AUC between models were bootstrapped with 7-day moving blocks over days (1,000 resamples). The seven nodes in Metro Manila (NCR) and CALABARZON (Region IV-A) were analyzed separately, as a secondary analysis chosen after the all-node results; a merged node is assigned to the region of most of its members. The daily series cover 1 January 2025 to 31 October 2026 without gaps.

Reference-station comparison

Daily 24-hour $PM_{2.5}$ means from DENR-EMB continuous ambient air quality monitoring stations were obtained under a freedom-of-information request[46]: 16 stations with 2025 data and 23 with January–March 2026 data in Metro Manila, Rizal, Laguna, Bulacan, Zambales, La Union, Ilocos Norte, Abra and Baguio, plus two manual samplers in Dagupan and San Carlos (one 24-hour sample about every sixth day; 43 samples each, January–September 2025). EMB states that its January–March 2026 data for Metro Manila and Region IV-A had not completed quality assurance, so 2026 is reported separately. Station coordinates were read from the addresses in the release (uncertainty about 2 km). Each station was matched to the node whose query point is nearest, and each mislocated city's station also to the series carrying its name. For each station and period we report the shared days, means, bias, Pearson and Spearman correlations, the Pearson correlation of day-to-day changes, mean absolute and root-mean-square errors and, for station exceedances of 15 and 25 $\mu g\ m^{-3}$, the hit rate, false-alarm ratio and Heidke skill score of model exceedance; stations with fewer than 20 shared days are omitted. Correlation intervals come from a moving-block bootstrap over calendar days (7-day blocks, 1,000 resamples; days missing from the record stay missing), which allows for serial dependence; 14-day blocks are a sensitivity check. Lag scans correlate each station with at least 100 days in 2025 with the model shifted by − 2 to + 2 days. For the regional test we averaged, on days with at least four of them, the nine Metro Manila stations with the highest mutual correlations (Ateneo, Marikina, Makati, both Caloocan sites, Pasig, Pateros, Parañaque, Valenzuela) and the three Metro Manila nodes. The boundary-layer adjustment uses the daily mean of ERA5 boundary-layer height at each node (Methods, Wind benchmark) and the cap of 3.17 from an earlier draft's calibration against a low-cost sensor in Makati. For comparison, Copernicus CAMS global $PM_{2.5}$[53] (0.4°) was obtained through the Open-Meteo air-quality API[54], whose archive serves the day-0 run of the CAMS global forecast (issue time not controlled by us; hourly values averaged to Philippine-time days), both at each station's coordinates and at the stated coordinates of the cities forming each node, so that CAMS can be matched to stations in the same way as the product. The product itself was not re-queried at station coordinates. The two products were compared on identical station-days, and differences in correlation, bias and mean absolute error were bootstrapped with 7-day moving blocks over days (1,000 resamples). Sensitivity analyses repeated the regional comparison with stricter presence rules and with fixed

station sets, reporting each set's days and date range, and compared stations by distance to their node's query point and by number of valid days.

Reproducibility

The analysis is implemented in Python (NumPy, SciPy, statsmodels, scikit-learn, NetworkX); the network code agrees with statsmodels' Granger test to within 10^{-11}. Scripts (folder code/ of the deposit): network_discovery.py (discovery pipeline, effective distance), lagcorr_distance.py (lagged-correlation distance), integrity_audit.py (query locations, concentration diagnostics and duplicates), network_calibration.py (variants, nulls, diagnostics, HAC), wind_benchmark.py (ERA5 benchmark, direction tests, contiguous December–February, meteorological conditioning), edge_set_comparison.py (tests passed by each edge, comparisons between edge sets), network_prediction.py (prediction), network_structure.py (stability and partitions), monthly_activation.py (monthly activation), forecast_skill.py (forecast models), early_warning.py, early_warning_metro.py and early_warning_operational.py (early warning), station_comparison.py, station_diagnostics.py, station_blh_adjustment.py, station_cams.py, station_products_paired.py and station_sensitivity.py (reference stations, the CAMS comparison and sensitivity analyses), and make_tables.py and make_figures.py; the deposit's README maps every table and figure to its script and result files.

Use of AI tools

Claude (Anthropic), used through Claude Code, assisted with writing and running the analysis code, re-running and checking the analyses, preparing figures and tables, and drafting and editing the text and the deposit documentation. The authors directed the work, checked the code, results, text and references, and take full responsibility for the content. All reported results are produced by the deposited code and data.

Tables

Query points are identical in all eleven monthly copies of the city table in the 2026 edition[26]; the other 30 panel cities lie 0.0–4.9 km from their stated locations (Supplementary Table S2). Lucena's query point lies in Iloilo (Visayas), outside the panel, so its nearest panel cities are distant.

Each variant tests 1,140 pair–lags (20 × 19 ordered pairs × 3 lags). Nulls: 500 surrogate panels each; every node's $PM_{2.5}$ and co-pollutants are shifted together, by 30 days or more (circular) or by 7–21 days in either direction (local). Proportions are mean null edges divided by observed edges, reported as a sensitivity result, not as false-discovery proportions. The top-ten rule (an edge is kept if it is among its source's ten strongest outgoing or its target's ten strongest incoming edges) removed no edge from any network; it does not cap node degree (Supplementary Table S1). HAC: Newey–West covariance with seven lags, used in place of the classical F-test at every step.

Intervals: node-cluster bootstrap (2,000 resamples). p: node-permutation nulls (2,000 permutations of latitudes for shares; 500 permutations of node locations for alignment). Edges within 0.1° of latitude are excluded from shares. ERA5 meteorology: the target's daily wind components, boundary-layer height and precipitation.

Declarations

Competing interests.

The authors declare no competing interests.

Author Contribution

J.B.B. conceived the study, performed the analysis and wrote the manuscript. C.P.M. supervised the work and revised the manuscript.

Acknowledgement

J.B.B. is supported by a DOST-SEI ASTHRDP graduate scholarship. ERA5 data are produced by the Copernicus Climate Change Service at ECMWF and were accessed through Open-Meteo.

Data Availability

The source data are public: the Kaggle datasets Philippine Cities Air Quality Index Data 2025 (version 429; https://www.kaggle.com/datasets/bwandowando/philippine-cities-air-quality-index-data-2025) and Philippine Cities Air Quality Index Data 2026 (version 331; https://www.kaggle.com/datasets/bwandowando/philippine-cities-air-quality-index-data-2026), and ERA5 via the Open-Meteo archive API (https://doi.org/10.5281/zenodo.7970649). The daily station data were released by DENR-EMB for research and publication under Freedom of Information request CO-2026-024602 and are included in the Zenodo deposit. The balanced panel, the integrity flags and node definitions, the ERA5 and CAMS extracts and all derived results are deposited at Zenodo (https://doi.org/10.5281/zenodo.23938113) under the Open Database License 1.0. During review, the same files are supplied to the editor and reviewers as a supplementary archive (Bongalos_Monterola_Luzon_PM25.zip).

Code availability

All analysis code is deposited with the data at Zenodo[55] (https://doi.org/10.5281/zenodo.23938113) under the MIT License; its README maps every table and figure to the script and result files behind it. During review the code is supplied in the same supplementary archive.

References

1. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* 3rd edn (Wiley, 2016).
2. Jacob, D. J. *Introduction to Atmospheric Chemistry* (Princeton University Press, 1999).

3. Apte, J. S., Marshall, J. D., Cohen, A. J. & Brauer, M. Addressing global mortality from ambient $PM_{2.5}$. *Environ. Sci. Technol.* **49**, 8057–8066 (2015).
4. Byun, D. & Schere, K. L. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* **59**, 51–77 (2006).
5. Grell, G. A. et al. Fully coupled online chemistry within the WRF model. *Atmos. Environ.* **39**, 6957–6975 (2005).
6. Thunis, P. et al. Source apportionment to support air quality planning: strengths and weaknesses of existing approaches. *Environ. Int.* **130**, 104825 (2019).
7. Liu, Y. et al. Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing. *Environ. Sci. Technol.* **39**, 3269–3278 (2005).
8. Di, Q. et al. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. *Environ. Int.* **130**, 104909 (2019).
9. Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
10. Spirtes, P., Glymour, C. & Scheines, R. *Causation, Prediction, and Search* 2nd edn (MIT Press, 2000).
11. Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).
12. Runge, J. et al. Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.* **5**, eaau4996 (2019).
13. Runge, J. et al. Causal inference for time series. *Nat. Rev. Earth Environ.* **4**, 487–505 (2023).
14. Pearl, J. *Causality: Models, Reasoning, and Inference* 2nd edn (Cambridge University Press, 2009).
15. Henneman, L. R. F., Choirat, C. & Zigler, C. M. Accountability assessment of health improvements in the United States associated with reduced coal emissions between 2005 and 2012. *Epidemiology* **30**, 477–485 (2019).
16. Carslaw, D. C. & Ropkins, K. openair — an R package for air quality data analysis. *Environ. Model. Softw.* **27–28**, 52–61 (2012).
17. Carslaw, D. C. et al. Recent evidence concerning higher NO_x emissions from passenger cars and light duty vehicles. *Atmos. Environ.* **45**, 7053–7063 (2011).
18. Monks, P. S. et al. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys.* **15**, 8889–8973 (2015).
19. Crutzen, P. J. The role of NO and NO_2 in the chemistry of the troposphere and stratosphere. *Annu. Rev. Earth Planet. Sci.* **7**, 443–472 (1979).
20. Smith, S. J. et al. Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.* **11**, 1101–1116 (2011).
21. OpenWeather. Air Pollution, A. P. I. November. (2026). https://openweathermap.org/api/air-pollution(accessed 3.

22. bwandowando. Philippine Cities Air Quality Index Data 2025 (OpenWeatherMap data), version 429. Kaggle (2026). https://www.kaggle.com/datasets/bwandowando/philippine-cities-air-quality-index-data-2025 Dataset record on 3 November 2026: 4,564 views, 1,214 downloads, three public notebooks.
23. OpenWeather, A. November OpenWeather. (2026). https://openweather.co.uk/about(accessed 3.
24. OpenAlex. Works search, accessed 3 November 2026: 330 works for OpenWeatherMap air quality, 114 for OpenWeather API air pollution, 66 for OpenWeatherMap air pollution PM2.5. https://api.openalex.org/works?search=OpenWeatherMap+air+quality
25. Matsumoto, J., Olaguera, L. M. P., Nguyen-Le, D., Kubota, H. & Villafuerte, M. Q. Climatological seasonal changes of wind and rainfall in the Philippines. *Int. J. Climatol*. **40**, 4843–4857 (2020).
26. bwandowando. Philippine Cities Air Quality Index Data 2026 (OpenWeatherMap data), version 331. Kaggle (2026). https://www.kaggle.com/datasets/bwandowando/philippine-cities-air-quality-index-data-2026
27. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R Stat. Soc. Ser. B*. **57**, 289–300 (1995).
28. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Phys. D*. **58**, 77–94 (1992).
29. Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431 (1979).
30. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom*. **54**, 159–178 (1992).
31. Ljung, G. M. & Box, G. E. P. On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303 (1978).
32. Newey, W. K. & West, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708 (1987).
33. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R Meteorol. Soc.* **146**, 1999–2049 (2020).
34. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
35. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
36. Gower, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338 (1966).
37. Sokal, R. R. & Michener, C. D. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **38**, 1409–1438 (1958).
38. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, 1990).
39. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
40. Fix, E. & Hodges, J. L. *Discriminatory Analysis: Nonparametric Discrimination — Consistency Properties*. Report No. 4USAF School of Aviation Medicine, (1951).

41. Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989).
42. Tashman, L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int. J. Forecast.* **16**, 437–450 (2000).
43. World Health Organization. *WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide* (World Health Organization, 2021).
44. Department of Environment and Natural Resources & Administrative Order, D. E. N. R. No. 2020-14: Establishing the Breakpoints for Particulate Matter 2.5 ($PM_{2.5}$) Air Quality Index (AQI) and Amending Section 5(a) of DAO 2013-13 Establishing the Provisional National Ambient Air Quality Guideline Values for Particulate Matter 2.5 ($PM_{2.5}$)DENR, Quezon City, 21 October (2020).
45. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).
46. Environmental, M. & Bureau *PM2.5 Monitoring Data for Luzon, 2021–2026 (Continuous Ambient Air Quality Monitoring Stations)*. Released under Freedom of Information request CO-2026-024602 (Department of Environment and Natural Resources, Quezon City, 28 July 2026).
47. Schutgens, N. A. J. et al. Will a perfect model agree with perfect observations? The impact of spatial sampling. *Atmos. Chem. Phys.* **16**, 6335–6353 (2016).
48. Hauck, H. et al. On the equivalence of gravimetric PM data with TEOM and beta-attenuation measurements. *J. Aerosol Sci.* **35**, 1135–1149 (2004).
49. OpenWeather November. Air Pollution API announcement (15 December 2020). (2026). https://openweather.co.uk/blog/post/keep-your-finger-pulse-air-pollution-api(accessed 3.
50. Sofiev, M. et al. Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin. *Geosci. Model. Dev.* **8**, 3497–3522 (2015).
51. Zippenfenig, P. & Zenodo Open-Meteo.com Weather API. (2024). https://doi.org/10.5281/zenodo.7970649
52. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
53. Copernicus Atmosphere Monitoring Service. CAMS global atmospheric composition forecasts. ECMWF Atmosphere Data Store https://November (2026). ads.atmosphere.copernicus.eu/datasets/cams-global-atmospheric-composition-forecasts(accessed 3.
54. Open-Meteo. Air Quality API (CAMS global domain). November (2026). https://open-meteo.com/en/docs/air-quality-api(accessed 3.
55. Bongalos, J. B. & Monterola, C. P. Data and code for: Data integrity auditing of model-derived $PM_{2.5}$ series and reanalysis-referenced evaluation of lagged dependence networks in Luzon, Philippines. Version 1.1. Zenodo (2026). https://doi.org/10.5281/zenodo.23938113

Figures

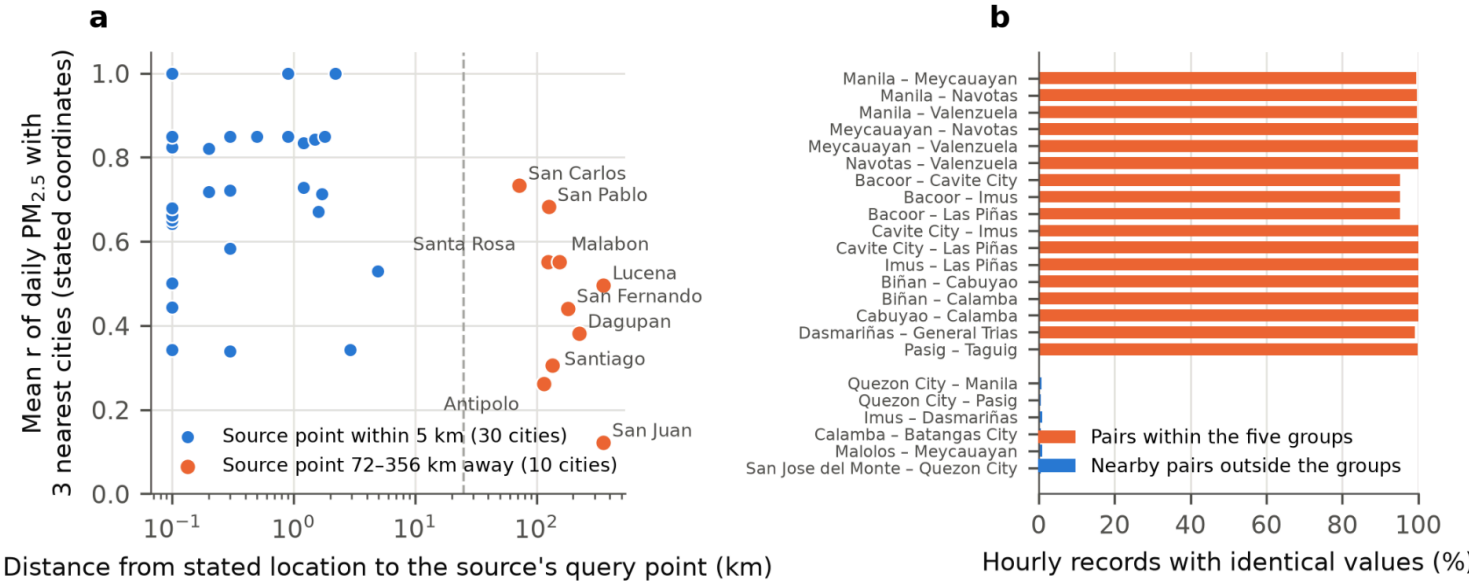

Figure 1

Two integrity problems in the city-labelled panel. (a) Distance from each city's stated location to the coordinates the source queried (publisher's city table), against the mean correlation of the city's daily $PM_{2.5}$ (2025) with its three nearest cities by stated coordinates. Cities queried at their stated location cluster on the left; the ten queried elsewhere lie at 72–356 km. Dashed line, 25 km. Cities with an offset of 0.1 km or less are plotted at 0.1 km. (b) Share of matched hourly records (2026, eight pollutants) with identical values: within the five duplicate groups (lowest pollutant per pair) and for nearby pairs outside the groups ($PM_{2.5}$).

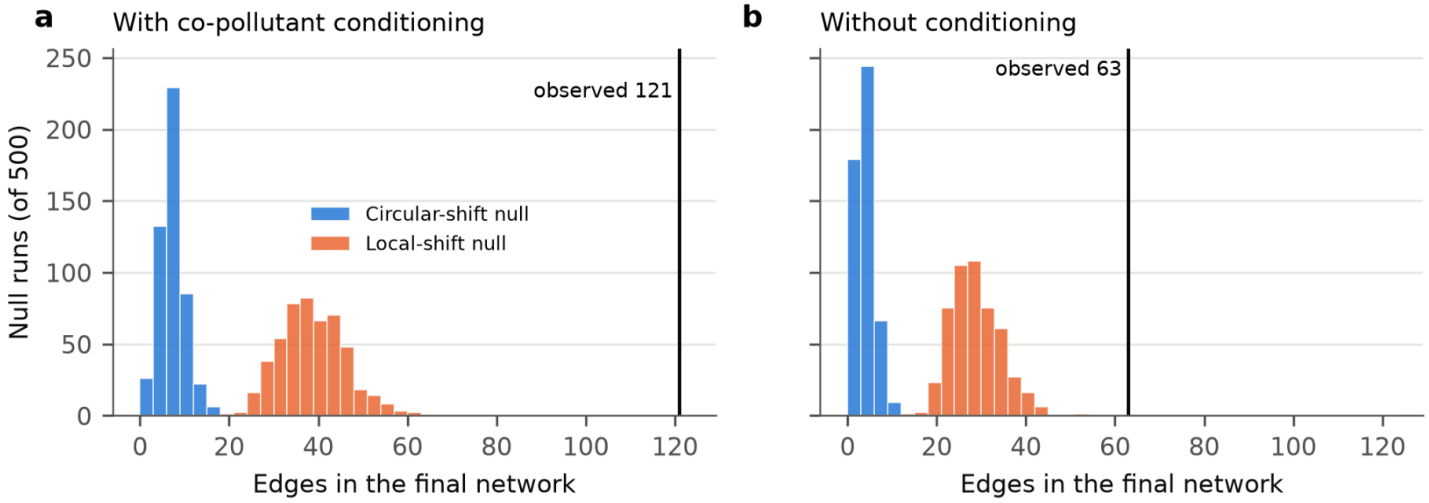

Figure 2

Edges produced by the full pipeline on surrogate panels. Final-network edge counts for 500 circular-shift surrogates (blue; shifts of 30 days or more, which also remove the shared seasonal cycle) and 500 local-shift

surrogates (orange; shifts of 7–21 days, which largely preserve seasonal structure), (a) with and (b) without co-pollutant conditioning. Vertical lines, observed networks.

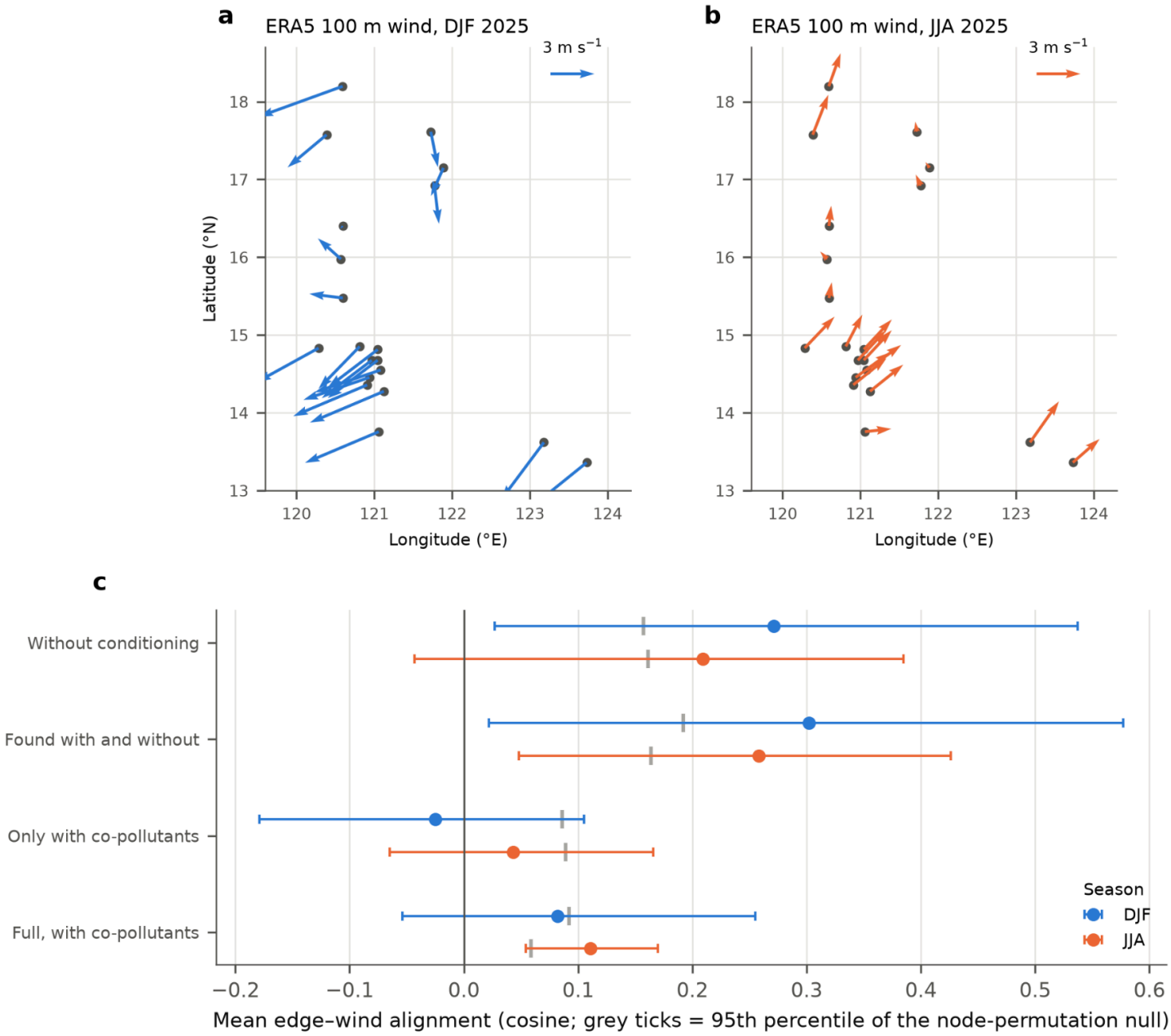

Figure 3

Edge direction against the monsoon wind. (a, b) Mean ERA5 100 m wind at the 20 nodes, December–February and June–August 2025; arrows point where the air moves. (c) Mean alignment (cosine) between edge bearing and the daily wind at the edge's endpoints, for season-specific networks: without conditioning, edges found with and without co-pollutant conditioning, edges found only with it, and the full conditioned network. Error bars, 95% node-cluster bootstrap intervals; grey ticks, 95th percentile of the node-permutation null.

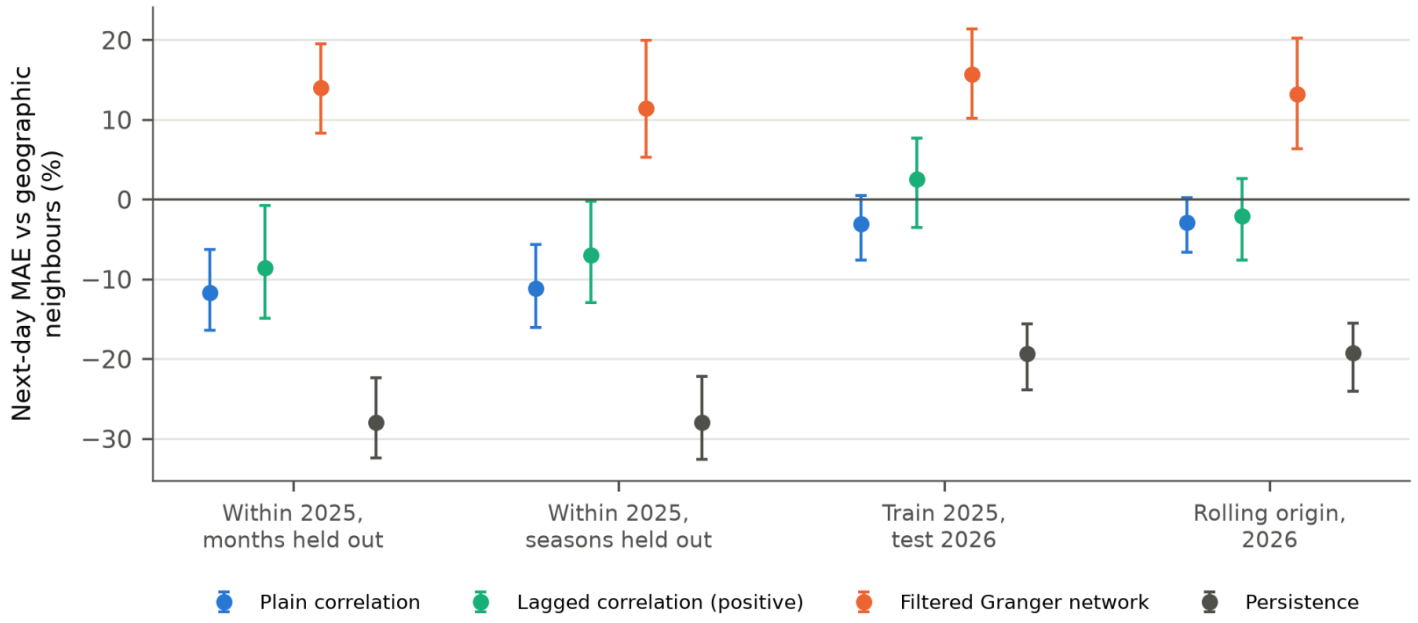

Figure 4

Next-day prediction relative to geographic neighbors. Change in mean absolute error relative to the five geographic nearest neighbors, with 95% moving-block bootstrap intervals (7-day blocks), for four evaluation schemes. Within-2025 schemes hold out one month (12 folds) or one season (4 folds) with training days on both sides; the forward-time schemes train on earlier days only. Negative values are better than geography. Distances re-estimated on the training days in every fold.

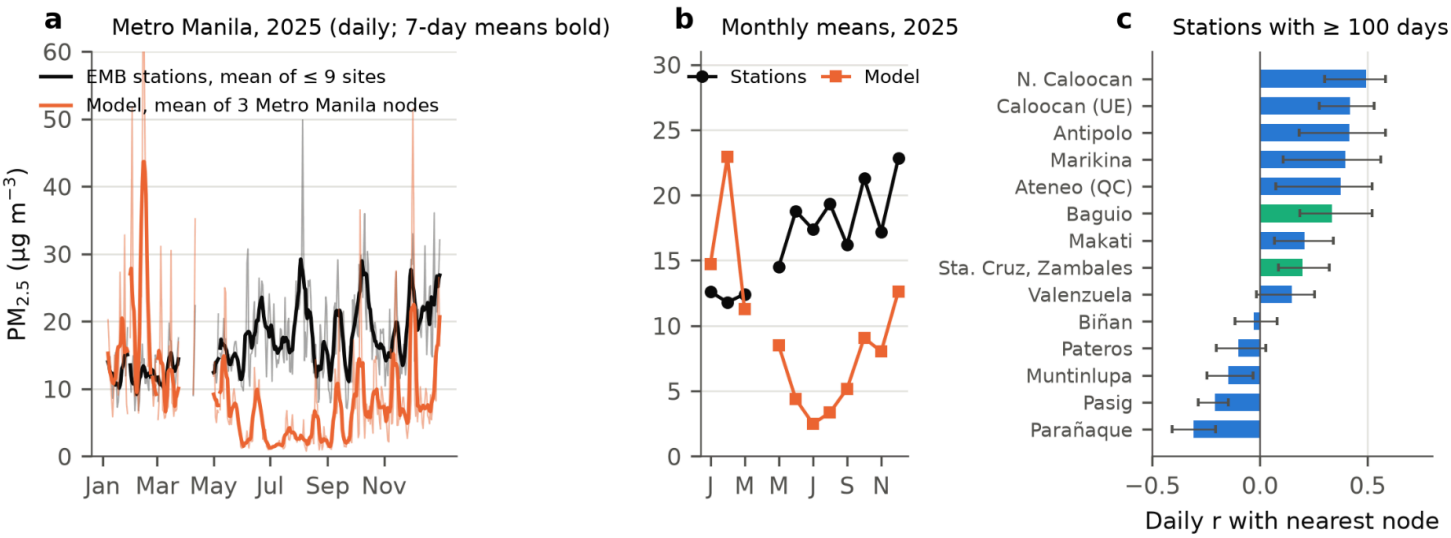

Figure 5

The product against DENR-EMB reference stations. (a) Daily $PM_{2.5}$ in Metro Manila, 2025: mean of up to nine EMB stations on days with at least four (black) and mean of the three Metro Manila nodes on the same days (orange); thin lines are daily values, bold lines 7-calendar-day means; gaps are days with fewer than four

stations. The stations contributing to the mean change through the year (Supplementary Table S13). (b) Monthly means of the same two series; April (3 days) is not shown. (c) Daily correlation between each station with at least 100 days in 2025 and its nearest node, with 95% moving-block bootstrap intervals (7-day blocks); blue, Metro Manila and CALABARZON; green, other Luzon.