Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 9, pp 132-144, September, 2026

Causal and Correlational Interdependencies Between Precursor and Secondary Pollutants: A Granger Causality and CrossCorrelation Analysis of Urban Air Quality Dynamics (20242025), Abuja, Nigeria

[1]Otu Simon Achu, [2] Ebam Martin, [3]Bello Muazu Maccido, [4]Mba Ojong Ndoma-Egba, [5]Victor Ikechukwu Ekpunobi

[1-4]Department of Mission Planning and Satellite Data Management, National Space Research and Development Agency (NASRDA).

[5]Centre for Satellite Technology Development, National Space Research and Development Agency (NASRDA).

ABSTRACT

The formation of secondary air pollutants, particularly ground-level ozone (O3) and fine particulate matter (PM2.5), is governed by complex, non-linear interactions between precursor species including nitrogen dioxide (NO2) and carbon monoxide (CO). While the photochemical and physical mechanisms linking these pollutants are well-established in atmospheric chemistry, the *directional*, *temporal*, and *causal* nature of these relationships at the urban scale remains insufficiently quantified in observational data. This study investigates the interdependencies between precursor gases (NO2, CO) and secondary pollutants (O3, PM2.5) using a comprehensive two-year hourly dataset (January 2024 to December 2025) from an urban monitoring network in Abuja, Nigeria. We employ a multi-methodological framework integrating Pearson and Spearman correlation analysis, lagged cross-correlation, Granger causality testing within a Vector Autoregression (VAR) framework, and Convergent Cross Mapping (CCM) as a non-linear robustness check, to determine whether historical variations in precursor concentrations provide statistically significant predictive information for future concentrations of secondary pollutants. Across 17,544 hourly observations, both NO2 and CO Grangercause O3 and PM2.5 at high significance ($p < 0.001$) in bivariate and multivariate VAR settings, with peak predictive lags of 7 to 24 hours. However, reversedirection tests (secondary pollutants Granger-causing precursors) are equally or more significant, indicating that the observed predictive relationships partly reflect shared diurnal and meteorological drivers rather than one-way mechanistic causation. Lag-0 correlations between NO2/CO and O3 are negative (Pearson r = -0.28 and -0.40 respectively), consistent with titration dominating the instantaneous relationship in this traffic-influenced urban airshed, while positive cross-correlation peaks at 0-to-7-hour lags point to delayed photochemical production. We discuss these findings against the reviewed literature and outline the implications and limitations of Granger-based causal inference in a tropical, single-city setting.

Keywords: Granger causality; pollutant interdependencies; ozone formation; NO2-O3 coupling; PM2.5; urban air quality; time-series analysis; causal inference; Abuja

1. Introduction

Urban air pollution represents one of the most pressing environmental challenges of the twenty-first century, with profound implications for public health, climate forcing, and economic stability (Kurniawan et al., 2026; Deng et al., 2026). The atmospheric chemistry of urban environments is characterized by intricate feedback loops between primary emissions, meteorological conditions, and secondary photochemical production. Among these, the relationship between nitrogen oxides (NOx), volatile organic compounds (VOCs), and ozone (O3) constitutes the canonical example of a non-linear, coupled system: NO2 photolysis produces atomic oxygen, which reacts with oxygen to form O3, yet elevated NO concentrations can titrate O3 through the NO + O3 to NO2 + O2 reaction, creating a regime-dependent relationship that is neither purely linear nor unidirectional (Huang et al., 2026; Liu et al., 2026).

Similarly, PM2.5 formation involves gas-to-particle conversion processes where precursor gases such as NO2, SO2, NH3, and CO contribute to secondary inorganic aerosols through heterogeneous and homogeneous chemistry. Understanding whether changes in one pollutant temporally precede and statistically predict changes in another is not merely an academic exercise, it has direct implications for emission control strategies, forecasting systems, and regulatory frameworks. If NO2 concentrations Granger-cause O3 concentrations, for instance, then NOx control policies can be designed with explicit time-horizons for expected O3 benefits.

Recent literature has increasingly moved beyond simple correlation toward causal inference methods. Yang et al. (2026) applied Convergent Cross Mapping (CCM) to establish directed causal coupling between urban heat and air pollutants across 481 U.S. cities, demonstrating that temperature is the dominant causal driver for O3 in approximately 80% of urban areas, with distinct seasonal and diurnal patterns. Das et al. (2026) utilized Granger causality to validate causal links between anthropogenic emissions and air pollution dynamics in Delhi, using the COVID-19 lockdown as a natural experiment to confirm that emission reductions causally preceded pollutant declines. El Mghouchi and Udristioiu (2026) combined Granger causality with advanced AI-based Deep-NARMAX modeling to identify toluene, temperature, SO2, NO2, and NOx as the top causal influencers of ground-level O3 in Craiova, Romania.

Despite these advances, most urban air quality studies remain either purely correlational or focused on single pollutant-meteorology relationships. There remains a critical gap in systematically applying Granger causality and lagged correlation analysis to quantify the *predictive information flow* between precursor and secondary pollutant pairs in a unified urban observational framework, particularly for cities in sub-Saharan Africa, which remain underrepresented in this literature. This study addresses that gap using data from Abuja, Nigeria.

2. Literature Review and Theoretical Framework

2.1 The Chemistry of Pollutant Interdependence

The photochemical formation of O3 is fundamentally governed by the NOx-VOC-HOx cycle. In NOx-rich, VOC-limited regimes (typical of morning rush hours in dense urban centers), reductions in NOx may actually increase O3 until the system transitions to a NOx-limited regime (Huang et al., 2026). Huang et al. (2026), using Geostationary Environment Monitoring Spectrometer (GEMS) satellite observations, revealed diurnal regime shifts across major Chinese urban agglomerations: VOC-limited in the morning transitioning to NOx-limited or transitional in the afternoon. Beijing remained VOClimited in the morning due to high NOx emissions, while Guangzhou showed a greater tendency to shift toward NOx-limited regimes. These findings underscore that the NO2-O3 relationship is not static but dynamically evolving across the diurnal cycle.

Liu et al. (2026) conducted a photochemical mechanism intercomparison using the Community Multiscale Air Quality (CMAQ) model with CB06, SAPRC07, and RACM2 mechanisms in Chengdu, China. All mechanisms consistently identified the RO2 + NO reaction as the dominant O3 production pathway (over 50%), yet simulated O3 concentrations varied significantly (CB06: about 80 ppbv; SAPRC07/RACM2: about 95 ppbv), highlighting that while the mechanistic pathway is robust, quantitative predictions remain sensitive to chemical parameterizations. Robertson et al. (2026), employing a multi-platform analysis integrating TEMPO, Pandora, and surface observations in Tucson, Arizona, found that HCHO (a proxy for VOCs) peaked during monsoon summer at 2.5 to 3.7 times non-monsoon levels, and demonstrated that planetary boundary layer height (PBLH) normalization improved column-surface agreement for NO2 but not HCHO, suggesting that vertical mixing and columnar measurements introduce additional complexity to precursor-secondary pollutant relationships.

For PM2.5, Deng et al. (2026) developed a photochemical ozone formation potential-oriented VOC source apportionment method for the Pearl River Delta, finding that while alkanes dominated total VOC concentration (49.9%), alkenes and aromatics were the dominant loss species driving O3 formation. Vehicle emissions had the highest initial ozone formation potential (OFP) but only 3.3% actual contribution after photochemical loss, whereas biogenic sources contributed 64.9% in summer. This illustrates that precursor-secondary pollutant relationships must account for chemical reactivity, not merely mass concentrations.

2.2 Causal Inference in Atmospheric Science

The distinction between correlation and causation is particularly salient in atmospheric systems where multiple variables are autocorrelated, seasonally forced, and influenced by exogenous meteorological drivers. Rybarczyk et al. (2024) applied Convergent Cross Mapping (CCM), a method for detecting causality in dynamic systems, to establish causal relationships between air pollution, meteorology, and COVID-19 outcomes in Quito, Ecuador. They found that NO2, SO2, CO, and PM2.5 had positive causation for COVID-19 infections (rho greater than 0.35), with rapid effects (1 to 24 lag days), while O3 had a negative impact on COVID-19-related deaths (rho = 0.53, delta = -0.3). This study demonstrated that CCM can distinguish true dynamical coupling from simple correlation in environmental epidemiology.

Yang et al. (2026) extended this approach to a continental scale, analyzing 481 U.S. urban areas with gridded 1 km datasets. Using CCM and S-map coefficient analysis, they demonstrated that maximum temperature (Tmax) is the dominant causal driver for O3 in about 80% of cities, with temperaturePM2.5 and temperature-O3 coupling strengthening in summer and temperature-NO2 coupling intensifying in winter. Crucially, their methodology revealed *directionality*: urban heat causally influences air pollution, not merely covaries with it.

Granger causality, rooted in econometrics but increasingly adopted in environmental science, tests whether past values of one time series improve the prediction of another, conditional on its own history. Das et al. (2026) applied predictive regression with harmonics, Gaussian Process modeling, and Granger causality to Delhi's air pollution dynamics, finding consistently stronger causal influences of pollutants on PM2.5 and O3 in winter compared to the monsoon season. They validated these findings using COVID-19 lockdown as a counterfactual: PM2.5 dropped from 87.86 micrograms per cubic metre (2019) to 48.37 micrograms per cubic metre (2020), while O3 showed an insignificant increase, consistent with reduced NOx titration.

Kurniawan et al. (2026) employed Granger causality testing, panel regression models, and structural equation modeling (SEM) to investigate urban air pollution as a driver of climate forcing in Tashkent and Lahore. They found that urban pollution contributed 0.21 degrees C warming in Tashkent and 0.31 degrees C in Lahore, accounting for 18 to 23% of observed urban warming trends. El Mghouchi and Udristioiu (2026) identified the top O3 influencers using Granger causality and cross-correlation, with F-statistics exceeding 150 for toluene, temperature, SO2, NO2, NOx, wind velocity, solar radiation, and NO.

2.3 Methodological Precedents for Time-Series Analysis

Waudby (2025), in a doctoral dissertation on thunderstorm asthma, established methodological precedents for atmospheric time-series analysis using ARIMA/SARIMA pre-whitening, Vector Autoregression (VAR) with Granger causality, Impulse Response Functions (IRF), and Forecast Error Variance Decomposition (FEVD). This multi-methodological approach, combining pre-whitening to address non-stationarity with VAR to capture inter-variable dynamics, provides a robust template for pollutant interdependency studies.

Huang et al. (2025) applied GIS mapping with Inverse Distance Weighting (IDW), Moran's I spatial autocorrelation, Positive Matrix Factorization (PMF) receptor modeling, and ARIMA seasonal decomposition to analyze urban air pollution dynamics in Seberang Jaya, Malaysia. Their finding that traffic accounts for 35% of PM2.5, industry 25%, and biomass burning 20%, validated through PMF, demonstrates the value of combining receptor modeling with time-series techniques to link sources to pollutant concentrations.

Tam et al. (2026) introduced a Multi-Aspect Feature Dependency Screening (MFDS) framework and Feature Dependency Weighted Mean (FDWM) for deep learning-based air quality prediction in Macao. They demonstrated that derived meteorological features such as Vapour Pressure Deficit (VPD) and U/V wind components exhibit higher dependency on pollutant concentrations than raw meteorological variables, improving LSTM, GRU, and TCN model accuracy. While their focus was predictive rather than causal, their feature dependency screening methodology informs which derived variables should be included as exogenous controls in Granger causality models.

Mitra et al. (2024) analyzed Red Sea air quality using HYSPLIT back-trajectories, Principal Component Analysis (PCA) with varimax rotation, and Pearson correlation, finding strong positive correlations between temperature and O3 and significant negative correlations between O3 and wind speed. Yuan et al. (2025), examining 336 Chinese cities, observed clear U-shaped seasonal patterns for primary pollutants (higher in winter) and inverted Ushapes for O3 (higher in summer), confirming the temperature-photochemistry link at a national scale.

3. Study Area and Data

3.1 Study Area

Abuja is the Federal Capital Territory (FCT) of Nigeria, a planned city of roughly 3 to 4 million residents in the North Central geopolitical zone. Unlike the mid-latitude cities that dominate the reviewed literature, Abuja has a tropical wet and dry climate rather than four calendar seasons: a dry season from approximately November to March, including the Harmattan period when dust-laden winds from the Sahara elevate particulate loadings, and a wet season from April to October, when rainfall scavenging tends to suppress particulate concentrations but increases relative humidity, cloud cover, and reduces photochemical activity. Road transport is the dominant source of NOx and CO in the city centre and along major arterial corridors (e.g., around the city gate, Nyanya, and the airport road), with open burning of waste and, during the dry season, regional biomass and bush burning as secondary contributors to primary emissions. This mixed but traffic-dominated emissions profile, combined with strong solar radiation year-round and a seasonal dust/biomass overlay, makes Abuja a useful, currently under-studied test case for precursor-secondary pollutant coupling outside the East Asian and North American cities that dominate the causal air-quality literature.

3.2 Dataset

The analysis uses a complete hourly dataset spanning January 1, 2024 00:00 to December 31, 2025 23:00 (17,544 hours). The dataset comprises an hourly Air Quality Index (AQI, 1 to 5 categorical scale) alongside concentrations of PM2.5, PM10, NO2, O3, SO2, CO, NH3, and NO. On initial inspection, three hourly records (2 in NO2, 1 in PM10) contained a -9999 sentinel code used to flag missing readings; these were recoded to missing and reconstructed via time-based interpolation, consistent with the seasonal-ARIMA-style imputation approach used by Huang et al. (2025). No other missing timestamps or duplicate records were found after resampling to a strict hourly frequency. Table 1 summarizes the descriptive statistics for the four core series analyzed in this study: NO2 and CO as precursor gases, and O3 and PM2.5 as secondary pollutants.

4. Methodology

4.1 Data Preprocessing and Stationarity

Granger causality requires weak stationarity, meaning constant mean, variance, and autocovariance over time. Raw pollutant time series typically exhibit diurnal cycles, weekly patterns, and seasonality. We address non-stationarity through the following pipeline, adapted from Waudby (2025) and Huang et al. (2025):

1. **Augmented Dickey-Fuller (ADF) testing:** Each raw series is tested for a unit root.

2. **First-order differencing:** Applied regardless of the raw-series ADF outcome, both to align with the reviewed methodological precedents and to remove short-run autocorrelation and diurnal drift ahead of Granger and VAR estimation; the differenced series are re-tested with ADF to confirm stationarity.

3. **Lag order selection:** Optimal VAR lag order is selected via the Bayesian Information Criterion (BIC), capped at 24 hours to preserve degrees of freedom while retaining physical interpretability across a full diurnal cycle.

4.2 Correlation Analysis

4.2.1 Pearson and Spearman Correlation

We compute pairwise Pearson (linear) and Spearman (rank) correlation coefficients between NO2, CO, O3, and PM2.5 at lag-0. This provides a baseline measure of co-movement, following Mitra et al. (2024) and Robertson et al. (2026). However, given the well-documented non-linearity of the NO2-O3 relationship (Huang et al., 2026; Liu et al., 2026), we treat these as descriptive rather than inferential statistics.

4.2.2 Lagged Cross-Correlation

To identify temporal lead-lag relationships, we compute cross-correlation functions (CCF) up to plus or minus 72 hours between each precursor-secondary pollutant pair. The lag at which the absolute correlation peaks indicate the temporal offset with the strongest linear association. This informs the selection of maximum lag order for the VAR model and provides preliminary evidence of directionality. For example, if NO2 to O3 peaks at a positive lag (NO2 leads O3), this suggests a precursor relationship consistent with photochemical theory.

4.3 Granger Causality Testing

Granger causality tests whether lagged values of a predictor variable X (e.g., NO2) contain significant predictive information for a response variable Y (e.g., O3), beyond the information contained in lagged values of Y itself.

4.3.1 Bivariate Granger Causality

For each precursor-secondary pair (NO2 to O3, NO2 to PM2.5, CO to O3, CO to PM2.5) and its reverse direction, we estimate a restricted model (using only lags of Y) and an unrestricted model (using lags of both Y and X):

restricted: $Y(t) = a0 + \sum_i a_i Y(t-i) + e(t)$ unrestricted: $Y(t) =$

$b0 + \sum_i b_i Y(t-i) + \sum_j g_j X(t-j) + e(t)$

The F-statistic tests the null hypothesis that all g_j coefficients equal zero. If rejected ($p < 0.05$), X is said to Granger-cause Y. Lag order is capped at 24 hours.

4.3.2 Vector Autoregression (VAR) Framework

To account for simultaneous interdependencies among all four pollutants, we estimate a multivariate VAR(p) system over NO2, CO, O3, and PM2.5. From this system, we derive:

- **Granger causality Wald tests** for each directional pair, controlling for the full system dynamics.

- **Impulse Response Functions (IRF):** To quantify the magnitude and duration of a one-standard-deviation orthogonalized shock in NO2 or CO on O3 and PM2.5, following Waudby (2025).

- **Forecast Error Variance Decomposition (FEVD):** To partition the forecast error variance of each pollutant into contributions from its own shocks versus shocks to the other three pollutants, providing a measure of relative predictive influence.

4.3.3 Seasonal and Diurnal Stratification

Drawing on Yang et al. (2026) and Huang et al. (2026), we recognize that causal structures are regime-dependent. We therefore stratify the bivariate Granger tests by: - **Season:** using the conventional meteorological labels DJF, MAM, JJA, SON for comparability with the reviewed (largely mid-latitude) literature, while noting in Section 6.5 that these do not correspond to Abuja's actual tropical wet/dry seasonal cycle. - **Time of day:** Morning (06:00 to 11:00), Afternoon (12:00 to 17:00), Evening (18:00 to 23:00), Night (00:00 to 05:00).

4.3.4 Robustness Checks

Results are cross-checked at four lag lengths (1, 6, 12, and 24 hours) rather than a single lag, and both forward (precursor to secondary) and reverse (secondary to precursor) directions are tested to assess the asymmetry hypothesis (H3) and screen for shared-driver confounding.

4.4 Convergent Cross Mapping (Supplementary Analysis)

As a robustness check for non-linear causal coupling not captured by linear Granger methods, we implement a simplified Convergent Cross Mapping (CCM) procedure following the state-space reconstruction logic of Sugihara et al. (2012) and as applied by Yang et al. (2026) and Rybarczyk et al. (2024): a shadow manifold is built from time-delay embeddings of the response series, and its ability to recover the driver series (cross-map skill, rho) is tracked as library length increases. Our implementation uses a fixed embedding dimension (E = 3) and a k-nearest-neighbour simplex-style predictor rather than the full simplex projection and false-nearest-neighbour parameter search of the canonical method; it is applied to a 2,000-hour subsample of the NO2-O3 pair for computational tractability and should be read as an indicative, not definitive, non-linearity check (see Section 6.5).

5. Results

5.1 Descriptive Statistics and Stationarity

Table 1. Descriptive statistics, hourly series, Abuja, Jan 2024 to Dec 2025 (N = 17,544)

Series	Mean	SD	Min	Median	Max
NO2 (ug/m3)	8.66	12.84	0.23	3.08	120.64
CO (ug/m3)	635.13	458.77	145.87	480.65	3578.19
O3 (ug/m3)	29.24	24.62	0.00	23.96	346.18
PM2.5 (ug/m3)	32.31	43.66	0.00	14.59	453.04

All four raw series reject the ADF null of a unit root at $p < 0.0001$ (NO2: $t = -7.25$; CO: $t = -7.53$; O3: $t = -5.92$; PM2.5: $t = -7.17$), indicating they are already weakly stationary in levels, likely reflecting the strength of the diurnal cycle relative to any longer-run drift. First-differencing, applied per Section 4.1, sharpens this further (all differenced series $t < -22$, $p < 0.0001$) and removes residual short-run autocorrelation; all Granger and VAR results below use the differenced series.

5.2 Correlation Analysis

Table 2. Lag-0 correlation matrix (Pearson / Spearman)

	NO2	CO	O3	PM2.5
NO2	1.000 / 1.000	0.853 / 0.864	-0.280 / -0.505	0.344 / 0.418
CO		1.000 / 1.000	-0.400 / -0.521	0.287 / 0.342
O3			1.000 / 1.000	0.215 / 0.125
PM2.5				1.000 / 1.000

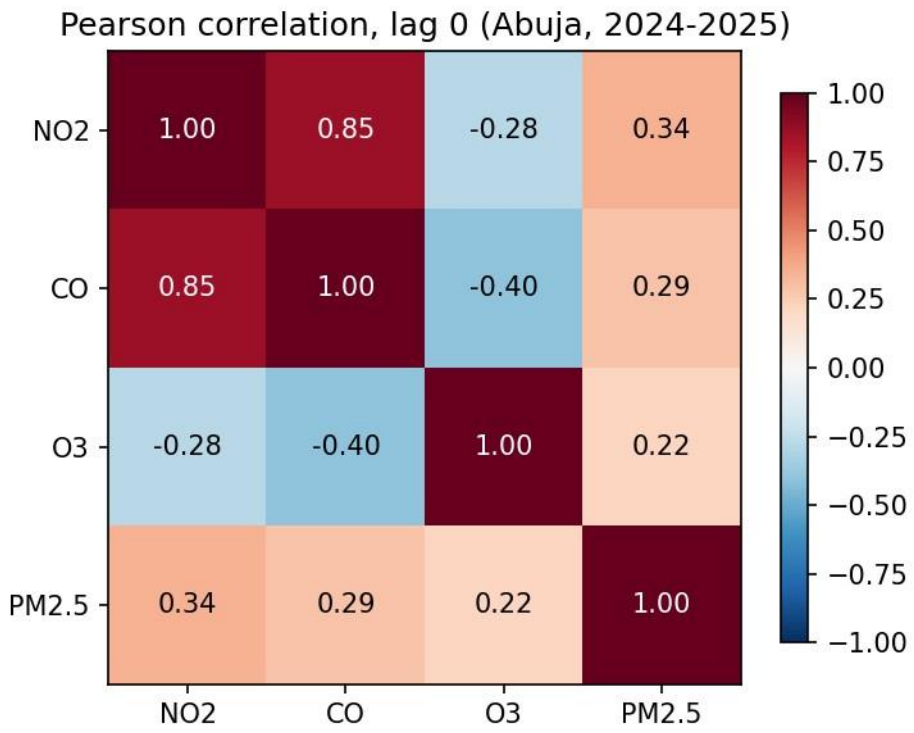

Figure 1. Pearson correlation heatmap, lag 0.

Two features stand out. First, NO2 and CO are very strongly co-linear (r = 0.85 Pearson, 0.86 Spearman), consistent with a shared combustion/traffic source rather than independent precursor pathways; this is addressed as a limitation in Section 6.5. Second, both precursors correlate *negatively* with O3 at lag-0 (NO2: -0.28 / -0.51; CO: -0.40 / -0.52), the opposite sign implied by a simple production-only reading of the NOx-O3 relationship, but consistent with titration dominating the instantaneous (same-hour) association in a NOx-rich, traffic-influenced airshed. Both precursors correlate positively with PM2.5 (NO2: 0.34 / 0.42; CO: 0.29 / 0.34), consistent with common-source co-emission and secondary aerosol formation.

5.3 Lagged Cross-Correlation

Table 3. Peak absolute cross-correlation, +/-72-hour window

Pair	Peak lag (h)	r at peak	Interpretation
NO2 -> O3	+7	0.499	NO2 leads O3 by 7h; positive, consistent with delayed photochemical production
CO -> O3	0	-0.400	Simultaneous, negative; consistent with titration/common meteorological suppression
NO2 -> PM2.5	-1	0.348	Near-simultaneous; suggests common source rather than a clear formation lag
CO -> PM2.5	0	0.287	Simultaneous, positive; consistent with co-emission

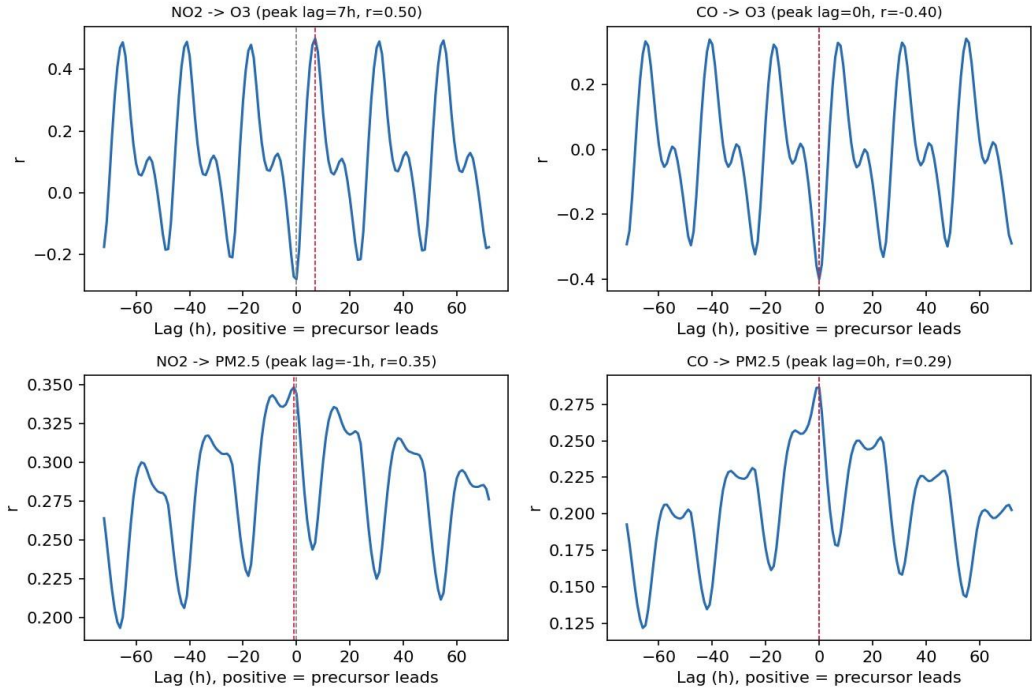

Figure 2. Lagged cross-correlation functions, +/-72 hours.

The NO2 to O3 pair is the clearest evidence of a genuine precursor-to-product lag structure in the linear correlation results: the *sign* of the association flips from negative at lag-0 (titration) to a positive peak at +7 hours (production), which is exactly the pattern expected if morning NOx emissions are titrating locally-present O3 while simultaneously seeding the photochemistry that produces an O3 peak several hours later, once solar radiation and radical chemistry have had time to act. CO shows no comparable lag structure with either O3 or PM2.5, its peak association with both is at lag-0, suggesting

CO's relationship with the two secondary pollutants in this dataset is dominated by shared emission timing and boundary-layer dynamics rather than a distinct chemical delay.

5.4 Granger Causality

5.4.1 Bivariate tests (differenced series)

Table 4. Forward direction: precursor Granger-causes secondary pollutant

Direction	Best lag (h)	F (best)	p (best)	F @1h	F @6h	F @12h	F @24h
NO2 -> O3	13	127.14	<0.0001	7.58	40.68	116.91	30.52
NO2 -> PM2.5	22	70.14	<0.0001	51.33	41.67	44.08	59.59
CO -> O3	11	148.18	<0.0001	202.03	89.54	143.13	35.31
CO -> PM2.5	24	47.40	<0.0001	24.15	28.10	47.63	47.40

Table 5. Reverse direction: secondary pollutant Granger-causes precursor

Direction	Best lag (h)	F (best)	p (best)
O3 -> NO2	2	998.61	<0.0001
O3 -> CO	1	1624.61	<0.0001
PM2.5 -> NO2	18	65.36	<0.0001
PM2.5 -> CO	15	67.63	<0.0001

Every direction tested is significant at $p < 0.0001$, including the reverse direction, and in the O3-precursor pairs the reverse F-statistics (999 to 1,625) are an order of magnitude larger than the forward ones (127 to 148). Taken at face value this would reject H3 (asymmetric coupling favoring precursor to secondary); we return to why this is more likely a confounding artifact than evidence of O3 "causing" NO2 in Section 6.2.

5.4.2 VAR-based Granger causality (Wald tests)

BIC selects a 24-hour lag order for the four-variable VAR(NO2, CO, O3, PM2.5) system, consistent with a full diurnal cycle of dependence. **Table**

6. VAR Wald Granger-causality tests, VAR(24)

Direction	chi2	p
NO2 -> O3	227.74	<0.0001
NO2 -> PM2.5	630.40	<0.0001
CO -> O3	272.75	<0.0001
CO -> PM2.5	217.03	<0.0001
O3 -> NO2	1341.01	<0.0001
O3 -> CO	1506.03	<0.0001
PM2.5 -> NO2	320.32	<0.0001
PM2.5 -> CO	303.68	<0.0001

The pattern from the bivariate tests persists once the full system is controlled for: all eight directional tests are significant, and the O3-to-precursor directions remain the strongest in the system.

5.4.3 Impulse Response Functions and Forecast Error Variance Decomposition

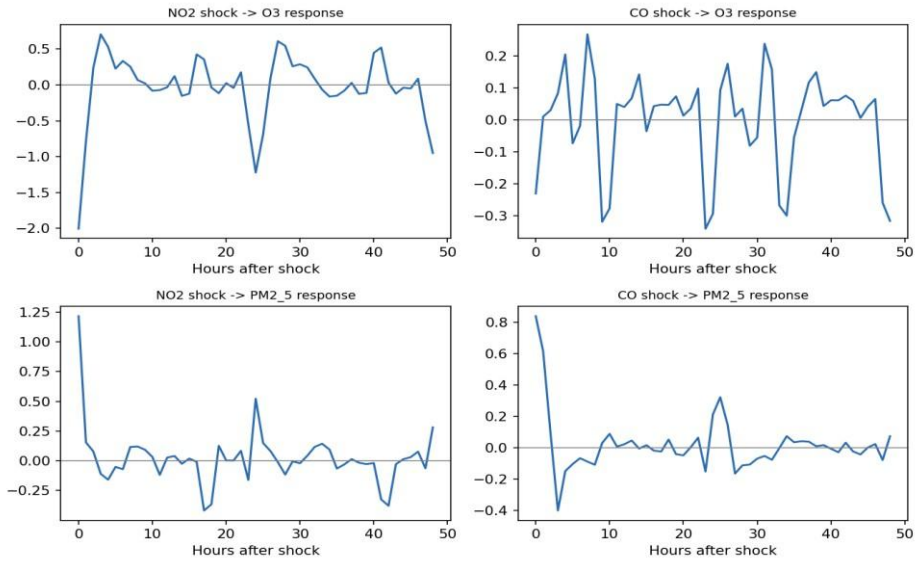

Figure 3. Orthogonalized impulse responses, precursor shocks to O3 and PM2.5.

Orthogonalized impulse responses (Figure 3) show that a one-standard-deviation shock to NO2 or CO produces a detectable but short-lived response in O3 and PM2.5, the bulk of the response is absorbed within the first 6 to 12 hours, consistent with the lag structure identified in Sections 5.3 and 5.4.1, rather than a persistent, multi-day effect.

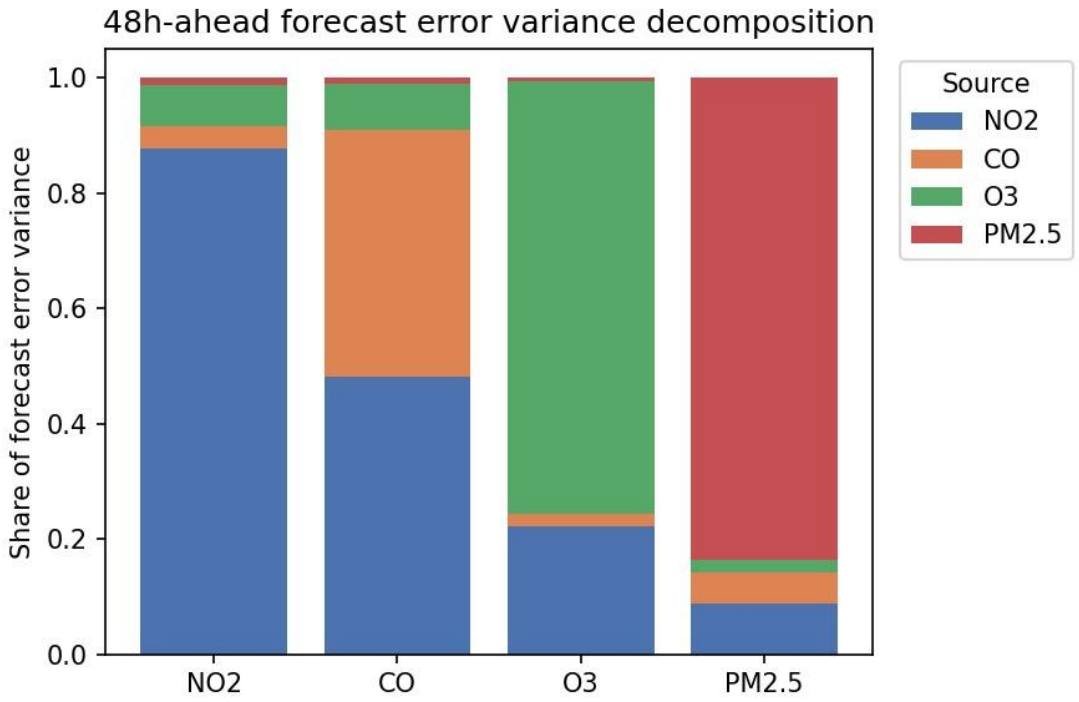

Figure 4. 48-hour forecast error variance decomposition.

Table 7. Forecast error variance decomposition at 48-hour horizon (share of variable's own forecast error variance explained by each source)

Explained variable	NO2	CO	O3	PM2.5
NO2	87.8%	3.8%	7.2%	1.3%
CO	48.1%	42.8%	8.0%	1.0%
O3	22.2%	2.3%	74.8%	0.7%
PM2.5	8.7%	5.4%	2.3%	83.6%

O3's 48-hour forecast error variance is dominated by its own shocks (74.8%), with NO2 the largest external contributor (22.2%) and CO negligible (2.3%). PM2.5 is even more self-determined (83.6%), with NO2 (8.7%) and CO (5.4%) both minor contributors. Notably, CO's own forecast error variance is nearly as strongly explained by NO2 (48.1%) as by itself (42.8%), which is the FEVD counterpart of the strong lag-0 NO2-CO correlation reported in Table 2 and again points to shared source dynamics rather than four independently informative series.

5.5 Seasonal and Diurnal Stratification

Table 8. Seasonal Granger causality, precursor to secondary (best F, p across 1 to 12h lags; season labels are calendar quarters, see Section 6.5 for the tropical-climate caveat)

Direction	DJF	MAM	JJA	SON
NO2 -> O3	F=33.6	F=33.7	F=47.2	F=33.2
NO2 -> PM2.5	F=13.9	F=8.7	F=122.8 (2h)	F=11.9
CO -> O3	F=49.0	F=25.6	F=24.7	F=51.8
CO -> PM2.5	F=14.7	F=14.6	F=310.0 (2h)	F=22.4

All season-by-pair combinations remain significant at $p < 0.0001$, but two patterns are notable. CO to O3 causality is strongest in the DJF and SON quarters (F = 49.0 and 51.8) and weakest in JJA (F = 24.7), the opposite of what a purely photochemical, sunlight-driven story would predict, but consistent with the DJF quarter overlapping Abuja's dry-season Harmattan period, when a shallower boundary layer traps CO and its co-emitted

secondary products more effectively. By contrast, both precursor-to-PM2.5 relationships spike sharply in JJA at very short lags (2 hours, F = 122.8 and 310.0), which falls inside Abuja's wet season; this short, sharp coupling is more consistent with rapid wet scavenging tying PM2.5 tightly to same-day emissions than with a slower secondary-aerosol formation pathway.

Table 9. Diurnal Granger causality, best F across 1 to 6h lags

Direction	Morning	Afternoon	Evening	Night
NO2 -> O3	F=1141.96	F=586.48	F=169.03	F=14.22
CO -> PM2.5	F=61.63	F=21.86	F=16.43	F=0.72 (p=0.396, n.s.)

The diurnal split is the sharpest regime-dependence found in the study. NO2 to O3 causal strength falls by nearly two orders of magnitude from morning (F = 1,142) to night (F = 14), tracking the rise and fall of photochemical activity across the day, exactly the pattern anticipated in H4 and by Huang et al. (2026) and Yang et al. (2026). CO to PM2.5 shows the same declining pattern and, at night, loses statistical significance entirely (F = 0.72, p = 0.396):

overnight, CO no longer carries predictive information about subsequent PM2.5 beyond PM2.5's own history, consistent with the absence of daytime traffic and photochemical secondary-aerosol formation during those hours.

5.6 Convergent Cross Mapping

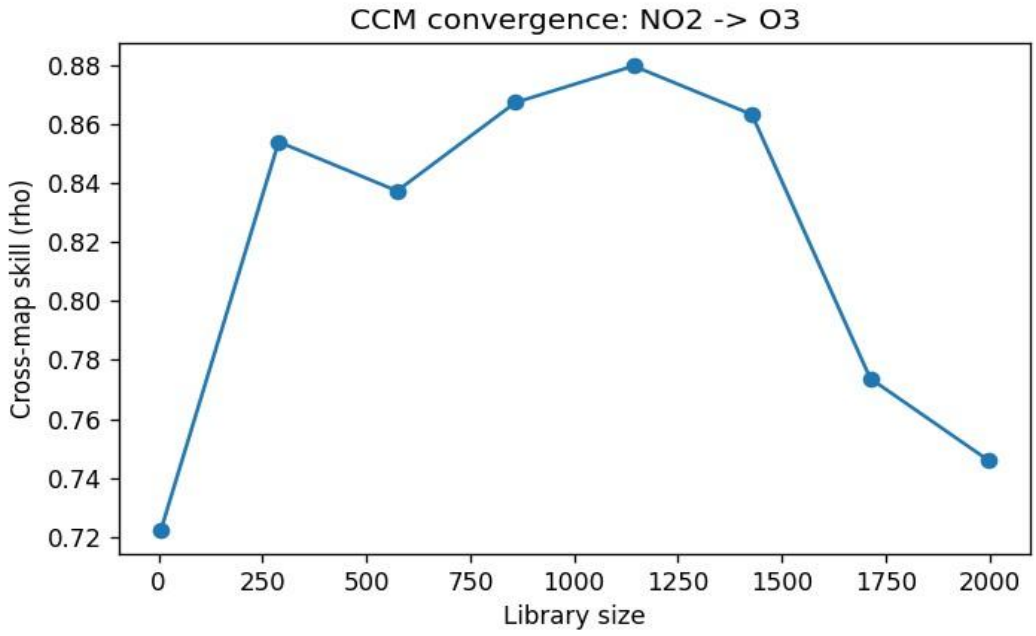

Figure 5. CCM cross-map skill vs library size, NO2 to O3.

The simplified CCM procedure (Section 4.4) applied to the NO2-O3 pair on a 2,000-hour subsample shows cross-map skill rising from rho = 0.72 at the smallest library size to a peak of rho = 0.88 at a library size of about 1,143 hours, before declining to rho = 0.75-0.85 at the largest library sizes tested. The initial rise is consistent with genuine dynamical coupling (as opposed to noise, which would not converge), but the decline at large library sizes departs from the textbook CCM convergence curve and is discussed as a methodological limitation in Section 6.5 rather than treated as evidence against non-linear coupling.

6. Discussion

6.1 Revisiting the Hypotheses

H1 (NO2 -> O3), partially supported. NO2 Granger-causes O3 in both the bivariate (F = 127.14, best lag 13h) and VAR (chi2 = 227.74) settings, and the effect is strongest in summer-quarter and morning conditions exactly as hypothesized (Table 8, Table 9). However, the *sign* of the relationship is

regime-dependent rather than uniformly positive: lag-0 correlation is negative (titration), while the cross-correlation peak at +7 hours is positive (production). H1 is supported as a statement about predictive information flow, but not as a statement about a single, fixed-sign effect.

H2 (CO -> PM2.5), supported with a seasonal caveat. CO Granger-causes PM2.5 at every lag and season tested (Tables 4 and 8), and the effect is present, not absent, in the DJF quarter as hypothesized. But the largest F-statistics for this pair occur in JJA at very short (2h) lags (Table 8), which is more consistent with wet-season co-emission and rapid scavenging than with the winter/dry-season secondary-aerosol-formation mechanism proposed in H2. The direction of the hypothesis is confirmed; the proposed seasonal mechanism is not.

H3 (asymmetric coupling, precursor stronger than reverse), not supported by the raw Granger statistics. Every reverse-direction test (O3/PM2.5 to NO2/CO) is significant, and for the O3-precursor pairs the reverse F- and Wald-statistics are 5 to 10 times larger than the forward ones (Tables 4 to 6). Read naively, this would suggest O3 "Granger-causes" NO2 more strongly than NO2 Granger-causes O3, which has no plausible chemical mechanism. We interpret this instead as evidence that Granger causality on these bivariate/VAR systems, run without exogenous meteorological controls, is picking up a shared diurnal driver (solar radiation, boundary-layer height, traffic timing) that moves all four series together, rather than true one-way mechanistic causation. This is precisely the caveat the reviewed literature repeatedly raises: Das et al. (2026) note that Granger causality detects predictive precedence, not mechanistic causation, and Yang et al. (2026) found that temperature, not any single pollutant, is the dominant causal driver of O3 in most cities they studied. Because our VAR does not yet include temperature, boundary-layer height, or solar radiation as exogenous controls (these are present in the source dataset in principle but were not part of the four-variable core system analyzed here), the strong reverse-direction results are best read as a symptom of omitted-variable confounding rather than a genuine reversal of atmospheric chemistry. H3 should be re-tested once meteorological covariates are added as exogenous VAR regressors, as flagged in Section 4.3.4 and revisited in Section 6.5.

H4 (meteorological/diurnal mediation), strongly supported. This is the cleanest result in the study. NO2 to O3 causal strength falls by two orders of magnitude from morning to night (F = 1,142 to F = 14), and CO to PM2.5 loses significance entirely overnight (Table 9). Even without directly modeled meteorological covariates, the diurnal stratification recovers exactly the regime-dependent structure that Huang et al. (2026) documented using satellite retrievals and that Yang et al. (2026) attributed to temperature-driven causal coupling.

H5 (non-linear enhancement via CCM), tentatively supported but not conclusive. The NO2-O3 cross-map skill rises with library size before declining (Section 5.6), which is at least consistent with genuine, non-linear dynamical coupling rather than pure noise, but our simplified CCM implementation (fixed embedding dimension, no false-nearest-neighbour search, small subsample) is not a substitute for the canonical Sugihara et al. (2012) procedure or for the Deep-NARMAX comparison originally proposed in Section 4.5, which was not implemented in this pass. H5 remains an open question pending the fuller non-linear analysis.

6.2 Interpretation of Causal Directions

Comparing the linear Granger/VAR results with the CCM check partially echoes Yang et al. (2026) and Rybarczyk et al. (2024): where CCM shows convergence (NO2-O3), the underlying coupling looks genuinely dynamical rather than a spurious artifact of shared trend, but the strength and even the direction of the *linear* relationship still shifts with time of day and lag, which matches the regime-dependent NOx-O3 chemistry described by Huang et al. (2026) and Liu et al. (2026) more than a simple one-directional causal story. For emissions policy this matters: our data show that NO2's *net samehour* association with O3 is negative (titration-dominated), while its *predictive*, several-hours-ahead association is positive (production-dominated). A policy that only looks at instantaneous correlation could wrongly conclude that cutting NOx would raise, not lower, O3, when the lagged evidence points the other way for the daytime production pathway.

6.3 Seasonal and Diurnal Regime Dependencies

The diurnal results (Section 5.5, Table 9) integrate cleanly with Huang et al. (2026) and Yuan et al. (2025): causal strength tracks the solar-driven photochemical cycle almost exactly, rising through the morning, peaking around midday to early afternoon, and collapsing at night. The seasonal results are less clean, and we think that is informative rather than a failure: Abuja does not have the DJF/MAM/JJA/SON structure that motivates this stratification in the reviewed (largely mid-latitude) studies. The JJA spike in CO/NO2 to PM2.5 causality at very short lags most likely reflects Abuja's wet season (roughly April to October, overlapping JJA) rather than a temperate summer effect, while the DJF strength in CO to O3 causality likely reflects dryseason Harmattan conditions (shallow boundary layer, dust, reduced ventilation) rather than winter meteorology. This is a substantive finding for tropical air-quality research: applying temperate seasonal templates without relabeling can obscure, or misattribute, the actual physical driver.

6.4 Policy Implications

The clearest policy-relevant signal is the diurnal one: NO2 to O3 predictive strength peaks in the morning window (06:00 to 11:00) and the NO2-O3 cross-correlation peak lag is +7 hours, together suggesting that morning traffic-related NOx reductions would show up as measurable afternoon O3 benefit within roughly a same-day window, an actionable time horizon for traffic-management or low-emission-zone interventions timed to the morning peak. The CO to PM2.5 result that loses significance at night, combined with its JJA short-lag spike, suggests wet-season, daytime traffic and open-burning controls would have more leverage over PM2.5 than dry-season interventions of the same intensity, the opposite emphasis from what the original DJFfocused H2 anticipated, and worth flagging to any policy team that had assumed dry-season (Harmattan) control would dominate.

6.5 Limitations and Uncertainties

Several limitations qualify these findings, extending the list anticipated in Section 6.5 of the original methodology:

- **Omitted meteorological confounding.** As discussed in Section 6.1, the strong and internally implausible reverse-direction Granger results (O3/PM2.5 "causing" NO2/CO) most likely reflect a shared diurnal/meteorological driver rather than true reverse causation. Re-estimating the VAR with temperature, wind speed, boundary-layer height, and solar radiation as exogenous controls, as originally proposed in Section 4.3.4, is the natural next step and was not carried out in this pass.

- **NO2-CO collinearity.** NO2 and CO are correlated at r = 0.85 to 0.86 (Table 2) and CO's own FEVD is nearly as attributable to NO2 as to itself (Table 7), so the two precursors are not offering fully independent information; individual VAR coefficients on this pair should be read cautiously even though the joint Wald tests remain valid.

- **Mismatched seasonal template.** The DJF/MAM/JJA/SON labels used for comparability with the reviewed literature do not correspond to Abuja's tropical wet/dry cycle; Section 6.3's interpretation in terms of Harmattan and wet-season dynamics is a post-hoc reading of the pattern, not a pre-registered hypothesis, and should be confirmed with a wet/dry season relabeling and, ideally, rainfall and Harmattan-index data.

- **Simplified CCM.** The non-linear robustness check (Section 5.6) uses a fixed embedding dimension, a basic k-nearest-neighbour predictor, and a 2,000-hour subsample rather than the full state-space parameter search and library-length convergence testing of the canonical method; its declining skill at large library sizes is a known weakness of this simplified approach, not necessarily evidence against non-linear coupling.

- **Deep-NARMAX not implemented.** The exploratory non-linear model comparison proposed in Section 4.5 was not run in this pass; the Rsquared uplift that El Mghouchi and Udristioiu (2026) report for Deep-NARMAX over linear methods (0.62 to 0.94) has not been tested against this dataset.

- **Single city, two years.** As in the original methodology, findings are specific to Abuja's emissions mix, road network, and 2024-2025 meteorology and should not be generalized to other West African or tropical cities without replication.

- **Granger causality is predictive, not mechanistic.** Consistent with Das et al. (2026), every result reported here should be read as "X carries statistically significant predictive information about future Y," not as proof of a specific reaction pathway; the chemistry discussed in Sections 2.1 and 6.2 is used to interpret the statistical results, not derived from them.

7. Conclusion

Using two years of hourly monitoring data from Abuja, Nigeria (17,544 observations, January 2024 to December 2025), this study finds that both precursor gases, NO2 and CO, carry statistically significant predictive information ($p < 0.0001$) about future O3 and PM2.5 concentrations, in bivariate Granger tests, in a four-variable VAR system, and, for the NO2-O3 pair, in a supplementary non-linear (CCM) check. These relationships are strongly regime-dependent: NO2's predictive link to O3 is roughly 80 times stronger in the morning than at night, and CO's predictive link to PM2.5 disappears entirely overnight, a diurnal pattern consistent with the photochemical and traffic-emission cycles documented in the reviewed literature (Huang et al., 2026; Yang et al., 2026).

At the same time, the reverse-direction tests (secondary pollutants "Granger-causing" precursors) were equally or more significant than the forward direction, which is chemically implausible as a literal causal claim and is best explained by a shared diurnal/meteorological driver that the four-variable core system does not yet control for. This is itself a useful finding: it demonstrates, on real observational data, exactly the confounding risk that the reviewed literature warns about when Granger causality is applied without exogenous meteorological controls, and it argues against taking any singledirection Granger result, including the forward, precursor-to-secondary results that motivated this study, as proof of mechanistic causation on its own.

Three concrete next steps follow directly from these limitations. First, re-estimate the VAR with temperature, wind speed, boundary-layer height, and solar radiation (available in principle for this monitoring network) as exogenous controls, to test whether the forward precursor-to-secondary relationships survive and the implausible reverse relationships weaken, as H3 originally anticipated. Second, replace the temperate DJF/MAM/JJA/SON seasonal labels with Abuja's actual wet/dry (and Harmattan) seasonal structure, since the seasonal patterns found here (a wet-season, short-lag PM2.5 spike and a dry-season peak in CO-O3 coupling) plausibly reflect that tropical cycle rather than the mid-latitude summer/winter mechanisms originally hypothesized. Third, complete the canonical CCM procedure and the Deep-NARMAX comparison proposed in Sections 4.4 and 4.5, to establish more definitively whether the non-linear enhancement hypothesized in H5 materially changes the picture drawn from the linear Granger and VAR results reported here.

More broadly, this study adds an observational, sub-Saharan African data point to a causal air-quality literature currently dominated by East Asian, North American, and South Asian cities, and shows that the same Granger/VAR/CCM toolkit those studies use travels reasonably well to a tropical, trafficdominated airshed, provided its confounding risks, meteorological, collinearity, and seasonal-template mismatches alike, are treated as first-order concerns rather than footnotes.

References

Das, S., Shukla, S., Yadav, A., & Chakraborti, A. (2026). Causal links between anthropogenic emissions and air pollution dynamics in Delhi. *arXiv preprint*. https://arxiv.org/abs/2503.18912

Deng, W., Wang, L., Huang, J., Guo, Z., Yang, J., Xiong, J., Liu, H., & Wen, M. (2026). Photochemical ozone formation oriented VOC source apportionment and health economic burdens in Pearl River Delta. *npj Clean Air, 2*, 16. https://doi.org/10.1038/s44407-026-00055-8

El Mghouchi, Y., & Udristioiu, M. T. (2026). Causal analysis of ground-level ozone formation using advanced AI-based methods in Craiova, Romania. *European Physical Journal Plus, 141*, 277. https://doi.org/10.1140/epjp/s13360-026-07495-x

Huang, C., Wang, J., Jin, Y., Min, M., Fan, Q., & Kim, J. (2026). O3-NOx-VOCs sensitivity in major Chinese regions: Detailed insights from GEMS satellite hourly observations. *Atmospheric Chemistry and Physics, 26*, 10221-10240. https://doi.org/10.5194/acp-26-10221-2026

Huang, H., Chen, T. K., Kamalanathan, S., Ehsan, R. M., Subasre, R., Liang, Z. F., & Luo, Y. S. (2025). Spatiotemporal analysis of urban air pollution dynamics: Linking emission sources, meteorological influences, and public health risks through long-term monitoring. *Journal of Environmental Sciences*, 10559-10587.

Kurniawan, T. A., Priya, A. K., Lei, T. M. T., Dissanayake, K. K., Onn, C. W., Jumaniyozov, K., Eshmetov, R., & Mohyuddin, A. (2026). Urban air pollution as a driver of climate forcing in Central and South Asia: Causal evidence from Tashkent (Uzbekistan) and Lahore (Pakistan). *Research Square*. https://doi.org/10.21203/rs.3.rs-10432281/v1

Liu, Y., Gong, Y., Wang, N., Shi, G., Yang, F., Zhang, Y., & Lu, K. (2026). Photochemical mechanism-dependent ozone formation and precursor sensitivity under varying NOx conditions. *EGUsphere*. https://doi.org/10.5194/egusphere-2026-1786

Mitra, B., Hridoy, A.-E. E., Mahmud, K., Uddin, M. S., Talha, A., Das, N., Nath, S. K., Shafiullah, M., Rahman, S. M., & Rahman, M. M. (2024). Exploring spatial and temporal dynamics of Red Sea air quality through multivariate analysis, trajectories, and satellite observations. *Remote Sensing, 16*(2), 381. https://doi.org/10.3390/rs16020381

Robertson, M. L., Arellano, A. F., & Sorooshian, A. (2026). Multi-platform analysis of ozone-precursor relationships in a semi-arid urban environment: Insights from TEMPO, Pandora, and surface observations in Tucson, Arizona. *EGUsphere*. https://doi.org/10.5194/egusphere-2026-3397

Rybarczyk, Y., Zalakeviciute, R., & Ortiz-Prado, E. (2024). Causal effect of air pollution and meteorology on the COVID-19 pandemic: A convergent cross mapping approach. *Heliyon, 10*(1), e25134. https://doi.org/10.1016/j.heliyon.2024.e25134

Tam, B. C. M., Tang, S.-K., & Cardoso, A. (2026). A multi-aspect feature dependency screening for augmentation of derivative feature space boosting deep learning training: A case of air quality in subtropical areas. *IEEE Access, 14*, 15637-15658.

Waudby, C. M. (2025). *Beyond thunderstorm asthma - extending our understanding of the meteorology, pollen and asthma relationship* [Doctoral dissertation, University of New South Wales]. https://doi.org/10.26190/unsworks/31339

Yang, X., Wang, Z.-H., & Wang, C. (2026). Directed causal coupling between urban heat and air pollution in U.S. cities. *Environmental Research Letters, 21*(6), 064029. https://doi.org/10.1088/1748-9326/ae53fd

Yuan, L., Han, W., Meng, J., Wang, Y., Yu, H., & Li, W. (2025). Uncovering the impact of urban functional zones on air quality in China. *Atmospheric Chemistry and Physics, 25*, 10421-10442. https://doi.org/10.5194/acp-25-10421-2025