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.

Clustering Daily Ozone Cycles as Comparable Regimes Across Stations, Years, and Independent Studies: The Querétaro Metropolitan Area, Mexico

Aldo Ledesma-Durán

aldo_ledesma@ciencias.unam.mx

Universidad Nacional Autónoma de México

Dara Salcedo

Instituto de Ciencias de la Atmósfera y Cambio Climático

Research Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11472208/v1

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

Additional Declarations: No competing interests reported.

Clustering Daily Ozone Cycles as Comparable Regimes Across Stations, Years, and Independent Studies: The Querétaro Metropolitan Area, Mexico

Aldo Ledesma-Durán[1*] and Dara Salcedo[1,2]

[1*]Unidad Multidisciplinaria de Docencia e Investigación (UMDI), Facultad de Ciencias, Universidad Nacional Autónoma de México (UNAM), Blvd. Juriquilla 3001, Santiago de Querétaro, 76230, Querétaro, Mexico.

[2]Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad de México, 04510, CDMX, Mexico.

*Corresponding author(s). E-mail(s): aldo_ledesma@ciencias.unam.mx;

Abstract

Ozone pollution in the Querétaro Metropolitan Area, a rapidly growing but understudied region of central Mexico, is monitored through datasets that differ in spatial coverage, measurement period, and instrumentation, limiting direct comparison across stations, years, and independent studies. We evaluated whether daily ozone cycles could be represented by discrete regimes providing a common basis for such comparisons. K-means clustering was applied to two non-overlapping datasets: an intensive single-site campaign at Juriquilla (2015–2016) and a multiyear, five-station operational network (2021–2024). A two-regime partition, distinguishing high- and low-ozone daily cycles, was the most robust solution across methods and datasets; an exploratory five-regime partition added a wind-ventilation signature linked to reduced precursor accumulation and a flatter diurnal cycle. Regime labels showed substantial cross-station agreement (Cohen's kappa 0.47–0.62) and changed measurably over 2021–2024 at four of five stations, mostly toward more frequent hot, dry, high-ozone regimes. Cross-scenario comparisons showed that independently trained models identified corresponding regimes not only in ozone-cycle shape but also in meteorological and precursor characteristics not used for clustering, a correspondence also evident in regime-transition dynamics; however, cross-model similarity showed no clear association with longer-term trends in the underlying variables, so regime correspondence should not substitute for direct trend analysis. These results show that regime-based representation offers a common framework for comparing heterogeneous ozone datasets across stations, years, and independent campaigns, preserving site-specific and temporal differences while enabling comparison between routine monitoring and chemically richer campaigns – an approach whose extension to other regions should be tested, not assumed.

Keywords: Tropospheric ozone, K-Means clustering, Cross-scenario comparability, Categorical regime labeling

Introduction

The Metropolitan Area of Querétaro (MAQ, ≈1.5 million inhabitants) is a rapidly growing but comparatively understudied region in central Mexico's Bajío, roughly 200 km north of Mexico City. Its semi-arid climate, intense sunlight, and weak ventilation, combined with traffic, industrial emissions (including from the nearby Salamanca refinery), and agriculture, shape a distinctive atmospheric environment. Official monitoring of O_3, NO_2, CO, SO_2, $PM_{2.5}$, and PM_{10} across five stations Secretaría de Medio Ambiente y Recursos Naturales (2012, 2019) shows marked recent deterioration: poor-air-quality days rose nearly 83% in 2024 relative to 2023, driven by particulate matter and ozone Secretaría de Desarrollo Sustentable (SEDESU) (2024). Ozone rises in the warm season with solar radiation and precursor emissions (NO_x, VOCs) Leighton (1970); $PM_{2.5}$ and PM_{10} dominate the dry season (January–May), from mobile sources,

agricultural burning, and resuspended dust CeMCAQ (2025); Secretaría de Desarrollo Sustentable (SEDESU) (2024).

Querétaro's atmospheric chemistry has been characterized from several angles Castañeda-Miranda et al. (2014); Rozanes-Valenzuela et al. (2023); Rodríguez-Trejo et al. (2024); Liñán-Abanto et al. (2019). The most detailed characterization comes from an intensive campaign at Juriquilla (March 2015–February 2016) using an Aerosol Chemical Speciation Monitor with Positive Matrix Factorization (ACSM/PMF) to resolve organic, sulfate, nitrate, ammonium, chloride, and black-carbon aerosol components Salcedo et al. (2023). The study found that the secondary fraction of $PM_{2.5}$ was the largest of the three, followed by the resuspended and primary fractions, consistent with an earlier study at the same site Olivares Salazar et al. (2021); both studies emphasize the importance of regional sources and processing in the atmospheric chemistry of air pollution at MAQ. The campaign described in Salcedo et al. (2023) is one of the two datasets analyzed here, referred to throughout as Scenario A. A more recent study at the same site reported the first nitroaromatic compounds in this environment and recurring biomass-burning influence Olivares-Salazar et al. (2025). Together, these studies show that Querétaro's photochemistry is far from generic, yet recurrent daily ozone behavior has not been systematically characterized. The comparison across stations, years, or independent studies, in any urban area, is difficult, and the problem is not specific to Querétaro.

A continuously operated network running across many stations and years, and an occasional, more detailed research campaign confined to a single site and a short period, rarely overlap in time, instrumentation, or even the variables measured. This study uses two such non-overlapping Querétaro datasets as a concrete case of that general problem: from the Juriquilla campaign above (Scenario A, 2015–2016, 22 chemical and meteorological variables) and from the operational network of the Centro de Monitoreo de Calidad del Aire de Querétaro (CeMCAQ), five stations recording the standard set of criteria pollutants from 2021 to 2024 (Scenario B, 9–12 variables). Scenario A allows regimes to be interpreted mechanistically in terms of photochemical activity, precursor environments, aerosol loading, and boundary-layer processes, whereas Scenario B trades this chemical depth for the broader spatial and temporal coverage typical of operational urban air-quality monitoring.

Clustering daily ozone cycles provides a natural framework for addressing this problem Boleti et al. (2020); Lyapina et al. (2016); Wang et al. (2022); Zhu et al. (2024); Tamas et al. (2016); Soares et al. (2018); Alahamade et al. (2021). Existing applications of clustering to ozone time series have compared regimes across stations Lyapina et al. (2016); Boleti et al. (2020); Gouveia et al. (2015) and, less commonly, across years or aggregation periods Nelson and Drysdale (2025). Almost none ask whether a *regime label* – the discrete category a trained clustering model assigns to a single day's cycle – produced by one model is comparable to a label produced by another, independently trained model on a different dataset. That is precisely the situation in a network like Querétaro's: five stations that began operating on different dates, so any comparison across stations or years must contend with unequal record lengths and no shared calibration period.

The advantage of a categorical label over a continuous similarity measure is structural. Measures such as functional principal component analysis (PCA) or dynamic time warping (DTW) require a coordinate system computed jointly across the datasets being compared – in practice, the same variables, measured the same way. Scenario A and Scenario B share no such space: one is chemically comprehensive, the other ozone-primary. A regime label sidesteps this: once a clustering model is trained, it converts any new day into a discrete symbol without recomputing anything relative to a comparison target, so comparing two datasets reduces to comparing label frequencies – a problem with standard statistical tools (chi-squared, Cohen's kappa, Jensen–Shannon divergence (JS)), not one that requires a bespoke distance metric for every new pair of datasets.

This study analyzes daily ozone cycles in the MAQ through clustering of their full 24-hour evolution, with two complementary objectives. The first is diagnostic: characterizing recurrent daily atmospheric behaviors and their chemical, meteorological, seasonal, and spatial signatures, and identifying what conditions produce a given shape of the ozone cycle. The second, and the central methodological contribution of this study, is comparability: every day–station ozone cycle is reduced to a single regime label within a defined clustering solution, and that label supports direct comparison across monitoring stations, across years, and even across independently trained clustering solutions fitted to non-overlapping datasets – three axes of space, time, and model – without overlapping records, shared instrumentation, or a bespoke similarity metric for each new comparison. The approach is directly relevant to rapidly growing, resource-limited monitoring networks across the Bajío and comparable regions, where operational networks and intensive campaigns rarely coincide in time or design.

Data and Methods

Study areas and instrumentation

This study draws on two non-overlapping ozone monitoring datasets from Querétaro, Mexico, referred to

throughout as Scenario A and Scenario B. The two scenarios differ in geographic coverage, campaign timing, instrumentation, and the number of variables available for interpretation, summarized in Fig. 1 and Table 1 and detailed below.

Scenario A is an intensive measurement campaign conducted at a single site in Juriquilla, Querétaro, from 1 March 2015 to 29 February 2016, yielding a complete hourly record (8784 hours, 366 daily ozone cycles). Instrumentation details are described in Salcedo et al. (2023), including the aerosol mass spectrometer used to derive the AMS/PMF aerosol tracers characterized in this study. In addition to ozone, the site measured temperature, relative humidity, solar radiation, wind speed, wind direction, and precipitation, together with the gas-phase and aerosol species NO, NO_2, CO, SO_2, PM_{10}, $PM_{2.5}$, black carbon, and the AMS/PMF tracers SO_4, NO_3, NH_4, Chl, Hydrocarbon-like Organic Aerosol (HOA), Low-Volatility Oxygenated Organic Aerosol (LVOOA), Semi-Volatile Oxygenated Organic Aerosol (SVOOA), and m/z 60, for a total of 22 variables.

Scenario B is the operational network of CeMCAQ, comprising five stations within the municipalities of Corregidora and Querétaro: Corregidora (COR), Félix Osores Sotomayor (FEO), Epigmenio González (EPG), Josefa Vergara y Hernández (JOV), and Carrillo Puerto (CAP). The stations came online on different dates – CAP and COR on 1 January 2021, FEO on 3 February 2021, and EPG and JOV on 1 July 2021 – and all recorded through 31 August 2024, with complete hourly ozone coverage from each station's commissioning date onward. A common core of 9 variables is available at all five stations: ozone, temperature, relative humidity, wind speed, wind direction, NO_2, CO, SO_2, and barometric pressure. Three additional variables bring the Scenario B variable set to 12, but are measured only at a subset of stations – $PM_{2.5}$ at all but CAP, PM_{10} at CAP and FEO only, and precipitation at CAP and FEO only – and are therefore excluded from cross-station comparisons but retained where a single-station analysis is possible. None of the aerosol-phase species available in Scenario A are measured in Scenario B.

Clustering procedure and normalization

Daily ozone cycles were represented as 25-hour vectors (00:00–24:00), where the 24:00 value closes the diurnal cycle and is set equal to the following day's 00:00 value for the same station. Each scenario's full matrix of ozone values (all days and hours pooled) was then scaled to the [0, 1] range using that scenario's own single minimum and maximum, prior to clustering; the two scenarios were scaled independently of each other, and scaling was not performed

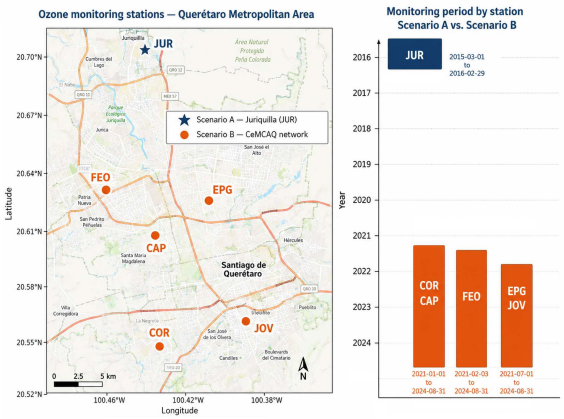


Fig. 1: Location of the monitoring sites and campaigns used in this study. Scenario A corresponds to the Juriquilla site, whereas Scenario B comprises the five CeMCAQ monitoring stations distributed across the Metropolitan Area of Querétaro (MAQ); the main urban road network is shown for geographical reference. The accompanying timeline shows each station's monitoring period, highlighting the temporal gap and staggered commissioning dates between Scenario A and Scenario B.

per day or per station. Clustering was performed exclusively on this 25-hour ozone cycle; all ancillary chemical and meteorological variables were used only afterward, to characterize the resulting clusters, and never entered the clustering algorithm itself. Clustering was performed using K-Means MacQueen (1967), with identical configuration for Scenario A and Scenario B; parameter and reproducibility details are given in Supplementary Material A.

The number of clusters was determined from multiple independent criteria. K-Means, agglomerative clustering with Ward linkage, and Self-Organizing Maps (SOM) were evaluated using the silhouette score, Calinski–Harabasz index, and Davies–Bouldin index, which assess within-cluster cohesion and between-cluster separation. These criteria consistently supported $k = 2$ as the optimal solution in both scenarios. Across the alternative solutions, $k = 5$ provided the strongest secondary support in both scenarios, driven primarily by the K-Means inertia elbow criterion, although agreement among the other methods was comparatively weak. This pattern is consistent with the tendency of internal validation indices to favor parsimonious solutions when continuous data exhibit weak or diffuse cluster structure. The Hopkins statistic confirmed genuine clustering structure in both scenarios. The Gap statistic supported $k = 2$ for Scenario A and, more conservatively, $k = 1$ for Scenario B, with only a narrow margin. Full validation results, method agreement, and elbow- and Gap-statistic comparisons are reported in Supplementary Material A.

Scenario	Code	Municipality / delegation	Latitude	Longitude	Altitude (m a.s.l.)
A	JUR	Juriquilla	20.7034	−100.4474	1900
B	COR	Corregidora	20.5497	−100.4428	1814
	FEO	Félix Osores Sotomayor	20.6382	−100.4642	1874
	EPG	Epigmenio González	20.6301	−100.4097	1890
	JOV	Josefa Vergara y Hernández	20.5686	−100.3904	1880
	CAP	Felipe Carrillo Puerto	20.6101	−100.4376	1805

Table 1: Geographic location of Scenario A and Scenario B monitoring sites. Altitude values are Open-Elevation estimates, not direct measurements.

Thus, $k = 2$ provides the most statistically robust and interpretable partition for both scenarios, corresponding to a high-ozone/low-ozone distinction developed in the following section. However, the recurrence of $k = 5$ across both scenarios, despite weaker statistical support, provides a useful exploratory resolution of the data. It reveals differences in the level and shape of the diurnal ozone cycle that are not apparent at $k = 2$, without implying that five represents a privileged or "correct" number of regimes. Section 3 reports both partitions; Section 4, which focuses on regime-based comparability across space, time, and model, adopts $k = 5$ because its finer resolution better serves that goal, while $k = 2$ remains appropriate for analyses concerned only with a high/low-ozone distinction.

Results

Results for both scenarios are presented under two partitions, $k = 2$ and $k = 5$. For each scenario, the $k = 2$ partition is presented first, followed by the finer $k = 5$ resolution. Their implications for station-level and cross-year comparability are examined in Section 4.

Scenario A

Primary regime structure ($k = 2$)

The primary, statistically robust partition of Scenario A separates daily ozone cycles into two regimes, corresponding to low- and high-ozone days. To assess whether individual variables differed between clusters, we used one-way analysis of variance (ANOVA), which compares between-cluster and within-cluster variability through the F-statistic; larger F values indicate stronger separation. Because multiple variables were tested simultaneously, p-values were adjusted using the Benjamini–Hochberg (BH) procedure to control the expected proportion of false discoveries among significant results Benjamini and Hochberg (1995). Of the 22 variables tested, 20 differed significantly between the two clusters after BH correction (all but wind speed and CO); ozone concentration showed the strongest discrimination ($F = 551.7$, based on its daily mean), followed by solar radiation ($F = 114.0$). Figure 2 shows the mean diurnal ozone cycle for the two primary regimes, illustrating the low- and high-ozone character of the two clusters described above.

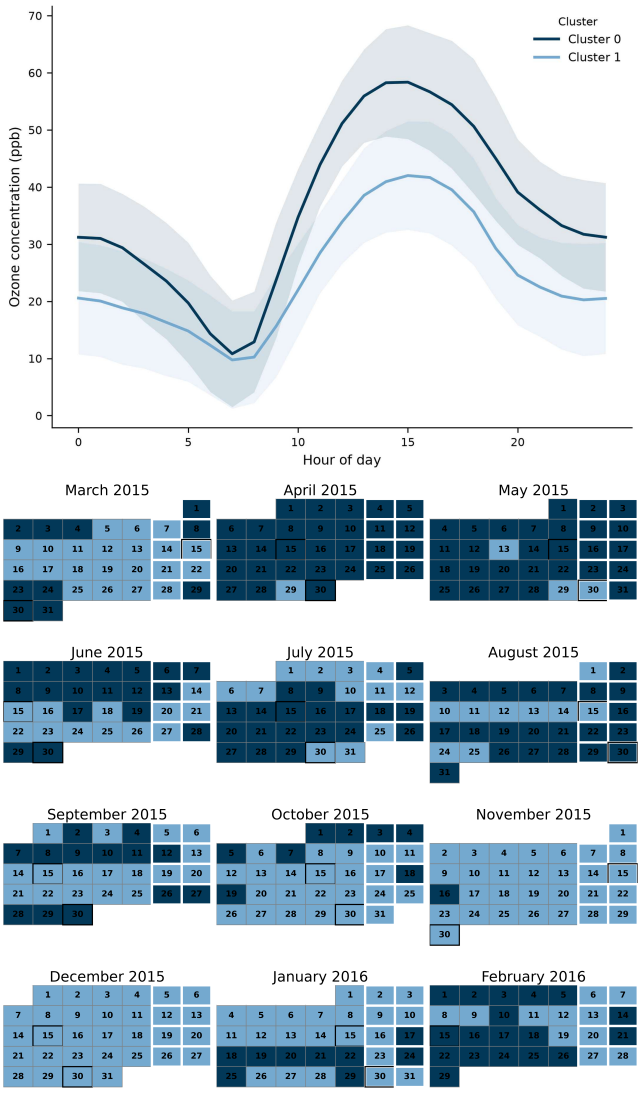


Fig. 2: Mean diurnal ozone cycle (±1 SD) and calendar heatmap of daily cluster assignment for the primary two-cluster partition ($k = 2$), Scenario A.

Ozone was negatively correlated with NO and NO_2 (Pearson $r = -0.377$ and -0.223 respectively;

both correlations remain significant after BH correction for multiple comparisons, applied throughout this manuscript unless noted otherwise), consistent with titration of ozone by freshly emitted NO_x (exact statistics, together with NO_3 and wind speed, which showed no significant correlation with ozone, are given in Supplementary Material B). Precipitation differed significantly between the two clusters (one-way ANOVA). Figure 3 makes this mechanism directly visible: NO peaks sharply during the morning rush hour, coinciding with the daily ozone minimum, and then declines as solar radiation rises through the morning while ozone climbs toward its afternoon peak.

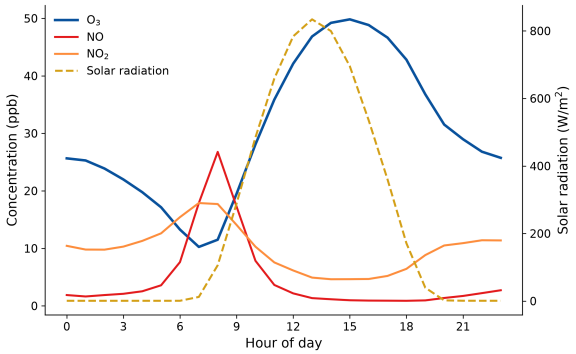


Fig. 3: Mean diurnal cycles of O_3, NO, NO_2, and solar radiation for Scenario A (full 366-day record, not split by cluster; complete hourly coverage with no data gaps, see Section 2.1).

Cluster occurrence showed clear seasonal and calendar structure (Figure 2). Cluster 1 (low-ozone) occurred more frequently between October and January, Querétaro's dry-cold season, consistent with reduced photochemical activity under lower solar radiation and lower temperature during this period.

We also defined an operational criterion to identify days characterized by high ozone and enhanced secondary photochemical processing. A day was classified as meeting this criterion when its daily maximum O_3 was above the 75th percentile, its daily mean NO_x (NO + NO_2) was below the 25th percentile, and either LVOOA or SO_4 was above the 75th percentile. This multivariate criterion combines high ozone, reduced NO_x titration, and evidence of secondary aerosol formation, providing a more stringent definition than ozone concentration alone. Under this criterion, qualifying days were rare (1.6% of the 366-day record) and fell, without exception, in the high-ozone cluster. No literature-derived benchmark exists for this specific combination of criteria, so it is reported as an explicit operational definition, not an established standard.

Secondary exploratory layer ($k = 5$)

The five-cluster partition reveals within-regime atmospheric and chemical structure not visible at $k = 2$.

The five-cluster solution was compared with the $k = 2$ partition to assess how much of the finer structure corresponds to the broader high-ozone/low-ozone separation. The two partitions show limited agreement: three $k = 5$ clusters are nested within a single $k = 2$ cluster at $> 90\%$ membership (clusters 0, 1, and 4), whereas the remaining two show substantial mixing (cluster 3, 55.0%/45.0%; cluster 2, 80.2%/19.8%). Overall agreement is low (adjusted Rand index, ARI = 0.245; normalized mutual information, NMI = 0.345), indicating that the finer partition captures structure that is not simply a subdivision of the $k = 2$ solution. Precipitation, which distinguished clusters at $k = 2$, no longer does so at $k = 5$ (the adjusted p-value is $p_{\text{adj}} = 0.294$).

The main atmospheric differences among the five clusters are summarized in Table 2 and Figures 4 and 5. The corresponding detailed characterization is provided in Supplementary Material B, including cluster-mean values for the complete set of ancillary variables, the calendar of daily cluster assignments, diurnal cycles for all variables, and ozone–species phase-lag diagnostics. The following characterization therefore focuses on the principal atmospheric contrasts and their physical interpretation, while the supplementary analyses provide the broader observational basis for these interpretations.

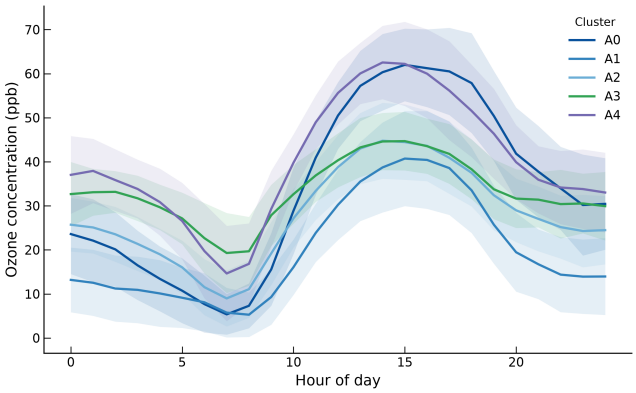


Fig. 4: Mean diurnal ozone cycle (±1 SD) for each of the five Scenario A clusters ($k = 5$).

Cluster 0 ($n = 67$) is the most heavily loaded of the five clusters in several respects at once: it combines high ozone (34.0 ppb) and the highest solar radiation of the five clusters with the driest conditions (lowest relative humidity, 51.3%) and the highest levels of NO_2, SO_2, and PM_{10} (Supplementary Figure B2), together with the highest secondary-aerosol load (NO_3, SVOOA, and black carbon all maximal; Supplementary Figure B3). This

Cluster	n	O_3 (ppb)	Temp (°C)	Rad (W/m^2)	RH (%)	WSpeed (m/s)	Wdir (°)	NO (ppb)	LVOOA (μg/m^3)
0	67	34.0	26.5	277.9	51.3	2.66	357	7.31	2.26
1	78	20.1	23.4	194.8	62.2	2.62	51	8.29	1.57
2	101	28.0	24.2	236.4	62.9	2.94	55	3.19	1.44
3	60	33.4	22.8	232.4	63.9	3.17	62	1.63	1.51
4	60	40.4	27.0	275.6	54.4	2.71	57	2.53	2.39

Table 2: Key atmospheric variables by cluster, Scenario A ($k = 5$). All variables show the cluster daily mean; Wdir is the circular mean wind direction (degrees).

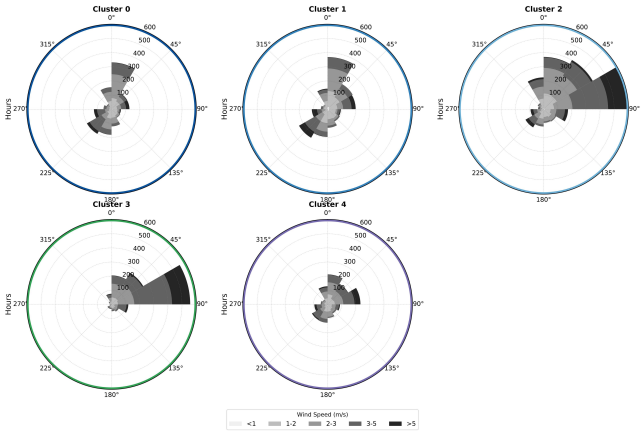


Fig. 5: Wind-direction and wind-speed distribution for each of the five Scenario A clusters ($k = 5$). Except for cluster 0's distinct northerly/bimodal regime, the remaining four clusters share a similar NE-ENE mean wind direction and differ primarily in how concentrated that flow is.

combination of elevated primary precursors alongside maximal secondary-aerosol markers is consistent with a dry, high-insolation regime in which substantial photochemical processing accompanies rather than replaces primary-pollutant accumulation. Its wind pattern is bimodal – calm-to-light flow from the north together with a secondary gusty sector from the southwest – consistent with the late-peaking, fast-rising ozone cycle shown in Fig. 4 and analyzed in depth in Fig. B7, along with other shape descriptors included in the Supplementary Material.

Cluster 1 ($n = 78$) is the least photochemically active: lowest ozone, lowest temperature and solar radiation, and highest relative humidity, yet the highest NO and NO_2 of the five clusters – more precursor available but less radiation to convert it, directly consistent with a radiation-mediated mechanism. It also shows the highest CO and the highest primary organic aerosol markers (HOA, m/z 60) of the five clusters, in agreement with the accumulation of fresh combustion emissions under low photochemical turnover – a pattern that matches this cluster's concentration in November–December (Calendar in Fig. B8 of the Supplementary Material). Its wind pattern is comparatively weakly organized, with elevated high-speed fractions concentrated in the SW and WSW sectors, suggesting weaker or less consistent ventilation.

Cluster 2 ($n = 101$, the largest) shows no extreme in any atmospheric variable and a strongly unidirectional wind from the ENE, making it the closest approximation to a typical campaign day.

Cluster 3 ($n = 60$) – the structurally atypical cluster examined in Section C.1 – shows an even more strongly unidirectional wind regime than cluster 2, from a near-identical ENE sector, with the highest daily-mean wind speed of the five clusters – differing from every other cluster by a wide margin (one-way ANOVA, $F = 15.11$, $p_{\text{adj}} = 2.5 \times 10^{-11}$; see Fig. 5 for the full wind-speed-bin breakdown). This wind regime coincides with the lowest levels of most measured precursor and aerosol species – NO, NO_2, CO, PM_{10}, $PM_{2.5}$, black carbon, NO_3, Chl, HOA, m/z 60, and SVOOA (Supplementary Figures B2 and B3) – consistent with wind-driven dispersal limiting local pollutant accumulation.

Cluster 4 ($n = 60$) shows the highest ozone, temperature, and radiation of the five clusters, together with the highest SO_4 and PM2.5, consistent with the campaign's days of maximal photochemical loading; SO_2, in contrast, is highest in cluster 0 (above), reflecting its greater primary-pollutant loading. Its wind pattern is unremarkable in both speed and direction – a moderate wind speed (2.71 m/s) from a moderately concentrated NE-ENE sector, similar in mean direction to clusters 1-3 but without the strong unidirectionality of clusters 2 and 3.

For each cluster, the lag between ozone and every other variable was estimated by cross-correlating their mean diurnal cycles and taking the hour of maximum correlation (full method in Supplementary Material B); this identifies relative phase, not a causal delay. Across all five clusters, ozone lags NO, CO, and black carbon by several hours (7–10 h for NO and CO) and leads PM_{10} by 6–9 h, while temperature and radiation are essentially synchronous with ozone (lag 0–2 h). Because this ordering is shared across clusters rather than distinguishing among them, we treat it as a common cross-cluster pattern. The observed phase relationships are also compatible with a nighttime NO_3-radical oxidation pathway reported for this site

Olivares Salazar et al. (2021); Salcedo et al. (2018), although the lag analysis alone does not establish this mechanism. Phase-lag Figures B5 and B6, along with the daily cycle of all variables for comparison B4, are presented in the supplementary material.

Two additional patterns emerge from the $k = 5$ partition. First, the apparent monotonic relationship between ozone and regional photochemical aging breaks down in cluster 3, essentially indistinguishable from cluster 2 in LVOOA (Table 2) despite higher ozone. Second, several pairs of clusters show statistically equivalent NO while differing significantly in both solar radiation and ozone – a recurring pattern consistent with a radiation-mediated photochemical mechanism rather than a precursor-driven one (full pairwise statistics in Supplementary Material B).

Taken together, these observations support for Scenario A a near-mechanistic interpretation of the five regimes, linking ozone-cycle behavior to meteorology, precursor availability, radiation, and aerosol composition. This interpretation remains observational and correlational and does not substitute for targeted experimental or process-based evidence.

Scenario B

Primary regime structure ($k = 2$)

As in Scenario A, the primary partition of Scenario B is a high-/low-ozone split: cluster 1 (high ozone) is predominantly observed between late March and mid-June at the five stations, whereas cluster 0 (low ozone) covers the rest of the year (calendars not shown). Figure 6 shows the corresponding mean diurnal ozone cycles.

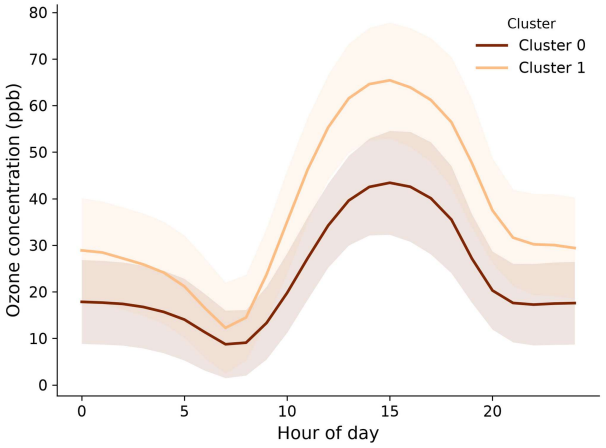


Fig. 6: Mean diurnal ozone cycle (±1 SD) for the primary partition into two clusters ($k = 2$), Scenario B.

As in Scenario A, temperature was the most discriminating variable ($F = 3044$), followed by relative humidity and $PM_{2.5}$; precipitation was the weakest, though still significant, discriminator ($F = 4.6$) among the ten variables tested.

At $k = 2$, station-level cluster shares were close to the network-wide average: pooling all five stations, 64.0% of station-days fell in cluster 0 (low-ozone) and 36.0% in cluster 1 (high-ozone). Each station's cluster-0 share was close to the 64.0% network-wide value, with CAP showing the largest deviation (−5.18 percentage points). Thus, the $k = 2$ partition in Scenario B captures a clear high/low-ozone distinction but provides limited station-level differentiation, motivating the finer $k = 5$ resolution examined next.

Secondary exploratory layer ($k = 5$)

The $k = 5$ partition was examined as an exploratory refinement of the operational network, independently of the $k = 2$ solution. Three of the five clusters nest cleanly ($> 90\%$) within a single $k = 2$ cluster, whereas the remaining two are split across both groups (cluster 2: 69.8%/30.2%, discussed below; cluster 3: 8.6%/91.4%).

Seasonally, the clusters show distinct temporal patterns (numbering as in Table 3): cluster 0 concentrates in winter (October–March, 82.5% of its days); cluster 1 in May (41.0%, with April and June as shoulders); cluster 2 spans June–September (peaking in August at 16.0%) and is the only cluster over-represented on weekends and Mondays; cluster 3 concentrates in spring (March–May, 46.6%); and cluster 4 shows a bimodal July/December pattern (12.2% and 12.5%, respectively), separated by a January–May trough. The full station-by-cluster distribution and annual calendar are shown in Table 3 and Figure 7.

Station-level structure was more pronounced at $k = 5$. The most frequent cluster at each station reveals a 3+1+1 pattern: EPG (34.5%), FEO (26.6%), and JOV (30.6%) are dominated by cluster 2, whereas CAP is dominated by cluster 0 (28.3%) and COR by cluster 4 (27.9%). Thus, no single regime dominates all five stations, and the finer partition reveals a three-station block together with two station-specific patterns. Within the three-station block, EPG deviates most from the network-wide cluster-2 share (25.9%; +8.67 percentage points; Table 3). This grouping may reflect differences in surrounding land use: EPG, FEO, and JOV are located within the denser urban core, whereas CAP lies near an industrial zone and COR in an outlying municipality (Table 1); this association is not tested formally here, and the station-level differences it might explain are discussed further in Section 4.

The five Scenario B clusters differ in temperature, humidity, wind, precursor levels, and ozone–precursor phase relationships (Supplementary Material C, Fig. 8, and Table 4). ANOVA identified

Station / Cluster	0	1	2	3	4
CAP	28.3	14.0	19.3	22.7	15.6
COR	24.3	7.5	20.2	20.2	27.9
EPG	20.3	11.5	34.5	17.0	16.7
FEO	22.1	13.2	26.6	15.8	22.4
JOV	22.7	9.2	30.6	16.8	20.7

Table 3: Cluster occurrence by station, Scenario B, $k = 5$ (percentage of station-days).

temperature as the most discriminating variable ($F = 1042$), followed by NO_2 and relative humidity; precipitation was the weakest discriminator ($F = 9.3$) among the ten variables tested.

Wind direction does not split cleanly into consistent and inconsistent clusters: cluster 2 is the only cluster whose dominant sector (60–90°, 25.4%) is not shared by any other, while clusters 0 and 3 share a dominant sector (210–240°) and clusters 1 and 4 share another (90–120°). No cluster shows full agreement across all five stations; cluster 2, the closest to a shared pattern, in fact reflects a station-block split (CAP, COR, and FEO favoring 90–120°; EPG and JOV favoring 60–90°) rather than individual-station consistency, with inter-station variability the norm across the network rather than the exception (full station-by-station wind breakdown in Supplementary Material, Figure C9). Wind speed, by contrast, varies far more by station than by regime: across all five clusters, the same ranking holds – EPG and JOV are consistently the windiest stations, CAP and COR sit in an intermediate, roughly comparable range, and FEO is consistently the calmest by a wide margin, roughly threefold slower (exact values in Figure C9). This gap holds regardless of which regime a given day is assigned to, indicating an apparent siting effect – presumably reflecting local exposure or topography at each station – rather than a property of the regimes themselves.

Cluster 0 shows the highest NO_2, CO, and SO_2 of the five clusters, together with the weakest wind of the five – consistent with a stagnant regime favoring accumulation of primary pollutants. Cluster 1, the smallest of the five by sample size ($n = 16{,}800$ station-hours), is the hottest and driest cluster and shows by far the highest ozone of the five – a low-humidity, high-photochemical-activity regime consistent with this region's meteorology. Cluster 2, the largest by sample size ($n = 39{,}120$ station-hours), shows the lowest NO_2 and CO of the five clusters together with a strongly unidirectional easterly wind, consistent with its role as the network's largest and most typical regime. It also carries the highest wind speed in the network ($F = 158.68$, $p = 8.0 \times 10^{-129}$), combined with the lowest precursor and particulate load among the five clusters – the same wind-driven dispersal signature independently identified in Scenario A's cluster 3 (Section 5.1), examined

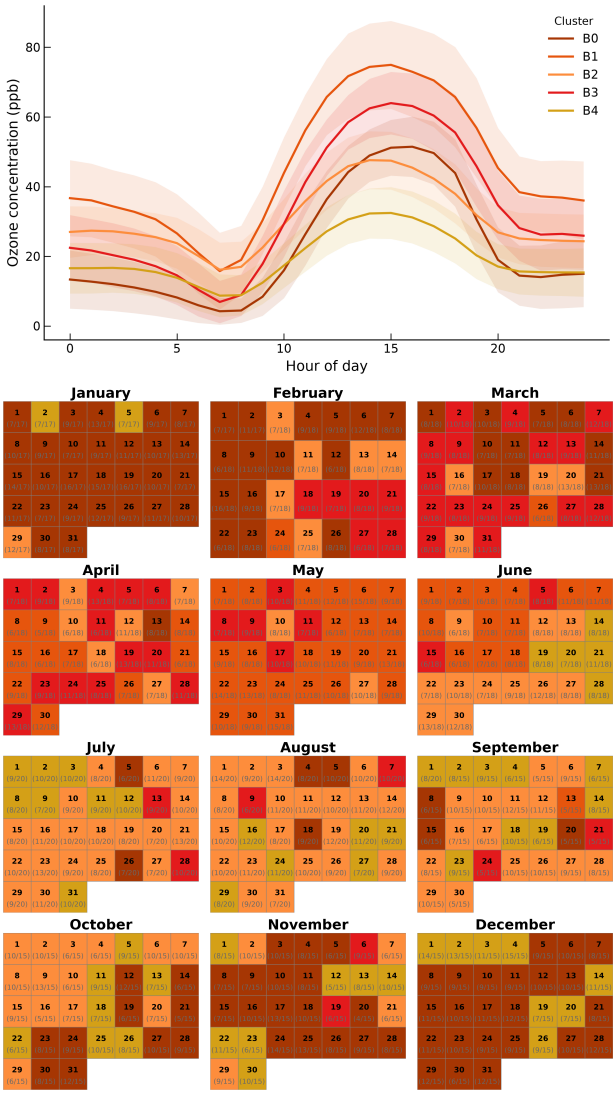


Fig. 7: Mean diurnal ozone cycle (±1 SD) and annual calendar of the daily modal cluster across the five monitoring stations, Scenario B ($k = 5$). For each calendar day, assignments are pooled across stations and years (2021–2024; typically 15–20 observations); the cell colour indicates the modal cluster and the fraction in parentheses its fraction of the pooled observations.

further in Supplementary Material C.1. Cluster 3 is warm and dry with intermediate-to-high precursor levels. Cluster 4 shows the lowest SO_2, PM_{10}, and $PM_{2.5}$ of the five clusters together with nearly double the rainfall of any other cluster, the lowest ozone, and the most humid conditions; unlike clusters 0 and 2, its dominant wind sector is not the same across all five stations, reflecting station-to-station heterogeneity rather than a within-station unstable signal. Full atmospheric detail, including exact values and station-level wind breakdown, is given in Supplementary Material C.

Cluster	n (h)	O_3 (ppb)	Temp (°C)	RH (%)	WSpeed (m/s)	NO_2 (ppb)	Precip. (mm)	Wdir (dominant)
0	35,746	22.9	18.5	50.0	0.34	17.1	0.034	210–240° (14.8%)
1	16,800	45.6	24.3	40.9	0.48	11.4	0.026	90–120° (19.3%)
2	39,120	30.8	20.2	56.6	0.91	8.9	0.054	60–90° (25.4%)
3	28,128	33.6	21.8	45.3	0.38	13.6	0.045	210–240° (15.1%)
4	31,392	19.5	18.4	64.1	0.85	10.3	0.107	90–120° (23.8%)

Table 4: Key atmospheric variables by cluster, Scenario B ($k = 5$). All variables show the cluster mean, with the single exception of WSpeed, reported as the cluster mean of the daily minimum wind speed, given its direct mechanistic interpretation as a floor on sustained ventilation. Wdir (dominant): most frequent 30° wind-direction bin and its share of cluster-hours; each station-hour is treated as an independent observation, i.e. simultaneous readings from different stations are not averaged into a single regional value before computing the mode. n is given in station-hours, from the same station-level assignment.

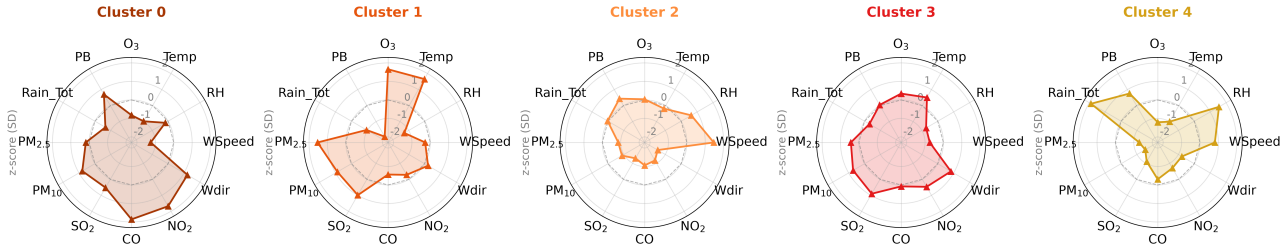


Fig. 8: Atmospheric characterization of the five Scenario B clusters ($k = 5$): z-scores of cluster means relative to the five-cluster mean, for the same twelve variables shown in Figures C10 and C11, presented jointly as a radar for direct cross-variable comparison within each cluster.

In addition, the NO_2–ozone lag structure divides the five clusters into two groups (Figure C13): clusters 1 and 2 show a net NO_2–ozone lag of 7–8 hours *before* the ozone peak, while clusters 0, 3, and 4 show a net lag of 7–9 hours *after* it. NO_2 follows a bimodal diurnal cycle, with a morning and an evening peak, in all five clusters; because the lag reported here is the single hour of maximum cross-correlation with the (unimodal) ozone cycle, this split more likely reflects *which of the two NO_2 peaks is comparatively larger* in each cluster than a genuine precursor-to-product delay that differs by regime. We therefore report it here as a descriptive, within-network contrast among the five clusters rather than as evidence of a different precursor-product ordering. Supplementary Material C gives the complete atmospheric characterization behind this summary: per-variable diurnal cycles and normalized cluster means for all twelve variables, together with the full station-by-station wind-direction and NO_2–ozone lag breakdowns.

This mechanistic characterization of the five Scenario B clusters is necessarily less detailed than that of Scenario A, since the operational network measures a smaller set of chemical and meteorological variables. Its value lies elsewhere: spanning five stations over four years gives it a spatial and temporal breadth that the single-site, one-year Scenario A campaign cannot offer. Whether – and how – this coarser but broader regime structure can be compared meaningfully against Scenario A's finer one is examined next, in Section 4.

Regime comparability: a practical measure across space, time, and independently trained models

A categorical regime label is only useful for comparison if it means the same thing wherever it is applied. The most direct test of this is within Scenario B, where every station is measured by the same operational network, on the same instrumentation, and classified by a single trained model: Sections 4.2 and 4.3 use this shared framework to show that regime labels compare cleanly across space and time, since any difference in regime prevalence between stations or years reflects a genuine atmospheric difference rather than an artifact of how the labels were built. A harder problem remains, and it is the central methodological question of this study: whether the same kind of comparison can be made between two scenarios that share nothing – no stations, no years, no instrumentation, and independently trained clustering models, as is the case between Scenario A and Scenario B, where cluster indices carry no inherent correspondence to begin with (there is no statistical reason to expect, for instance, that Cluster 0 in Scenario A represents the same atmospheric regime as

Cluster 0 in Scenario B). Section 4.4 addresses exactly this: a way to quantify how closely the shape of the daily ozone cycle in one scenario's regimes resembles that of the other's, without assuming any correspondence between the two independently trained models.

Cluster identity reflects cycle shape, not only ozone level

A necessary condition for the regime-label interpretation is that cluster membership encodes information about the *shape* of the diurnal ozone cycle – how fast ozone rises and falls, when it peaks, how large the swing is between its daily minimum and maximum – and not merely the daily mean or maximum ozone level. Two extremes bound the choice of unit for comparing ozone regimes. At one end, reducing each cycle to a single level statistic discards exactly the shape information that gave the atmospheric characterization in Sections 3.2.2 and C.1 its interpretive power: comparing such labels across stations or years would add nothing beyond comparing daily ozone means directly. At the other end, comparing full raw diurnal time series preserves that information but is itself difficult to operationalize – it requires aligning cycles in time, is sensitive to instrument noise, and offers no natural discrete unit to tally and compare. A shape-aware cluster label sits between these extremes: coarse enough to reduce every cycle to a single, countable category, yet rich enough to retain the geometric information a level statistic alone would discard.

To assess whether shape adds information beyond level, we ran the following test. For each Scenario B day–station record with a well-defined rise and fall segment ($n = 6300$), we computed two level descriptors (daily mean and maximum ozone) and three shape descriptors capturing cycle geometry independently of level: relative amplitude, morning rise rate, and afternoon fall rate. A multinomial logistic classifier predicting the $k = 5$ cluster label performed substantially better using level and shape descriptors together than using either alone, closing roughly 40% of the classification error left by a level-only model (5-fold cross-validation; balanced-accuracy values in Supplementary Material D). Because level and shape descriptors are calculated from disjoint aspects of the same cycle, this gain indicates that shape carries information about cluster identity that is not recoverable from level alone; an analysis of covariance corroborates this from a different angle, with the cluster effect on each shape descriptor remaining significant after adjusting for daily mean ozone (full statistics in Supplementary Material D) – these shape features distinguish clusters beyond what daily mean ozone alone would predict.

Peak timing – the hour at which each day–station's ozone cycle reaches its maximum – fell outside the typical afternoon window in 106 of 6300 records (1.7%), disproportionately concentrated in cluster 4 (5.1% of its own records) and, to a lesser extent, cluster 2 (2.3%), versus $\leq 0.14\%$ in clusters 0, 1, and 3 – cluster 4 being the same cluster identified in Section C.1 as having the flattest diurnal amplitude among the five Scenario B regimes, see Fig. 7. The atypical timing is an expected consequence of cluster 4's flat diurnal structure: when a cycle shows little contrast between its daily minimum and maximum, small fluctuations can shift where the apparent peak falls, making peak hour an unreliable descriptor for that cluster rather than evidence of a data-quality problem (full diagnostics in Supplementary Material D).

This is what makes the regime label a meaningful unit of comparison in the sections that follow rather than a relabeling of the daily ozone mean. Figure 9 shows both sets of descriptors side by side: cluster B1 combines the highest ozone level with a below-average relative amplitude and the fastest morning rise, while B4 shows the opposite level pattern but expresses its flat diurnal structure through the peak-timing instability noted above rather than through relative amplitude.

Cross-station regime correspondence

A regionally shared ozone regime implies that two stations should fall into the same cluster on the same day more often than expected by chance; station-specific idiosyncrasies alone would not produce such cross-station agreement. We tested this across all 10 possible pairs of the five Scenario B stations, considering only calendar days on which both stations had data.

To assess this effect, two tests were applied to each station pair: Cohen's kappa Cohen (1960), which measures the extent to which same-day agreement exceeds that expected from each station's own cluster frequencies (i.e., agreement expected by chance); and a chi-squared test of independence on the 5×5 contingency table, to determine whether the two stations' cluster labels were statistically independent. A BH correction was applied across the 10 pairs to control for multiple testing. All 10 pairs passed both tests. Observed same-day agreement (58–70%) far exceeded the 20–23% expected by chance and remained significant after correction (Table 5). Kappa ranged from 0.470 (COR–FEO, the weakest pair) to 0.622 (COR–JOV, the strongest). According to the benchmarks of Landis and Koch (1977) (0.41–0.60, "moderate"; 0.61–0.80, "substantial"), most pairs fell within the moderate range, while the two strongest pairs (COR–JOV and EPG–JOV) reached the substantial range.

However, a high kappa could still mask an uneven pattern, with one cluster driving most of the agreement while the other four rarely coincide between

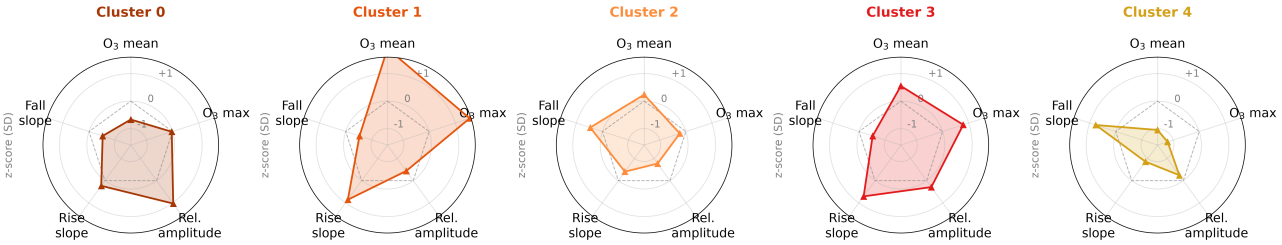


Fig. 9: Shape and level descriptors of the diurnal ozone cycle by cluster, Scenario B, $k = 5$ expressed as z-scores relative to the full population. The equivalent figure for Scenario A appears in Supplementary Material B.

Station pair	m	% Obs. agree. (% Exp.)	κ
COR–JOV	1158	70.4 (21.6)	0.622
EPG–JOV	1158	68.9 (22.5)	0.599
CAP–FEO	1306	67.1 (20.3)	0.587
CAP–EPG	1158	65.7 (20.4)	0.569
COR–EPG	1158	65.7 (21.3)	0.564
EPG–FEO	1158	65.2 (21.8)	0.555
CAP–COR	1339	63.2 (20.8)	0.535
CAP–JOV	1158	63.0 (20.6)	0.534
FEO–JOV	1158	61.7 (22.0)	0.509
COR–FEO	1306	58.2 (21.1)	0.470

Table 5: Cross-station same-day regime agreement, Scenario B, $k = 5$. m: number of overlapping calendar days for the pair; agreement column shows observed percentage with the chance-level expected percentage under independence in parentheses; κ: Cohen's kappa. chi-squared independence test, 16 d.f.).

stations. To assess this possibility, we examined the standardized residuals of the contingency tables for the best- and worst-agreement pairs only (COR–JOV and COR–FEO; the remaining eight pairs were not examined in this way). We tested whether the large positive residuals—cells with more same-day matches than expected under independence—were concentrated on the diagonal (matching clusters) or scattered across off-diagonal combinations. In both pairs, all large positive residuals occurred on the diagonal, whereas off-diagonal residuals were negative and small (exact ranges are provided in Supplementary Material D). Thus, agreement was distributed across all five regimes rather than being driven by a single cluster, at least for these two representative pairs.

In plain terms, on any given day, two Scenario B stations were far more likely to report the same ozone regime than chance alone would predict, and this pattern held for every station pair in the network, not only for nearby stations. Moreover, the association was not driven by a single easily identifiable regime, such as unusually high-ozone days: even in the pair with the weakest overall agreement, all five regimes showed same-day agreement across stations beyond that expected by chance. Ozone regimes in Scenario B therefore behaved as a regional condition, independently identified at each station, rather than as station-specific labels that happened to share a common classification.

If this regional signal follows the expected pattern of atmospheric transport, nearby stations sharing the same air mass more consistently should agree more than distant ones. This expectation received only partial, inconclusive support: kappa was negatively correlated with the great-circle distance between station pairs (Pearson $r = -0.63$, $p = 0.051$; Spearman $\rho = -0.56$, $p = 0.090$), but the relationship fell short of significance and was not straightforward—the closest pair (CAP–EPG, 3.7 km) did not have the highest kappa. Cross-station agreement is therefore better interpreted as a shared regional exposure signal than as a simple function of geographic proximity.

Cross-year regime stability and change

A complementary use of the same regime labels is to track changes in a station's regime composition over time, a different question from same-day cross-station agreement because there is no day-to-day pairing across periods. The five stations were commissioned on different dates in 2021 – CAP and COR in January, FEO in February, and EPG and JOV in July (Section 3.2.1) – while the record ends on 31 August 2024 at all stations. Consequently, no calendar year has complete and equal coverage across the network: 2021 is only partially covered at three stations, and 2024 is truncated to January–August at all five. Direct comparison of calendar years could therefore confound genuine interannual changes with seasonal sampling differences; for example, a partial year missing the cooler, wetter months could appear spuriously shifted toward the hotter, drier regime.

To avoid this, we instead defined three rolling twelve-month periods, anchored to the last available date and each covering a full annual cycle at every station:

- Period 1: 1 September 2021 – 31 August 2022
- Period 2: 1 September 2022 – 31 August 2023
- Period 3: 1 September 2023 – 31 August 2024

All five stations have complete daily coverage in all three periods ($n = 1096$ station-days each). For each station, we compared the distribution of cluster assignments across the three periods using a

chi-squared test of homogeneity and pairwise Jensen–Shannon (JS) divergence between consecutive periods.

Four of the five stations—CAP, COR, EPG, and FEO—rejected the null hypothesis of homogeneous cluster distributions across the three periods (Table 6); JOV did not ($p_{\text{adj}} = 0.491$), making it the only station for which no statistically detectable change in regime composition was found over the study period. Beyond significance, Cramér's V (Table 6) measures how strong a station's shift is, not just whether it is detectable. By this measure, CAP shows the strongest interannual restructuring in the network ($V = 0.190$), and EPG the weakest among the four significant stations ($V = 0.087$).

Station	n	χ^2	p_{adj}	V
CAP	1096	79.4	3.3×10^{-13}	0.190
COR	1096	53.5	1.5×10^{-8}	0.156
EPG	1096	16.7	0.042	0.087
FEO	1096	65.7	8.8×10^{-11}	0.173
JOV	1096	7.4	0.491	0.058

Table 6: Chi-squared test of homogeneity of cluster distribution ($k = 5$) across the three complete-coverage periods, Scenario B, by station. d.f. = 8 for all five stations. p_{adj}: BH-corrected across the five stations. V: Cramér's V (effect size).

Figure 10 shows what that shift looks like, with the exact per-period percentage for every cluster labeled directly on each bar. The general, interpretable pattern is: at CAP, COR, and FEO, the coolest/wettest regime (cluster 4) shrinks and the hottest/driest regime (cluster 1) grows from Period 1 to Period 3, consistently enough to be the dominant network trend; EPG and JOV do not follow it: EPG's shift, though statistically significant (Table 6), does not point in this direction — its cluster 4 share grows instead of shrinking — while JOV shows no significant shift at all, consistent with its non-significant omnibus test above. These results indicate sensitivity to interannual changes in regime composition at four stations, but do not imply that such changes are universal across the network; JOV provides a clear counter-example.

As a complementary analysis to the three-period test above, we also tested each pair of consecutive periods separately, yielding 10 station-period comparisons. Six were significant; the four exceptions were COR between Periods 1 and 2 ($p_{\text{adj}} = 0.088$), EPG between Periods 2 and 3 ($p_{\text{adj}} = 0.319$), and JOV in both consecutive pairs ($p_{\text{adj}} = 0.828$ and 0.147). Figure 11 summarizes the JS divergence for these period pairs, with pairwise significance indicated directly on the heatmap.

Cross-model comparability: independently trained clustering solutions

Cross-station and cross-year comparability (Sections 4.2, 4.3) compares regime labels produced by a *single* trained model. A more stringent test asks whether regimes identified by two *independently* trained solutions – based on different data sources, time periods, and preprocessing pipelines – can themselves be meaningfully related. Because each model assigns its own arbitrary label alphabet, we address this question in two steps. First, we test whether the two sets of clusters occupy comparable positions in a shared geometric space (Step 1). Second, we assess whether this geometric correspondence is consistent with the actual ordered, day-by-day structure of the data (Step 2). To illustrate this approach, we compare Scenario A with data from FEO station, the Scenario B station nearest to the Juriquilla site.

Figure 12 makes explicit what is being compared: the daily ozone cycles in Scenario A and in each of FEO's three complete years, together with the cluster composition of each series. The histograms are intended as a visualization aid rather than as a direct statistical comparison: they place the independently trained clusters from Scenario A and Scenario B side by side on the same ozone axis, even though the two models never saw each other's data and did not share a common label space. The three FEO years visibly differ in cluster composition, not only in their overall ozone levels. In particular, the cooler, low-ozone cluster (B4) becomes less frequent from 2021–2022 to 2023–2024, whereas the hottest, driest cluster (B1) becomes more frequent. This is the same regime shift documented in Section 4.3, but here it is visible directly in the distribution of individual daily values rather than through annual percentages.

The cross-model analysis that follows is intended as a framework for comparing the daily ozone-cycle structure captured by two independently trained clustering solutions, not as evidence that corresponding clusters represent identical or complete atmospheric regimes. The object being compared is deliberately narrower than the full atmospheric state: the shape and level of the daily ozone cycle. This representation is more informative than comparing daily mean ozone alone, because cycle shape retains the diurnal structure that a single mean discards, and more robust than comparing the raw time series directly, which is sensitive to small temporal offsets and day-to-day noise. By reducing the data to a limited set of recurring cycle shapes, the cluster-based representation provides a more interpretable basis for comparison. Read in this way, the correspondences reported below indicate that Scenario A's and Scenario B's ozone-cycle regimes recur with comparable frequency and in a comparable order, without implying that the

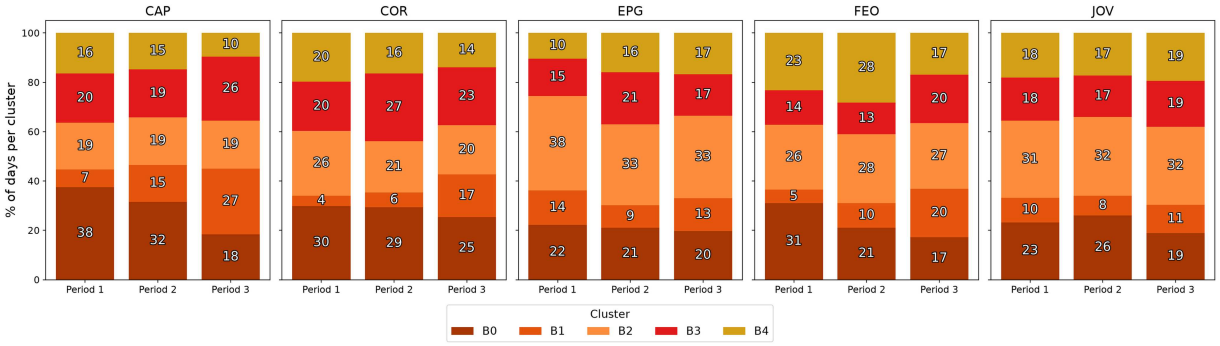


Fig. 10: Cluster composition ($k = 5$) by station across the three complete-coverage periods; percentage of days per cluster labeled on each bar segment.

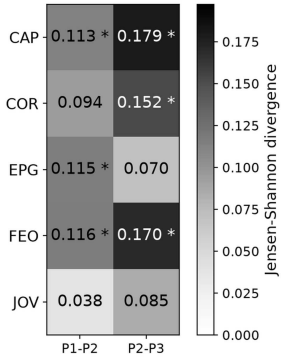


Fig. 11: JS divergence between consecutive complete-coverage periods, by station, Scenario B ($k = 5$). Asterisk: significant pairwise chi-squared test of homogeneity (family of 10 comparisons, Periods 1–3).

underlying atmospheric drivers – including emissions, boundary-layer dynamics, or synoptic forcing – are themselves identical across the two datasets.

Step 1: geometric correspondence.

We projected Scenario A's and Scenario B's daily ozone cycles into a single shared space via a joint PCA, fitted once on the pooled, unnormalized ppb data. The analysis yielded three components explaining 55.1%, 18.9%, and 6.4% of the variance, respectively. Four of the five cluster pairs were mutual nearest neighbors: A0↔B3, A1↔B4, A2↔B2, and A4↔B1 (Figure 13). Only A3 lacks a reciprocal partner: its nearest neighbor, B2, already reciprocates with A2, leaving B0 as the only Scenario B cluster with no reciprocal match in Scenario A. One pair carries a secondary suitor worth flagging here, since it resurfaces below: after its own reciprocal partner B1, A4's next-closest Scenario B cluster is B3 – roughly four times farther than A0's tight reciprocal link to B3, but close enough to leave a faint secondary pull on an otherwise clean match. The four reciprocal pairs agree not only in the shape of their mean diurnal ozone cycles but also in absolute level, differing by only 1–5 ppb in each pair. This is a demanding match because the joint PCA was fitted on unnormalized data, so differences in absolute ozone level were retained rather than discounted.

Because the joint PCA in Step 1 is fitted on ozone cycles alone, the correspondence above is, on its own, partly circular: clusters defined to separate different ozone regimes will naturally appear close in a space constructed from ozone. A stronger test is therefore whether the same geometric proximity also predicts similarity in variables not used in the clustering. We compared each cluster's mean meteorological and precursor fingerprint (temperature, relative humidity, wind speed, NO_2, CO, SO_2, PM_{10}, $PM_{2.5}$, precipitation, and wind direction – the variables measured comparably in both the Scenario A campaign and the Scenario B operational network, with ozone itself excluded to avoid reintroducing this circularity) using a Mantel-style test. Each variable was z-scored across the ten cluster fingerprints, after which the 5×5 Euclidean distances between Scenario A and Scenario B fingerprints were computed and their 25 entries correlated with the Step-1 geometric distance matrix. Significance was assessed against an exact permutation null consisting of all $5! = 120$ relabelings of the Scenario B side. The two distance matrices were significantly correlated ($r = 0.39$, exact $p = 0.017$; rank agreement $\rho = 0.41$, $p = 0.044$; Figure 14): cluster pairs that are close in ozone-cycle space are also, independently, close in their weather and precursor loading. This is visible directly in the four Step-1 reciprocal pairs (diamonds in Figure 14), which cluster toward the low end of both axes rather than scattering among the remaining 21 combinations. Figure 15 makes this correspondence concrete for the four reciprocal pairs: temperature, humidity, wind speed, and precursor levels track each other within each pair, not only their ozone cycles.

This also gives the Step-1 ambiguity around cluster B3 a second, independent line of evidence, rather than leaving it as a possible ozone-only artifact: in the auxiliary-variable space, B3's closest Scenario A match is again A0 (its Step-1 reciprocal partner),

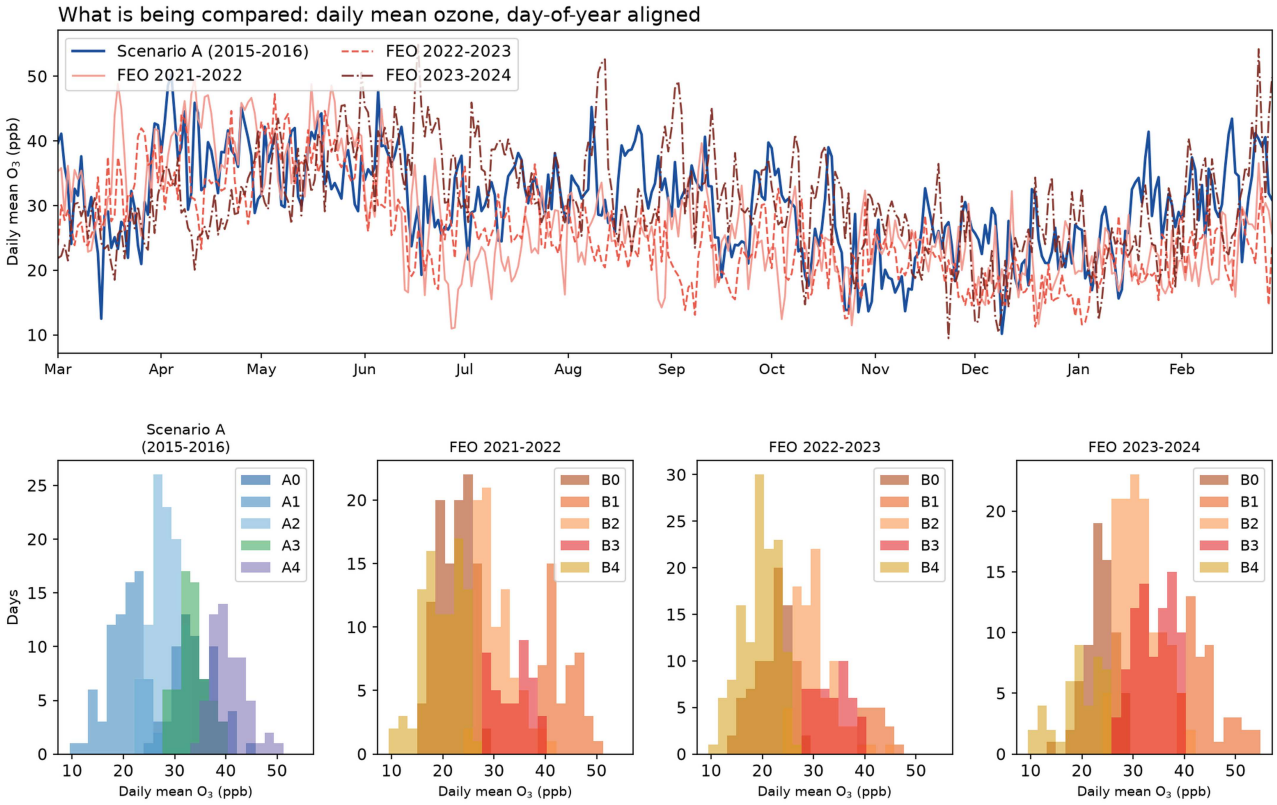


Fig. 12: Daily mean ozone for Scenario A and each of FEO's three complete years, aligned by day of year (365 days, March-to-March). *Top:* time series. *Bottom:* distribution of daily mean ozone for each series, with its own cluster composition overlaid as semi-transparent histograms (Scenario A: clusters A0–A4; FEO: clusters B0–B4, same palettes as Figure 13).

and its second-closest is again A4 (B3's unreciprocated secondary suitor from Step 1, flagged above) – the same fuzziness reappears in a completely independent set of variables (Supplementary Figure D17), consistent with B3 genuinely sitting at a blend of regime types rather than the clustering being unable to separate them only because of how ozone alone was measured.

Step 2: does the geometry hold up dynamically?

Geometric proximity between centroids is a static, cross-sectional statement: it does not establish whether the *sequences* that produced those centroids also behave similarly over time, with comparable frequencies, without systematic phase offsets, and with similar patterns of regime succession. To test this, we constructed three complementary statistics, all using the Step-1 distance matrix as their common measure of physical similarity between clusters.

The first is a distance-weighted version of Cohen's kappa, the same statistic used for cross-station comparability in Section 4.2, but with its usual exact-match scoring replaced by a graded penalty proportional to centroid distance. Thus, days assigned to reciprocal pairs contribute as near-matches rather than mismatches. This statistic captures day-aligned physical similarity while correcting for chance agreement given each series' own cluster frequencies.

The second, used as a more conservative corroborating check, is a banded DTW alignment, which uses the same distance matrix as a local substitution cost to align the two label sequences within a ±14-day window (the timescale of synoptic persistence), testing whether a plausible temporal offset improves the match.

The third is a transition-matrix comparison. Each series' first-order Markov transition matrix – representing the day-to-day probability of switching from one regime to another – is translated into the other series' label space through a soft, distance-weighted mapping, after which the two matrices are compared entrywise. Unlike the other two statistics, this measure is insensitive to day-level alignment and to the overall frequency of the regimes. It therefore tests specifically whether the pattern of regime succession is similar in the two series.

Before applying these statistics to a comparison whose answer was not known in advance, we validated all three against two positive controls – cases where genuine recurrence was expected on independent grounds – each designed to isolate a different source of uncertainty in the pipeline. The first, FEO's

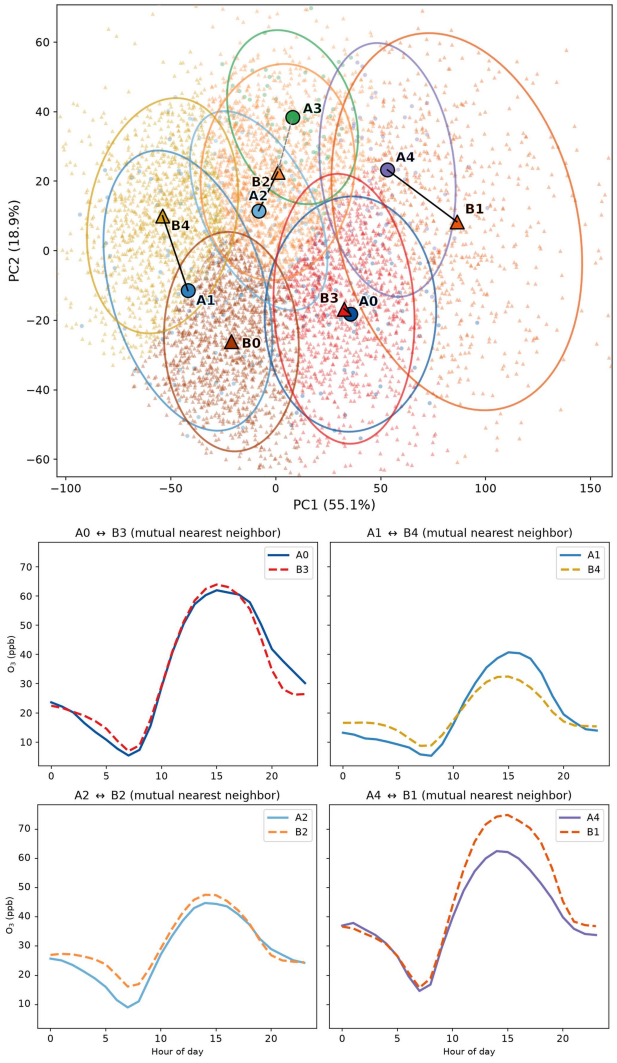


Fig. 13: Cross-model correspondence between Scenario A and Scenario B clusters ($k = 5$). *Top:* joint PCA space (PC1: 55.1% variance; PC2: 18.9%), fitted once on the pooled, unnormalized daily ozone cycles of both scenarios. Circles: Scenario A clusters (A0–A4); triangles: Scenario B clusters (B0–B4). Solid lines connect the four mutual nearest-neighbor pairs (A0↔B3, A1↔B4, A2↔B2, A4↔B1); the dashed arrow shows A3's one-directional nearest match (B2, whose own nearest neighbor is A2, not A3), leaving B0 the one cluster claimed by neither. *Bottom:* mean diurnal ozone cycle for each of the four reciprocal pairs.

Period 1 versus Period 3 using the same trained model, requires no cross-model mapping at all, since both periods already share one label alphabet; it tests whether the three statistics themselves can detect known recurrence under the simplest possible conditions. The second, FEO's Period 2 versus Period 3 using two *independently retrained* k=5 models with their own freshly computed distance matrix, adds back the full cross-model mapping step – exactly as

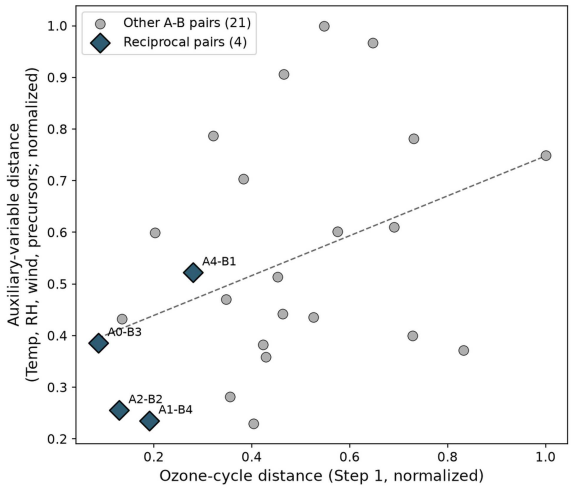


Fig. 14: Ozone-cycle geometric distance (joint PCA) versus auxiliary-variable distance (temperature, RH, wind speed, NO_2, CO, SO_2, PM_{10}, $PM_{2.5}$, precipitation, and wind direction), one point per Scenario A×Scenario B cluster pair (25 pairs total). Diamonds: the four mutual nearest-neighbor pairs from Step 1. Dashed line: linear fit.

required for the Scenario A versus B comparison – while still comparing periods from the same station, where recurrence remains the expected outcome (full protocol in Supplementary Material D). Distance-weighted kappa and the transition-matrix comparison detected the expected recurrence reliably in both controls. Banded DTW detected it in the simpler case but lost significance once the cross-model mapping was added, pinpointing that step, rather than the statistic itself, as its weak point; accordingly, we report its result below as corroborating rather than primary evidence.

Application: which of Scenario B's years most resembles Scenario A?

Scenario A spans a single year (2015-03-01 to 2016-02-29). We compared it against each of the three complete March-to-March years available at FEO, aligned by day-of-year (365 days each). All three years show significant kappa and significant transition-matrix correspondence with Scenario A, yielding a stronger and more consistent result than either validation case considered separately (Table 7, Figure 16). The two statistics nevertheless disagree on which year is "most similar": kappa ranks 2021–2022 highest (day-level physical similarity, 0.166), whereas the transition-matrix comparison ranks 2023–2024 highest (switching dynamics, $r = 0.400$). This mirrors the disagreement between complementary criteria observed for cluster B3 in Step 1 and underscores that "more similar" is not a single quantity, but depends on the aspect of similarity being considered.

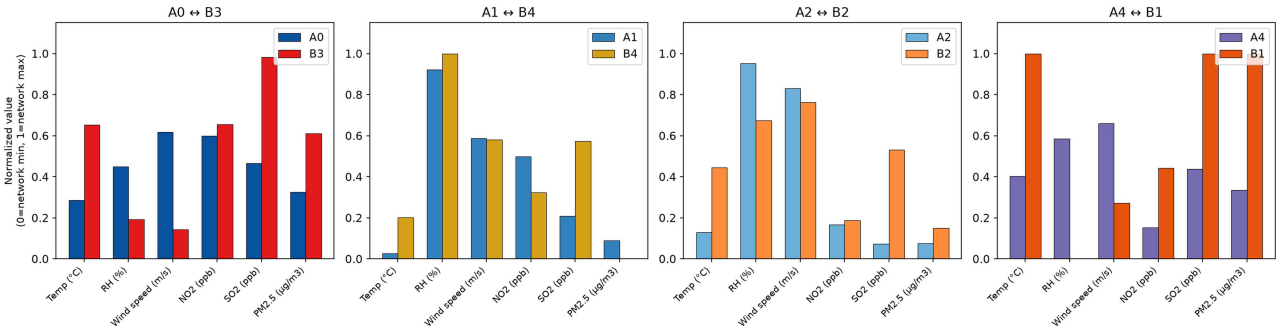


Fig. 15: Meteorological and precursor fingerprint for each of the four mutual nearest-neighbor pairs from Step 1 (Scenario A cluster: solid; Scenario B cluster: adjacent bar), each variable normalized 0–1 across the pooled ten-cluster range. Companion to Figure 13, which compares the same four pairs' ozone cycle shape only.

DTW shows the same overall tendency but reaches significance only for 2021–2022: of the three years, only this comparison rejects the null at $p_{\text{adj}} < 0.05$. This is consistent with DTW being the least sensitive of the three statistics once genuine cross-model uncertainty is introduced (Supplementary Material D).

FEO year	Kappa	DTW z (p_{adj})	Transition r
2021–2022	0.166*	1.67* (0.033)	0.270*
2022–2023	0.148*	1.47 (0.123)	0.374*
2023–2024	0.111*	1.23 (0.148)	0.400*

Table 7: Scenario A vs. each complete March-to-March year at FEO. Asterisks denote $p_{\text{adj}} < 0.05$ (3 comparisons, 2000 permutations). DTW z is sign-flipped so that, as with kappa and the transition-matrix correlation, a larger positive value indicates stronger correspondence with the null.

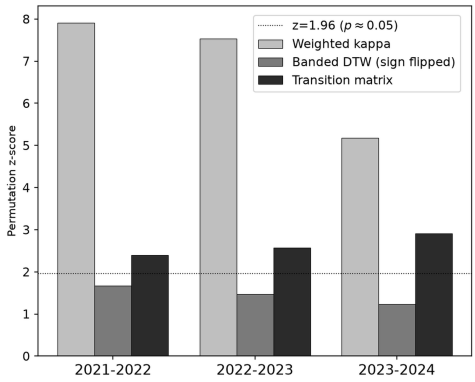


Fig. 16: Summary z-scores of Table 7, by year and statistic.

Two findings stand out from this comparison. First, we obtain an internally validated framework for comparing independently trained clustering solutions within the same airshed: distance-weighted kappa and the transition-matrix comparison successfully recovered known recurrence in the validation cases, supporting their application to the cross-model comparison above. Second, the empirical correspondence between Scenario A and Scenario B is stronger than the geometric correspondence in Step 1 alone would suggest: across every year tested, the regimes correspond not only in aggregate frequency but also in the day-to-day pattern of regime succession. This provides evidence that the regime-label approach generalizes across independently trained models fitted to non-overlapping data, rather than only across stations or years within a single model. For the specific question motivating this section, 2021–2022 emerges as the closest FEO year to Scenario A, agreeing with it on two of the three independent criteria: it has the highest day-level physical similarity (kappa) and is the only year for which DTW reaches significance. The transition-matrix comparison alone instead favors 2023–2024, underscoring that "most similar" depends on which aspect of similarity is being considered.

Discussion

The two scenarios analyzed here serve different purposes, and this discussion follows that division: Scenario A's short, chemically detailed campaign is where the daily ozone cycle can be interpreted as a signature of the underlying photochemistry; Scenario B's multi-year operational network is where that same signature becomes a tool for comparing stations and years directly; and their independent combination provides the most stringent test of the approach. We discuss these contributions in turn, then consider what the comparability result implies for air-quality monitoring beyond Querétaro and where the study design limits the conclusions that can be drawn.

What Scenario A reveals about ozone formation in Querétaro

Taken together, the five regimes identified in Scenario A (Section 3.1.2) point to two controlling factors operating in sequence rather than to a single dominant process. In the cool months, fresh emissions and their precursors accumulate through the morning without sufficient sunlight for efficient photochemical processing, so ozone remains low even when precursor availability is high. In these conditions, precursor supply is therefore not the primary limiting factor. In the hot months, a similar or even smaller precursor supply is processed much more efficiently, producing a faster rise and a substantially higher afternoon peak. What distinguishes an ordinary day from an ozone episode in Querétaro, in other words, appears to be less the amount of morning traffic emissions than the amount of sunlight available to process them. A third, independent axis—wind-driven ventilation—can override either seasonal pattern by dispersing accumulated precursors and secondary aerosol regardless of season. Querétaro's ozone dynamics are therefore better understood as the result of radiation and ventilation acting on a comparatively stable precursor supply, rather than as a system directly limited by precursor availability.

This ventilation effect is concrete rather than inferred. In Scenario A it isolates cleanly in cluster 3, whose near-unidirectional wind above 5 m/s coincides with the lowest levels of nearly every measured precursor and particulate species among the five clusters (Section 3.1.2). The same signature – elevated wind speed paired with depressed precursor and particulate loading – reappears independently in Scenario B's cluster 2 (Section 3.2.2), a network measured at different sites, in a different decade, and processed by an independently trained clustering model. This convergence across two datasets that share no stations, years, or instrumentation indicates that wind-driven ventilation is a genuine, transferable control on Querétaro's ozone cycle rather than an artifact of either dataset (two further, less resolved structural signatures in Scenario B are examined in Supplementary Material C.1).

The cool-season/warm-season contrast described above – muted photochemistry and primary-pollutant buildup in the former, radiation-driven ozone and secondary loading in the latter – parallels a dichotomy that Salcedo et al. (2023) established directly from aerosol composition at this site: primary, combustion-dominated aerosol in the dry-cold season versus peak ozone with regionally influenced secondary loading in the dry-warm season. The parallel holds at a finer timescale than the original comparison: here it emerges at the scale of individual daily cycles, rather than seasonal averages.

What the daily-cycle partition adds to this seasonal picture is resolution at the scale of individual days: the primary/secondary dichotomy is not only a seasonal average but also a recurring day-type structure that can be identified from the ozone cycle alone. The wind-ventilated regime further shows that this structure can be interrupted even within an otherwise favorable season, revealing a layer of variability that seasonal binning cannot resolve. Whether the co-occurrence of high-ozone and high-secondary-aerosol days reflects the same nighttime NO_3 pathway or a distinct, more directly photochemical process remains unresolved. Salcedo et al. (2023) found nitrate formation at this site to be essentially decoupled from the same night's ozone concentration, so the day-level association reported here should not be interpreted as evidence for that specific nocturnal mechanism without nighttime chemical speciation resolved at the level of individual regimes—a natural target for future measurements, but one that this dataset alone cannot resolve. The airshed has also continued to change since these regimes were characterized: a repeat campaign at the same site in 2022 measured a substantially higher aerosol burden than during the 2015–2016 period analyzed here Olivares-Salazar et al. (2025). This places Scenario A's day-type structure within a documented period of change in the local airshed, which Section 4.3 examines directly for ozone using the longer, multi-station record.

Scenario B: one network, not five separate stories

On any given day, Querétaro's five monitoring stations show strong agreement in which regime is occurring (Section 4.2): a hot, dry, high-ozone day at one site is substantially more likely to coincide with the same regime at the other stations than with any alternative regime. High- and low-ozone days therefore appear to unfold as metropolitan-scale events, consistent with a regional meteorological driver rather than being primarily determined by station-specific conditions. For air-quality monitoring, this regional synchrony is useful because each station provides information about the broader atmospheric state while retaining site-specific information about how frequently different regimes occur at that location.

That agreement, however, concerns *which* regimes occur, rather than *how frequently* they occur. Each station exhibits a distinct distribution of regimes, with a greater frequency of the hottest and driest regime at the denser urban core and industrial site, and a greater frequency of the cooler and wetter regime at the outlying municipality. These differences likely reflect local land-use characteristics in addition to variations in synoptic meteorological conditions.

Reducing the network to a single representative station would therefore preserve the shared regional signal but obscure an important aspect of spatial variability: the site-specific frequency with which a given location experiences conditions relevant to exposure. The five stations can thus be considered as a coherent regional network while retaining meaningful differences in regime occurrence that would be lost through spatial aggregation.

The same representation turns interannual variability from a source of noise to be averaged away into a diagnostic in its own right. Comparing regime composition across periods shows that several station–period pairs, concentrated at the industrial site and one urban station, cannot be reliably pooled with earlier years without obscuring meaningful changes in regime occurrence, whereas others—including one urban station throughout the record and most periods at another—show sufficient consistency to support pooling. Where pooling is not appropriate, the temporal shift is physically interpretable: these stations spent a measurably larger fraction of recent years under hot, dry, rapidly rising, high-ozone conditions, and a correspondingly smaller fraction under cool, wet, weakly varying ozone cycles, than they did earlier in the record. Because a similar shift occurs at three geographically separated stations, the pattern is more consistent with a regional tendency—potentially associated with warmer and drier interannual conditions—than with a change confined to a single site. The most direct implication is therefore not simply year-to-year variability, but a shift in the regional frequency of conditions associated with elevated ozone, with potential consequences for ozone exposure.

Cross-model comparability: a non-trivial result for monitoring

Everything described above derives from a single model trained on a single dataset. A more stringent test is whether the same regimes remain identifiable when two entirely independent datasets—from different sites, different time periods, and different measurement systems, with models trained independently—are used to characterize the same atmospheric system. This comparison brings together Scenario A's short, chemically rich campaign and Scenario B's multi-year operational network (Section 4.4). No correspondence between the resulting regimes was assumed a priori: because the two models were trained independently, using data collected years apart at different sites and with different instrumentation, they could have identified substantially different partitions of the ozone cycle.

They did not. The regimes identified by the two models correspond not only in the shape of the ozone cycle—the feature used to distinguish the regimes—but also independently in temperature, humidity, wind, and precursor loading, none of which was used by either clustering model to define the regimes. The correspondence therefore extends beyond similarity in ozone-cycle shape: the regimes capture coherent atmospheric states in which meteorological conditions and precursor chemistry vary systematically with ozone across both datasets. This correspondence was also evident in the temporal dynamics, rather than only in average regime characteristics. When the practical question was posed of which recent years in Scenario B most closely resembled those in Scenario A, the results were consistent across independent comparisons based on both the similarity of individual days and the similarity of regime-to-regime transitions.

The practical implication is particularly relevant for monitoring networks that lack the chemical detail available in Scenario A. Most operational networks more closely resemble Scenario B: they provide continuous, lower-cost measurements but lack the instrumentation typically available during intensive research campaigns. The present results show that such networks can nevertheless be compared with chemically richer campaigns conducted at other sites, without requiring repeated deployment of the same intensive instrumentation. Once daily ozone behavior is represented by a regime label, comparisons across stations, years, and independent studies can be performed at the level of regime correspondence, rather than requiring identical measurement windows or instrumentation. In this sense, historical intensive campaigns retain analytical value beyond their original measurement period: the chemically resolved regimes they characterized can recur at other sites and at later times, and operational networks with more limited chemical measurements can identify their occurrence through the corresponding regime structure.

This result should not, however, be overinterpreted. As an exploratory check – not a confirmatory test, given the limited number of independent station-years available – we asked whether cross-model similarity could serve as an early-warning indicator of longer-term changes in temperature, humidity, or pollutant levels—an appealing extension given that regime correspondence is associated with these variables. Across all five Scenario B stations and every complete year available, we found no clear association with such changes. Regime labels summarize the *type* of day that occurred rather than the magnitude of any individual underlying variable, placing a clear limit on their use: the framework provides a basis for comparing atmospheric regimes, but it does not substitute for direct measurements when quantifying temporal trends in their constituent variables. The interannual shift in ozone documented above was

therefore established through more direct analyses and remains independent of this exploratory result.

One practical choice remains, and it is a methodological choice rather than an intrinsic property of the method: the resolution at which the regime space is partitioned. This choice is not symmetric: $k = 2$ is supported by every validation criterion in both scenarios (Supplementary Material A), whereas $k = 5$ rests on the elbow heuristic alone, with the remaining criteria weak or divided. The coarser and more robust partition is sufficient for comparing regime prevalence across stations or characterizing broad shifts in the hot–cool balance over time. Detecting subtler interannual changes or finer within-season structure, however, may require the exploratory finer partition, whose process-level interpretations in this study should accordingly be read as hypothesis-generating rather than as confirmed structure. Applications of this approach should therefore select the level of resolution according to the specific comparison or research question, rather than treating a single partition as universally optimal.

Limitations

This is an observational study rather than an experimental investigation: the interpretation of ozone mechanisms in Section 5.1 is based on convergent statistical associations rather than controlled manipulation, and several mechanistic questions therefore remain unresolved. Moreover, the analysis is restricted to a single metropolitan area. Whether the observed regime structure and cross-dataset comparability extend to other urban environments, monitoring networks, or pollutants remains an open question for future research.

A related point concerns the cross-model comparison of Section 4.4: because Scenario A and Scenario B measure the same regional atmosphere, that comparison tests transferability across instrumentation, time period, and independently trained models within a single airshed, not across genuinely independent urban environments. A stronger test of geographic transferability would compare against a chemically resolved campaign from a different city, which is beyond the scope of the present datasets.

Two further design choices were fixed a priori rather than validated against alternative specifications: the 25-hour representation of the daily cycle (Section 2.2), and the specific percentile thresholds (75th/25th) defining the operational high-ozone/low-NO_x/high-secondary-aerosol criterion (Section 3.1.1). We have no specific reason to expect either choice to qualitatively change the cluster structure reported here; both were subsequently tested for sensitivity, with results reported in Supplementary Material A.

A further limitation concerns the Scenario B period (2021–2024), which partially overlaps with Mexico's post-COVID economic and activity recovery. The year showing the greatest seasonal divergence across the network was also the earliest in this period, and the available data do not allow us to fully disentangle a genuine activity-related signal from effects associated with incomplete station coverage during that year. We therefore treat this factor as a potential confounder rather than attempting to resolve its contribution from the available data. Extending the record to include additional years before and after the recovery period would provide a stronger basis for distinguishing these alternative explanations.

Conclusion

This study addressed two questions concerning Querétaro's diurnal ozone cycle: what the cycle reveals about the processes shaping ozone variability in the region, and whether the same signal can be compared across stations, years, and independent studies without requiring each comparison to be developed anew. The two-cluster partition provided the robust framework for the latter objective, remaining consistent across clustering methods and validation criteria. The finer, exploratory five-cluster partition – supported primarily by the elbow heuristic rather than by the full set of validation criteria (Section 2.2) – provided additional resolution for the former, distinguishing radiation-driven and wind-driven ventilation regimes as two physically interpretable, though hypothesis-generating, controls on the diurnal ozone cycle. Together, these results complement and extend the seasonal and composition-based characterization previously established for this airshed.

The central result is the transferability of the regime representation across space and time. Same-day regime agreement across Querétaro's five monitoring stations is far above chance, while each station retains its own regime distribution – a distinction examined in Section 5.2. Regime composition also changes measurably over only a few years at four of the five stations, most often toward a greater frequency of hot, dry, high-ozone regimes, though not uniformly across the network – a shift consistent with increasing conditions conducive to ozone exposure rather than simple year-to-year variability. Most importantly, the same regimes identified independently by a short, chemically detailed campaign and a multi-year operational network correspond not only in ozone-cycle shape but also in meteorological conditions and precursor chemistry that were not used to define either clustering solution.

This latter result has broader implications for air-quality monitoring beyond Querétaro, where routine networks typically offer Scenario B's coarser measurement capabilities rather than Scenario A's

chemical depth. The present results show that such networks can remain comparable across years, neighboring stations, and chemically richer campaigns conducted at other sites and times, without requiring equivalent instrumentation. This comparability is not automatic and should be demonstrated rather than assumed, nor does it replace direct measurements when the objective is to quantify long-term trends in the variables underlying the regimes. Nevertheless, regime-based representation provides a common framework through which heterogeneous ozone monitoring datasets can be compared. Because Scenario A and Scenario B measure the same regional atmosphere (Section 5.4), the present results establish this transferability within a single airshed, across instrumentation and time; whether the approach extends to genuinely independent cities and monitoring networks remains to be tested rather than assumed.

Supplementary information. Beyond the core analyses presented in the main text, Supplementary Material provide additional methodological and diagnostic detail.

Acknowledgements. The author thanks Sara Ericka Salazar Olivares for helpful discussions that informed the atmospheric interpretation of the results presented in this study.

Declarations

Funding. This work was supported by the UNAM internal project UMDI-J-126526, "Mathematical Modeling of Complex Physicochemical Systems," and by the UNAM DGAPA-PAPIIT project IA100126, "Advanced Modeling and Machine Learning for Environmental Sustainability."

Conflict of interest/Competing interests. The authors declare no competing interests.

Ethics approval and consent to participate. Not applicable.

Consent for publication. Not applicable.

Data availability. The Scenario A dataset (Juriquilla, 2015–2016) is available through the publication of Salcedo et al., "PM chemical characterization at a semi-arid urban environment in Central Mexico" Salcedo et al. (2023). The Scenario B dataset (CeMCAQ operational monitoring network, 2021–2024) is held by the Centro de Monitoreo de Calidad del Aire de Querétaro, an agency of the Querétaro state government, and is available upon reasonable request to that institution Secretaría de Desarrollo Sustentable (SEDESU) (2024).

Materials availability. Not applicable.

Code availability. The analysis code will be made publicly available at https://github.com/aldoleo2/ozono-queretaro-paper upon publication of this manuscript.

Author contribution. A.L.D.: conceptualization, methodology, software, formal analysis, data curation, writing – original draft. D.S.: atmospheric interpretation, writing – review and editing.

References

Alahamade, W., Lake, I., Reeves, C.E., De La Iglesia, B.: Evaluation of multivariate time series clustering for imputation of air pollution data. Geoscientific Instrumentation, Methods and Data Systems **10**(2), 265–285 (2021) https://doi.org/10.5194/gi-10-265-2021

Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B **57**(1), 289–300 (1995) https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boleti, E., Hueglin, C., Grange, S.K., Prévôt, A.S.H., Takahama, S.: Temporal and spatial analysis of ozone concentrations in Europe based on timescale decomposition and a multi-clustering approach. Atmospheric Chemistry and Physics **20**(14), 9051–9066 (2020) https://doi.org/10.5194/acp-20-9051-2020

CeMCAQ: Informe mensual de calidad del aire, mayo 2025. Technical report, Centro de Monitoreo de la Calidad del Aire del Estado de Querétaro (CeMCAQ), Secretaría de Desarrollo Sustentable, Gobierno del Estado de Querétaro (2025). https://www.cemcaq.mx/images/informes/2025/IMCA%20May2025.pdf

Castañeda-Miranda, A.G., Böhnel, H.N., Molina-Garza, R.S., Chaparro, M.A.: Magnetic evaluation of tsp-filters for air quality monitoring. Atmospheric Environment **96**, 163–174 (2014) https://doi.org/10.1016/j.atmosenv.2014.07.015

Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**(1), 37–46 (1960) https://doi.org/10.1177/001316446002000104

Gouveia, S., Scotto, M.G., Monteiro, A., Alonso, A.M.: Wavelets-based clustering of air quality monitoring sites. Environmental Monitoring and Assessment **187**(11), 694 (2015) https://doi.org/10.1007/s10661-015-4901-z

Lyapina, O., *et al.*: Cluster analysis of European surface ozone observations for evaluation of MACC reanalysis data. Atmospheric Chemistry and Physics **16**(11), 6863–6877 (2016) https://doi.org/10.5194/acp-16-6863-2016

Liñán-Abanto, R.N., Peralta, O., Salcedo, D., Ruiz-Suárez, L.G., Arnott, P., Paredes-Miranda, G., Alvarez-Ospina, H., Castro, T.: Optical properties of atmospheric particles over an urban site in mexico city and a peri-urban site in queretaro. Journal of Atmospheric Chemistry (2019) https://doi.org/10.1007/s10874-019-09394-1

Leighton, P.A.: Photochemistry of Air Pollution. Academic Press, New York (1970)

Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159–174 (1977) https://doi.org/10.2307/2529310

MacQueen, J.: Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability **1**, 281–297 (1967)

Nelson, B.S., Drysdale, W.S.: Urban ozone trends in Europe and the USA (2000–2021). Atmospheric Chemistry and Physics **25**(22), 16009–16026 (2025) https://doi.org/10.5194/acp-25-16009-2025

Olivares Salazar, E., Alvarez-Ospina, H., Aguillón Vázquez, C., Salcedo, D.: Source apportionment of particulate matter in the metropolitan area of Querétaro (Central Mexico): First case study. ACS Earth and Space Chemistry **5**(9), 2347–2355 (2021) https://doi.org/10.1021/acsearthspacechem.1c00122

Olivares-Salazar, S.E., Bahreini, R., Lin, Y.-H., Castro, T., Alvarez-Ospina, H., Salcedo, D.: Aerosol composition in a semi-urban environment in central mexico: Influence of local and regional processes on overall composition and first quantification of nitroaromatics. Atmosphere **16**(7), 827 (2025) https://doi.org/10.3390/atmos16070827

Rodríguez-Trejo, A., Böhnel, H.N., Ibarra-Ortega, H.E., Salcedo, D., González-Guzmán, R., Castañeda-Miranda, A.G., Sánchez-Ramos, L.E., Chaparro, M.A.E.: Air quality monitoring with low-cost sensors: A record of the increase of pm2.5 during christmas and new year's eve celebrations in the city of queretaro, mexico. Atmosphere **15**(8), 879 (2024) https://doi.org/10.3390/atmos15080879

Rozanes-Valenzuela, D.A., Magaldi, A.V., Salcedo, D.: Regional flow climatology for central Mexico (Querétaro): A first case study. Atmósfera **36**(2), 239–252 (2023) https://doi.org/10.20937/ATM.53096

Salcedo, D., Alvarez-Ospina, H., Olivares-Salazar, S.E., Liñan-Abanto, R.N., Castro, T.: PM chemical characterization at a semi-arid urban environment in Central Mexico. Urban Climate **52**, 101723 (2023) https://doi.org/10.1016/j.uclim.2023.101723

Salcedo, D., Alvarez-Ospina, H., Peralta, O., Castro, T.: PM_1 chemical characterization during the ACU15 campaign, south of Mexico City. Atmosphere **9**(6), 232 (2018) https://doi.org/10.3390/atmos9060232

Secretaría de Medio Ambiente y Recursos Naturales: NOM-156-SEMARNAT-2012, Establecimiento y operación de sistemas de monitoreo de la calidad del aire. Diario Oficial de la Federación. Publicado el 16 de julio de 2012 (2012)

Secretaría de Medio Ambiente y Recursos Naturales: NOM-172-SEMARNAT-2019, Lineamientos para la obtención y comunicación del Índice de Aire y Salud. Diario Oficial de la Federación. Publicado el 20 de noviembre de 2019 (2019)

Secretaría de Desarrollo Sustentable (SEDESU): Datos de calidad del aire 2024. Technical report, Secretaría de Desarrollo Sustentable (SEDESU), Centro de Monitoreo de Calidad del Aire del Estado de Querétaro (CeMCAQ), Gobierno del Estado de Querétaro (2024)

Soares, J., Makar, P.A., Aklilu, Y., Akingunola, A.: The use of hierarchical clustering for the design of optimized monitoring networks. Atmospheric Chemistry and Physics **18**(9), 6543–6566 (2018) https://doi.org/10.5194/acp-18-6543-2018

Tamas, W., Notton, G., Paoli, C., Nivet, M.-L., Voyant, C.: Hybridization of air quality forecasting models using machine learning and clustering: An original approach to detect pollutant peaks. Aerosol and Air Quality Research **16**(2), 405–416 (2016) https://doi.org/10.4209/aaqr.2015.03.0147

Wang, X., Wang, S., Zhang, S., Gu, C., Tanvir, A., Zhang, R., Zhou, B.: Clustering analysis on drivers of O_3 diurnal pattern and interactions with nighttime NO_3 and HONO. Atmosphere **13**(2), 351 (2022) https://doi.org/10.3390/atmos13020351

Zhu, J., *et al.*: Why did ozone concentrations remain high during Shanghai's static management? A cluster analysis perspective. Atmospheric Chemistry and Physics **24**(14), 8383–8402 (2024) https://doi.org/10.5194/acp-24-8383-2024

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf