A Review of Statistical Air Quality Assessment Methods

Mingyuan Bian,[1] and Song Xi Chen[2]

[1]School of Mathematical Sciences, Peking University, Beijing 100871, China; mingyuanbian@stu.pku.edu.cn
[2]Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China; csx@gsm.pku.edu.cn

. :1–25

Keywords

Air pollution, Extreme value analysis, Meteorological adjustment, Spatiotemporal models, Trend analysis.

Abstract

Air-quality assessment is fundamental for quantifying health risks and evaluating emission-control policies. However, assessing air-quality data is challenging due to strong variability, heavy tails, complex spatiotemporal structures, and meteorological confounding, which call for principled statistical methods. This review provides an overview of existing statistical approaches under four themes: trend analysis for detecting systematic changes; extreme value analysis for characterizing high-pollution risk; hierarchical models for capturing spatiotemporal dependence; and meteorological adjustment methods for isolating emission-driven signals. We use air-quality data from North China to illustrate the various reviewed methods.

Corresponding author: Song Xi Chen; email: csx@gsm.pku.edu.cn

1. Introduction

Air pollution imposes substantial risks to human health, ecosystems, and the climate system worldwide (Abbasi-Kangevari et al. 2023, Pozzer et al. 2023). Effective pollution control, therefore, requires air-quality assessment methods that are accurate, transparent, and accompanied by principled uncertainty quantification. Broadly speaking, the assessment tools fall into two complementary classes: atmospheric physics- and chemistry-based dynamic numerical models and data-driven statistical approaches. This review focuses on the latter, namely statistical methods for inference and assessment from observed air-quality data, while noting key connections to physics-based modeling.

Chemical transport models (CTMs) and coupled meteorology–chemistry systems have advanced substantially by integrating emissions, physical transport, and atmospheric chemistry across multiple spatial scales. Representative examples include the Models-3/CMAQ system (Appel et al. 2021), WRF-Chem (Grell et al. 2005), and GEOS-Chem (Bey et al. 2001). At the same time, their performance depends on initial conditions, emission inventories, and physicochemical parameterizations, and high-resolution applications can be computationally demanding. These considerations motivate statistical methods that (i) leverage routine monitoring data for estimation, prediction, and mapping, (ii) quantify uncertainty under complex dependence, and (iii) formalize how meteorology and other drivers confound emission-related signals.

In operational settings, numerical modeling is often paired with data assimilation, which updates forecasts using observations through methods such as the Ensemble Kalman Filter (EnKF) (Houtekamer & Zhang 2016) and variational approaches including 4D-Var (Skachko et al. 2016). From a statistical perspective, data assimilation is a state-space inference problem that bridges mechanistic dynamics and noisy observations. In contrast, purely statistical approaches work directly with observed pollutant and meteorological data to estimate their relationships, interpolate in space and time, and forecast future concentrations, while offering computational simplicity, modeling flexibility, and accounting for stochasticity and measurement error.

Given the breadth of statistical methods used in air-quality studies, it is challenging to provide full coverage of these methods in a single article. We instead focus on methodologies that are widely used in air-quality assessment and supported by well-developed statistical theory. The remainder of this review is organized as follows. Sections 2–5 cover four core statistical methods in air-quality assessment: detecting long-term trends, characterizing extremes and high-pollution tails, modeling spatiotemporal dependence, and inferring emission-related signals under meteorological confounding. Section 6 discusses episodic assessment methods that treat high-pollution periods as dynamic units, complementing full-sample analyses by focusing on onset, persistence, and cumulative pollution loading. We conclude in Section 7 with directions for future research.

2. Trend Analysis

Trend analysis aims to determine whether concentrations exhibit systematic temporal changes and to gauge the effectiveness of emission-control policies. Suppose there are n pollution observations $\{Y_t\}_{t=1}^{n}$ from a monitoring site, where Y_t denotes the concentration at time t (e.g., hourly, daily, or annual). Trend analysis aims to answer whether the distribution of Y_t changes systematically over time and, if so, how large that change is in interpretable units (e.g., μg/m^3 per year). In air-quality applications, these tasks are com-

plicated by strong seasonality, serial dependence, outliers, missing data, and meteorological variability, all of which can shift the distribution of Y_t and obscure emission-driven change. Thus, trend-analysis methodology must provide robust evidence for monotonic increases or decreases under deliberately weak (hence robust) distributional assumptions.

Although many methods have been proposed for trend detection and testing, this review focuses on two mainstream classes of methods: the **regression-based** methods, which posit a parametric or semi-parametric relationship between Y_t and time (and possibly covariates); and **the nonparametric trend tests**, which use rank- or sign-based procedures to detect monotonic change under minimal distributional assumptions.

2.1. Regression-Based Trend Models

A starting point for pollution modeling is the linear regression of Y_t on time t and covariates X_t:

$$Y_t \;=\; \beta_0 \;+\; \beta_1\, t \;+\; \mathbf{X}_t^{\top}\boldsymbol{\gamma} \;+\; \varepsilon_t, \qquad 2.1.$$

where β_0 is the intercept, β_1 is the per-unit time change in concentration (the temporal trend), $\boldsymbol{\gamma}$ collects the effects of the covariates (meteorology or policy indicators), and ε_t is an error term with $\mathbb{E}[\varepsilon_t \mid t, \mathbf{X}_t] = 0$ and finite conditional variance.

In applications, 2.1. is fitted to daily means or season-specific averages, and $(\beta_0, \beta_1, \boldsymbol{\gamma})$ is estimated by ordinary least squares (OLS). Trend analysis focuses on β_1 via its OLS estimate $\widehat{\beta}_1$ with uncertainty (e.g., a 95% confidence interval) and testing whether β_1 is nonzero based on the $t-$ test. Zhai et al. (2019) regressed deseasonalized 10-day mean $PM_{2.5}$ anomalies on time to benchmark national $PM_{2.5}$ change over 2013–2018. Li et al. (2020) applied the regression framework to summertime daily maximum 8-hour average (MDA8) O_3 to characterize the increase in surface ozone across China during 2013–2019.

However, the residuals may not be independent and identically distributed (i.i.d.) as there may be autocorrelation and time-varying volatility. While the OLS point estimates remain consistent, the classical t-test and i.i.d.-based standard errors are typically invalid, motivating corrections that account for dependence and heteroskedasticity. One approach is to specify a parametric error structure (e.g., autoregressive, AR(p), errors) and estimate the model using **the** generalized least squares (GLS) with an estimated covariance model; see Reinsel et al. (1998) and Orsini et al. (2006) for air-pollution applications. A more direct alternative is to keep the OLS point estimates and use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors with the Newey-West estimator (Newey & West 1987) to counter the weak dependence and heteroskedasticity.

Beyond correcting standard errors, policy evaluation or regime characterization often requires relaxing the single-slope assumption in 2.1., which imposes a constant rate of change over the entire study period. Piecewise linear (segmented) trend and change-point models allow for level shifts, slope changes, or both, providing an interpretable summary of regime shifts in atmospheric time series (Aminikhanghahi & Cook 2017). These ideas are supported by popular software such as a Python package `ruptures` (Truong et al. 2020) and the R package `changepoint` (Killick & Eckley 2014). Ma et al. (2021) evaluated the April 2019 introduction of London's Ultra Low Emission Zone (ULEZ) using multi-year monitoring data from the London Air Quality Network, estimating the intervention effects that allowed for policy-induced level or slope changes.

2.1.1. Kolmogorov-Zurbenko (KZ) Filter. Air-quality series exhibit variability on multiple time scales (diurnal, synoptic, seasonal, and interannual). A widely used preprocessing approach is the Kolmogorov–Zurbenko (KZ) filter (Rao & Zurbenko 1994), which separates scales before fitting a trend model and applying a test for trend.

For a window length $m = 2h + 1$ for a positive h, define the centered moving average

$$(A_m Y)_t \;=\; \frac{1}{m} \sum_{j=-h}^{h} Y_{t+j},$$

whenever the averaging window lies within the data time range. The KZ filter with parameters (m, k) iterates A_m operator k times:

$$\mathrm{KZ}_{m,k} Y \;=\; A_m^{\circ k} Y, \quad \textit{where} \quad A_m^{\circ k} = \underbrace{A_m \circ A_m \circ \cdots \circ A_m}_{k \text{ times}}.$$

From a nonparametric perspective, $\mathrm{KZ}_{m,1}$ is a linear kernel smoother with a discrete-time uniform (boxcar) kernel of half-width h,

$$K(j) = \begin{cases} \dfrac{1}{m}, & |j| \le h, \\ 0, & |j| > h, \end{cases} \qquad j \in \mathbb{Z}.$$

The KZ filter $\mathrm{KZ}_{m,k}$ corresponds to an iterated boxcar smoother whose equivalent kernel is the k-fold convolution of K; for moderate to large k, this equivalent kernel is well approximated by a discrete Gaussian kernel (Härdle 1992). Hence, the KZ filter can be viewed in the frequency domain as a low-pass filter with transfer function $H(\omega) = \{\sin(m\omega/2)/[m \sin(\omega/2)]\}^k$, which motivates its use for scale separation.

In air-quality applications, KZ filters with two sets of parameters (m_L, k_L) and (m_S, k_S) are used to isolate long-term and seasonal scales so that the original time series is decomposed into $Y_t \;=\; L_t + S_t + W_t$ where

$$L_t = (\mathrm{KZ}_{m_L,k_L} Y)_t,\; S_t = (\mathrm{KZ}_{m_S,k_S} Y)_t - L_t,\; W_t = Y_t - L_t - S_t. \tag{2.2}$$

Trend analysis is then conducted on the long-term component $\{L_t\}$, either interpreting L_t directly as the trend estimate or applying further analysis to it. Because the smoothing alters the dependence structure, inference on L_t often still requires autocorrelation-robust uncertainty quantification (e.g., HAC or block resampling).

Choosing (m, k) is largely empirical and depends on the sampling resolution and the temporal scales one aims to separate. Increasing m and/or k yields stronger low-pass filtering, retaining progressively lower-frequency variation. A widely used specification in ozone studies is Eskridge et al. (1997), which applies $\mathrm{KZ}_{15,5}$ and $\mathrm{KZ}_{365,3}$ to separate short-term, seasonal, and long-term components in daily ozone; see also Rao et al. (1997) for empirical evidence of effective scale isolation in ambient ozone.

2.1.2. Generalized Additive Models (GAMs). While the KZ filter conducts implicit smoothing in the decomposition of the raw data series, the generalized additive model (GAM) (Hastie & Tibshirani 1986, Wood 2017) carries out explicit smoothing in a semi-parametric regression model with smooth trend and seasonal components:

$$Y_t = \beta_0 + s_{\mathrm{trend}}(t) + s_{\mathrm{season}}(\mathrm{doy}_t) + \sum_{j=1}^{p} s_j(X_{j,t}) + \varepsilon_t, \tag{2.3}$$

where $s(\cdot)$ are smooth functions represented via the penalized regression splines, and doy_t denotes the day-of-year associated with t. Specifically, $s_{\text{trend}}(t)$ captures the trend, $s_{\text{season}}(\text{doy}_t)$ reflects the within-year seasonality, and $s_j(X_{j,t})$ encodes meteorological responses. In contrast to the linear model, trend inference under a GAM is not based on a single slope parameter but on the shape of $\widehat{s}_{\text{trend}}(t)$. In practice, one can inspect $\widehat{s}_{\text{trend}}(t)$ together with pointwise or simultaneous confidence bands for modeling $s_{\text{trend}}(t)$, or perform hypothesis testing for a specific form of the trend. In the presence of residual autocorrelation, the block bootstrap method (Künsch 1989) may be used to calibrate uncertainty with the inference for the trend. Liang et al. (2015) used a moving-block bootstrap method to obtain standard errors for meteorology-adjusted $PM_{2.5}$ means and percentiles, illustrating a practical workflow for dependent air-quality time series.

The GAM trend and seasonal components $s_{\text{trend}}(t)$ and $s_{\text{season}}(\text{doy}_t)$ are analogous to the corresponding terms in the KZ decomposition 2.2.. The difference is that the KZ filter uses a fixed linear filter with user-chosen smoothing parameters (m, k), whereas GAM estimates these components directly from the data through penalized regression splines (Wood 2017). To illustrate how these approaches behave on real data, Figure 1 compares the KZ and GAM-based decompositions for daily mean O_3 at Beijing Aotizhongxin site over 2015–2020. It suggests that KZ filtering and GAM can lead to very similar decompositions of different scale components in practice, and the differences may be attributed to the different smoothing parameters used.

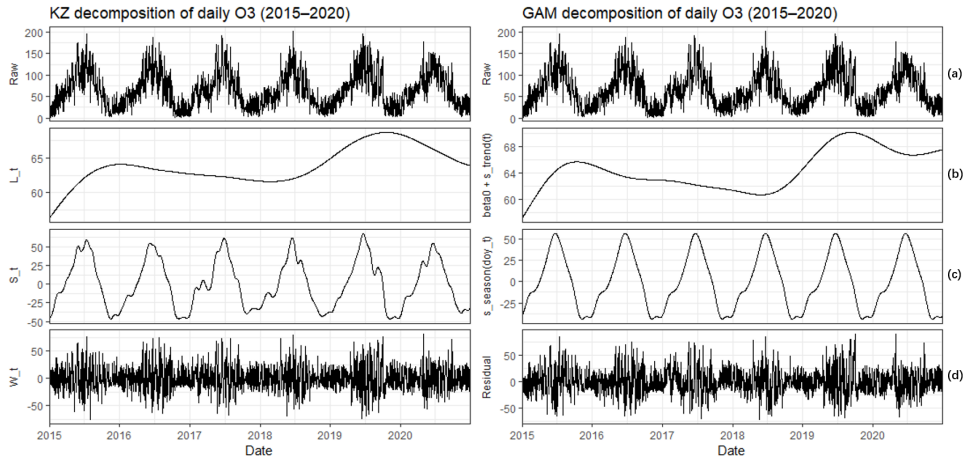

Figure 1

Comparison of the KZ (left panels) and GAM (right panels) decompositions for daily mean O_3 at the Beijing Aotizhongxin site (2015–2020). For the KZ-filter, $(m_L, k_L) = (365, 3)$ for the trend component (left 2nd upper panel) and $(m_S, k_S) = (15, 5)$ for the yearly seasonal component (left 3rd upper panel), and the corresponding components of the GAM (right panels), while the top (bottom) panels are the raw (residual) series.

GAMs are widely used for trend assessment with flexible meteorological adjustment; for instance, Pearce et al. (2011) use GAMs to quantify pollutant–meteorology relationships while controlling for long-term and seasonal structure, and Walker et al. (2023) formalize a general GAM-based framework (AirGAM) with low-rank smooth trends, smooth meteorological effects, and additional calendar terms. Recent deep learning models provide

flexible predictive models for concentrations (Xing et al. 2020), but most applications target short-horizon forecasting rather than uncertainty-quantified long-run trend inference or transparent decomposition into trend, seasonality, and weather-driven variability. An important direction for future work is to develop hybrid models that retain interpretable trend/seasonal components while leveraging deep architectures for nonlinear meteorological effects, together with calibrated uncertainty quantification for long-run summaries.

2.2. Nonparametric Trend Tests

Nonparametric approaches provide robust alternatives for trend analysis when the distributional assumptions underlying parametric or semi-parametric regression are restrictive. Ambient air-quality data are often skewed, heavy-tailed, and affected by outliers, making nonparametric rank-based procedures attractive. Among these, the Mann–Kendall (MK) test (Mann 1945, Kendall 1948) for detecting monotonic trends and the associated Theil-Sen slope estimator (Theil 1950, Sen 1968) for quantifying trend magnitude are the two most widely used in air-pollution studies.

2.2.1. Mann–Kendall (MK) Test. The MK test assesses whether later observations tend to be systematically larger (or smaller) than earlier ones. **The test** statistic aggregates all pairwise order comparisons:

$$S_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(Y_j - Y_i), \tag{2.4.}$$

with $\operatorname{sgn}(x) = 1$ if $x > 0$, 0 if $x = 0$, and -1 if $x < 0$. Under H_0 of no trend (identical continuous marginal distributions) and independence, S_n **as a U-statistic** is asymptotically normal with finite variance $\sigma^2(S_n)$. The standardized statistic $Z = S_n/\sigma(S_n)$ is then compared to $N(0,1)$ for one- or two-sided testing with $\sigma(S_n) = \{n(n-1)(2n+5)/18\}^{1/2}$ (Hirsch et al. 1982). Since S_n depends only on ranks, the MK test does not rely on specific distributional assumptions for Y_t, and it is robust to outliers and monotone transformations. It can also accommodate missing values naturally by working with the available observations, as long as the temporal ordering is preserved. These features are advantageous for ambient air-quality data. Examples include MK-based trend detection applied to (i) multi-pollutant monitoring records at an urban station in India (Chaudhuri & Dutta 2014), (ii) $PM_{2.5}$ concentrations in Rochester, NY (Emami et al. 2018), and (iii) city-level $PM_{2.5}$ series across South Asia (Fattah et al. 2023).

A practical caveat is that the classical MK variance is derived under independence; positive autocorrelation reduces the effective sample size and can inflate type-I error if unaddressed. Two common remedies are (i) modifying $\operatorname{Var}(S)$ using estimates of the autocorrelation function (Hamed & Rao 1998), and (ii) *prewhitening* (often in trend-preserving variants) to filter out short-memory dependence while preserving the monotonic signal (Kulkarni & von Storch 1992, Yue et al. 2002). Collaud Coen et al. (2020) used seven long-term atmospheric observation time series (10-60 years) to benchmark how prewhitening, time granularity, and temporal segmentation affect the MK trend detection.

Strong seasonality motivates a second extension, the seasonal MK test (Hirsch et al. 1982). Hirsch's procedure partitions observations into K blocks (e.g., seasons), computes

within-season MK statistics S_k, and aggregates

$$S = \sum_{k=1}^{K} S_k, \qquad \mathrm{Var}(S) = \sum_{k=1}^{K} \mathrm{Var}(S_k). \tag{2.5.}$$

The seasonal MK test then proceeds by constructing a standardized statistic $Z = S/\sqrt{\mathrm{Var}(S)}$ based on $\mathrm{Var}(S)$. The seasonal MK test is commonly used in air-quality trend assessment. Zeb et al. (2019) applied the seasonal MK test to assess decadal trends (2005-2014) based on satellite-derived CO, HCHO, NO_2, and O_3 over Pakistan.

2.2.2. Theil–Sen Slope. While the MK test assesses whether a monotonic trend is present, many applications require robust estimates of the magnitude of the trend (slope). The Theil-Sen (Sen's) estimator (Theil 1950, Sen 1968) provides a nonparametric slope estimator. Define the pairwise slopes over pairs of observations at time points $i < j$: $D_{ij} = (Y_j - Y_i)/(t_j - t_i)$, and the Theil-Sen estimator is

$$\widehat{\beta}_{\mathrm{Sen}} = \mathrm{median}\left\{D_{ij} : 1 \le i < j \le n\right\}$$

where t_i is the time associated with Y_i. Under mild conditions, $\widehat{\beta}_{\mathrm{Sen}}$ consistently estimates the linear slope and is robust to outliers. Uncertainty quantification can be obtained via either large-sample approximations connected to the MK statistic or through resampling. In practice, the bootstrap-based intervals (including block bootstrap for serial dependence (Künsch 1989)) are convenient and implemented in widely used tools such as the openair R package (Carslaw & Ropkins 2012). These tools admit standard extensions to handle seasonality and covariates, which are pervasive in air-quality records. de Jesus et al. (2020) applied the Theil-Sen slope estimator to quantify and compare long-term trends in urban $PM_{2.5}$ mass and particle number concentrations across multiple cities.

Regression-based models (linear regression or GAMs) offer a flexible, interpretable way to model trend components together with covariates and interventions, at the cost of stronger modeling choices (functional form, smoothing, and error structure). In contrast, rank-based procedures such as the Mann–Kendall test and the Theil–Sen slope estimator make weaker distributional assumptions and are robust to outliers and monotone transformations, but provide a more limited summary focused on monotonic change and are therefore often used as complementary robustness checks. Comparative studies such as Yue & Pilon (2004) emphasize this trade-off: parametric or semi-parametric approaches can be more efficient when their assumptions are approximately satisfied, whereas rank-based procedures tend to be more robust under non-normality, outliers, and model misspecification. Consequently, in some air-quality applications, linear regression was fitted to the KZ-filtered trend component and the MK plus Theil-Sen procedures were used to infer the trend, as in Yusoff et al. (2019) and Qayyum et al. (2022).

3. Extreme Value Analysis

Unlike air-quality assessments that focus on central tendencies or trends, *extreme value analysis* (EVA) targets the tail behavior of concentration distributions, because high-concentration events often pose the greatest risks to human health. Such extremes can arise from interacting emissions, meteorological stagnation, transport, or biomass burning, and are often not well characterized by models focused on mean behavior. This perspective

also aligns with regulatory practice: for example, the U.S. EPA 24-hour $PM_{2.5}$ standard is based on the 98th percentile of 24-hour concentrations, averaged over 3 years, and the ozone standard is based on the annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years (U.S. Environmental Protection Agency 2025). In a rapidly changing emission-control environment, characterizing extreme events is particularly important.

We review two EVA frameworks which are widely applied in air-quality assessment. The first is the *block-maxima* method, which models the distribution of periodic maxima using the generalized extreme value (GEV) family. The second is the *peaks-over-threshold* (POT) approach, which characterizes exceedances beyond a high threshold through the generalized Pareto distribution (GPD) (Smith 1989, Davison et al. 2012).

3.1. The Block-Maxima Method

The block-maxima method partitions the concentration time series $\{Y_t\}$ into K non-overlapping blocks (often calendar years or seasons) and analyzes the block maxima $Z_k = \max\{Y_t : t \in \text{block } k\}$, $k = 1, \ldots, K$. The block length entails a bias-variance trade-off: shorter blocks yield more maxima but can be more sensitive to seasonality and serial dependence, whereas longer blocks (e.g., annual) provide more stable extremes at the cost of fewer replications for tail inference.

The theoretical foundation of the EVA stems from the Fisher-Tippett theorem (Fisher & Tippett 1928), which states that properly normalized maxima of i.i.d. random variables converge in distribution to a generalized extreme-value (GEV) law. Although i.i.d. assumptions are idealized, analogous limits hold for weakly dependent stationary sequences under standard conditions (Leadbetter et al. 1983, Hsing et al. 1988). Accordingly, block maxima from non-overlapping blocks are often treated as approximately independent for likelihood-based inference in environmental studies.

The GEV cumulative distribution function (CDF) is

$$G(z;\mu,\sigma,\xi) = \begin{cases} \exp\Big\{ -[1+\xi(z-\mu)/\sigma]_+^{-1/\xi} \Big\}, & \text{if } \xi \neq 0, \\ \exp\Big\{ -\exp\big[-(z-\mu)/\sigma\big] \Big\}, & \text{if } \xi = 0, \end{cases} \tag{3.1.}$$

where $[a]_+ = \max(0, a)$, $\mu \in \mathbb{R}$ is the *location* parameter, $\sigma > 0$ is the *scale* parameter, and $\xi \in \mathbb{R}$ is the *shape* parameter governing tail heaviness. Parameters are typically estimated by the maximum likelihood, treating $\{Z_k\}$ as an independent $G(z;\mu,\sigma,\xi)$ sample. Uncertainty is quantified via the observed Fisher information.

The *return level* is a standard summary that translates fitted tail behavior into a measure of rareness. For block maxima at a given time scale, the return level associated with a return period of T blocks is the $1-T^{-1}$ quantile z_T that satisfies $G(z_T;\mu,\sigma,\xi) = 1-T^{-1}$. Return levels support statements such as "a once-in-10-year annual maximum $PM_{2.5}$ concentration". For annual maxima, the 10-year return level is the 90th percentile of the annual maximum distribution **with** $T = 10$.

The GEV modeling of block maxima is primarily operational through return levels and return periods for extreme pollution events. Su et al. (2012) fitted GEV models to the upper tail of VOC exposure measurements to quantify the risk of unusually high exposures. Barthwal & Acharya (2022) compared classical extreme-value distributions (Gumbel/Fréchet/Weibull) for urban AQI peaks and used the selected model to report exceeding probabilities for high AQI thresholds together with return-period summaries. Erçelebi &

Toros (2009) fitted a Gumbel (Type I) model to the maxima of SO_2 and NO_2 in Istanbul to assess extreme pollution risk and predict future peaks, and Garbatov et al. (2022) modeled wintertime NO_x extremes near the Varna seaport based on the Weibull distribution to estimate the return periods of high-concentration events.

3.2. The Peaks-Over-Threshold Method

In environmental assessment, modeling of high concentrations over a threshold level is desired. The *peaks-over-threshold* (POT) approach retains exceedances above a suitably high threshold u, thereby making fuller use of the tail observations (Coles 2001).

For a pollution series $\{Y_t\}$, given a high threshold u, define the exceedances $E_t = Y_t - u$ conditional on $Y_t > u$. Under broad regularity conditions, as u increases toward the upper endpoint, the conditional distribution of excesses converges to a generalized Pareto distribution (GPD) (Pickands 1975) for $u > 0$:

$$\Pr(E_t \le e \mid Y_t > u) = H(e; \sigma_u, \xi) = \begin{cases} 1 - \left(1 + \dfrac{\xi e}{\sigma_u}\right)_+^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\dfrac{e}{\sigma_u}\right), & \text{if } \xi = 0, \end{cases} \tag{3.2.}$$

where $\sigma_u > 0$ is a threshold-dependent scale parameter and ξ is the shape (tail) parameter. The GPD shares the same tail index ξ as the corresponding GEV limit for the block maxima, providing a coherent tail characterization across the two frameworks.

GPD-based inference typically relies on exceedances being approximately independent; however, in air-pollution time series, exceedances often occur in temporally clustered episodes due to serial dependence. A standard remedy is *declustering*, which summarizes each exceedance episode by a single representative before fitting a tail model; widely used approaches include runs-based rules and related cluster-inference ideas (Davison & Smith 1990, Ferro & Segers 2003). As an application, Masseran & Mohd Safari (2021) applied the *runs declustering* method to obtain an approximately independent sample of extreme events for return-level inference in a case study of unhealthy air-pollution episodes in Malaysia.

POT is closely aligned with regulatory and health-relevant questions framed in terms of exceedances over prescribed limits. A practical issue is threshold selection: if u is too low, the GPD approximation can be biased; if u is too high, too few exceedances remain for stable estimation. In air-quality applications, u is commonly chosen either as a policy-relevant cutoff or as a high-quantile threshold. For example, Al-Dhurafi et al. (2018) modeled Air Pollution Index (API) exceedances above the unhealthy limit $u = 100$ using a POT/GPD approach for three urban areas in Peninsular Malaysia (2005-2014). Martins et al. (2017) set the POT threshold at the 95th percentile when fitting GPD models to hourly pollutant concentrations in the Sao Paulo and Rio de Janeiro metropolitan regions to compare return periods of extreme events. Gouldsbrough et al. (2022) adopted a high-quantile threshold (around the 90th percentile) for POT modeling of UK surface ozone.

When no clear policy cutoff or domain-specific guidance is available for choosing u, practitioners often use diagnostic checks—most commonly mean residual life and parameter-stability plots—to identify a high-threshold range over which fitted tail parameters are reasonably stable (Coles 2001, Scarrott & MacDonald 2012). For example, Masseran & Safari (2020) used mean residual life diagnostics to support a GPD fit for exceedances when constructing risk summaries.

3.3. Nonstationary and Extended Models

Classical EVA is most naturally formulated for stationary processes, under which tail parameters remain constant across blocks or exceedances. In air-quality applications, however, extremes can evolve with changing meteorology, emissions, and policy interventions, motivating *nonstationary* EVA in which parameters depend on covariates x. To accommodate nonstationarity, the block-maxima (GEV) framework typically allows the location and/or scale parameters to depend on covariates x (Coles 2001), for example

$$\mu(x) = \beta_0 + \beta^\top x, \qquad \log \sigma(x) = \gamma_0 + \gamma^\top x,$$

or via smooth components when nonlinear effects are expected. Analogously, in the POT–GPD framework, the scale $\sigma_u(x)$ is often linked to x while the shape parameter ξ is kept constant (Davison et al. 2012). Empirically, linking tail parameters to covariates can improve the fit and enable physical interpretation. For surface ozone, Eastoe & Tawn (2009) incorporated seasonal and meteorological covariates for the extreme-value parameters to assess regime-dependent shifts in ozone extremes. Garcia-Aristizabal et al. (2015) linked tail behavior to time- and scenario-dependent climate covariates to evaluate how risk evolves under projected climate conditions and to propagate uncertainty in multi-hazard assessments. Moreover, Gyarmati-Szabó et al. (2017) fitted a POT model with a covariate-dependent scale parameter, quantifying how traffic and meteorological drivers modulate the magnitude of extreme pollution events.

Beyond covariate-driven nonstationarity in marginal tails, environmental extremes often involve dependence across time, variables, and space. Temporal clustering can be summarized via the extremal index (Ferro & Segers 2003). For multivariate extremes, conditional extremes models characterize tail dependence by modeling other variables given that one component is extreme (Heffernan & Tawn 2004). Spatial and spatiotemporal extensions further allow tail behavior to vary with location and time (Casson & Coles 1999, Wang et al. 2023). Together, these developments broaden the classical EVA toolkit toward more realistic representations of interdependent extreme events in air-quality applications.

4. Spatiotemporal Models

The spatiotemporal approach treats the observed concentrations as error-perturbed measurements of an underlying latent spatiotemporal process. This state-space form yields a framework for dynamic spatial mapping and uncertainty quantification. This section reviews the dynamic geostatistical state-space family and summarizes their adaptations for multivariate pollutants modeling and forecasting.

Huang & Cressie (1996) introduced the Hidden Dynamic Geostatistical (HDG) model for univariate state variables, which served as a baseline model in many comparative air quality studies (e.g., Cameletti et al. 2011). Conceptually, it treats air-quality observations as noisy measurements of a latent spatiotemporal field, whose temporal evolution is typically modeled by an autoregressive process AR(p) with spatially correlated innovations. Building on the HDG framework and the dimension-reduction ideas for large space–time problems (Wikle & Cressie 1999), Fassò et al. (2007) proposed the *Geostatistical Dynamical Calibration* (GDC) model to accommodate heterogeneous monitoring networks where instruments may exhibit systematic additive and/or multiplicative biases. The GDC can be viewed as a calibrated extension of the HDG framework tailored for heterogeneous monitoring networks. Specifically, it introduces a calibration layer that aligns measurements

from different instrument classes onto a common scale while preserving the reduced-rank state-space structure.

Let $Y_t(s)$ and $Z_t(s)$ be the observed and the latent "true" concentration at site s and time t, respectively, and $\mathbf{X}_t(s)$ contain a set of exogenous covariates. The calibrated observation equation of the GDC is

$$Y_t(s) \;=\; \alpha + \beta\, Z_t(s) \;+\; \mathbf{X}_t(s)^\top \boldsymbol{\gamma} \;+\; \varepsilon_t(s), \qquad \varepsilon_t(s) \sim \mathcal{N}\big(0, \sigma^2(s)\big), \tag{4.1.}$$

where (α, β) are calibration coefficients capturing additive and multiplicative bias, γ collects the coefficients to the exogenous covariates, and $\varepsilon_t(s)$ is an independent Gaussian white noise with mean zero and standard deviation $\sigma(s)$. Equation 4.1. represents a measurement equation where biases are corrected via the calibration parameters (α, β). In the heterogeneous-network setting considered by Fassò et al. (2007), they allowed α to drift over time such that $\alpha = \alpha(t)$, while β is treated as a constant. The bias-free reference sites typically had $\alpha = 0$ and $\beta = 1$, where the non-reference sites were calibrated with estimated (α, β).

To model spatiotemporal dependence on a common scale, GDC represents the latent field via a low-dimensional dynamic component plus a fine-scale residual:

$$Z_t(s) \;=\; \mathbf{B}(s)^\top \boldsymbol{\mu}_t \;+\; U_t(s), \tag{4.2.}$$

where $\mathbf{B}(s) \in \mathbb{R}^r$ is a set of spatial basis functions obtained from empirical orthogonal functions (EOF) decomposition. This low-rank representation via EOF basis functions substantially reduces the computational burden in large monitoring networks, thereby retaining computational tractability for high-dimensional space-time data. $\boldsymbol{\mu}_t \in \mathbb{R}^r$ is a shared state vector evolving as

$$\boldsymbol{\mu}_t \;=\; \mathbf{G}\, \boldsymbol{\mu}_{t-1} \;+\; \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{4.3.}$$

where $\mathbf{G}$ and $\boldsymbol{\Sigma}_\eta$ are often diagonal, reflecting the orthogonality of EOF components. The residuals $U_t(\cdot)$ capture the remaining fine-scale spatial structure and are often modeled as temporally independent, second-order stationary, and isotropic in space with covariance

$$\mathrm{Cov}\{U_t(s), U_t(s')\} \;=\; \sigma_U^2\, C_\theta\big(\|s - s'\|\big),$$

where $C_\theta(h)$ is a correlation function, often chosen to be exponential or Mat'ern class. For example, under an exponential correlation, $C_\theta(h) = \exp(-\theta h)$, with the range parameter $\theta > 0$ to be estimated.

As another extension of the HDG framework, Calculli et al. (2015) proposed a *multivariate* HDG model for modeling multiple pollutants, which can be viewed as a vector-valued counterpart of the hierarchical state-space template in Fassò & Cameletti (2009). Let $\mathbf{Y}_t(s) = \big(Y_t^{(1)}(s), \ldots, Y_t^{(q)}(s)\big)^\top$ be the q-dimensional pollutant vector such that

$$\begin{aligned} \mathbf{Y}_t(s) &= \mathbf{X}_{\beta,t}(s)\boldsymbol{\beta} \;+\; \mathbf{X}_{z,t}(s)\mathbf{A}\mathbf{Z}_t(s) \;+\; \boldsymbol{\varepsilon}_t(s), \\ \mathbf{Z}_t(s) &= \mathbf{G}\mathbf{Z}_{t-1}(s) + \boldsymbol{\xi}_t(s), \end{aligned} \tag{4.4.}$$

where $\mathbf{X}_{\beta,t}(s)$ is a known *fixed-effects* design matrix containing intercept and covariates with regression coefficients $\boldsymbol{\beta}$, and $\mathbf{X}_{z,t}(s)$ is a known *random-effects* design matrix that links the latent field to the available observations (often a weighting matrix). The matrix

$\mathbf{A}$ is a (typically diagonal) loading matrix that maps the latent field to the observation scale, and $\mathbf{Z}_t(s)$ is a latent q-variate dynamic field, $\boldsymbol{\varepsilon}_t(s)$ denotes Gaussian measurement error (with diagonal covariance), and $\mathbf{G}$ is typically diagonal for parsimony. Cross-pollutant dependence is captured through the innovation $\boldsymbol{\xi}_t(\cdot)$, a temporally independent multivariate Gaussian random field with

$$\mathrm{Cov}\{\boldsymbol{\xi}_t(s), \boldsymbol{\xi}_t(s')\} = \mathbf{V}\, C_\theta(\|s - s'\|),$$

so that dependence across pollutants is built via the cross-variable covariance matrix $\mathbf{V}$, a feature particularly useful for non-collocated (spatially misaligned) monitoring networks.

The hierarchical state-space framework allows likelihood-based inference for large monitoring networks with long time series where parameters are estimated by the likelihood approach via the EM algorithms (Fassò & Cameletti 2009). In this workflow, the Kalman smoothing provides the conditional moments of the low-dimensional latent state $\boldsymbol{\mu}_t$ required in the E-step, which naturally accommodates missing data. Software such as `D-STEM` and Wang et al. (2021) make these workflows operational by providing end-to-end routines for model fitting, dynamic spatial prediction/mapping, and uncertainty quantification on complex space–time monitoring data. For short range analysis, one can enrich the temporal component (e.g., response-side lags or shared temporal factors) while retaining the same state-space logic; for example, Wan et al. (2021) developed a dynamic spatiotemporal model for hourly $PM_{2.5}$ in Beijing that incorporates meteorology and geographically structured effects for short-horizon prediction.

These developments suggest that the HDG/GDC framework functions as a flexible state-space backbone rather than a fixed model specification. It can be adapted along different scientific and operational goals: enlarging the latent state to couple multiple pollutants; enriching the temporal dynamics (e.g., response-side lags) for forecasting; and incorporating structured spatial effects to reflect geographic drivers beyond a single global coefficient. These modifications preserve the core probabilistic state-space logic while allowing the model to better align with scientific objectives and decision-making needs. This backbone perspective also opens avenues for further methodological developments in scalable inference and nonlinear extensions. Its flexibility, together with computationally efficient likelihood-based inference via EM algorithms and Kalman smoothing, has facilitated a wide range of downstream applications in air quality research.

For population exposure assessment, Fassò et al. (2016) used the multivariate HDG model to reconstruct daily multi-pollutant (CO, NO_2, O_3, PM_{10}, and $PM_{2.5}$) surfaces over Europe from 2009 to 2011 with heterogeneous monitoring coverage and derived population exposure metrics. Finazzi et al. (2019) proposed a statistical emulator based on `D-STEM` for vector-valued model outputs with missing components (illustrated using yearly-average NO_2 predictions from an ADMS-Urban dispersion model), enabling fast predictions of partially observed output vectors with uncertainty quantification. For policy evaluation, Fassò et al. (2022) coupled the multivariate HDG model with variable-selection ideas to assess the air-quality impacts of the COVID-19 lockdown (Lombardy, Italy; PM_{10}, $PM_{2.5}$, and nitrogen oxides). Maranzano et al. (2023) developed a *functional* HDG model in which high-frequency pollutant trajectories are treated as functional responses (e.g., hourly intraday concentration profiles), with adaptive LASSO regularization to perform structured covariate selection in high-dimensional functional settings.

5. Methods for Inferring Emissions Under Meteorological Confounding

The key to air quality management is to reduce underlying emissions, which requires timely and accurate emission measurements. Emission inventories are a traditional tool for quantifying anthropogenic emissions, which compile information on industrial activity, energy use, and population statistics and then downscale these data to a desired spatial and temporal resolution (U.S. Environmental Protection Agency 2017). Emission inventories (Crippa et al. 2020) are key inputs to numerical atmospheric models, including the Community Multiscale Air Quality Modeling System (CMAQ) and other chemical transport models (CTMs). These numerical models simulate atmospheric chemical and physical processes, enabling dynamic forecasting and data assimilation. Errors in emission inventories can be a major source of uncertainty in model-based forecasts and assessments, including emission-fixed scenario analyses used to separate meteorological versus emission contributions (Wang et al. 2019, Zhang et al. 2019). Indeed, inventories are subject to measurement and reporting errors, and their compilation typically involves a delay of three to four years. Such delays limit their utility for timely air quality management and can undermine the accuracy of model-based assessments.

These limitations, together with the growing availability of high-frequency monitoring data on both air quality and meteorology, have motivated approaches that infer emission-related signals from concentration observations while adjusting for meteorological confounding (Xiao et al. 2021). This section reviews methods for extracting emission-related signals from pollutant concentrations measured at monitoring sites by removing meteorological variability.

5.1. Parametric Deweathering Methods

As shown in Section 2, air-quality studies routinely include meteorological covariates in the regression model to reduce weather-driven variability. Thompson et al. (2001) surveys regression- and smoothing-based meteorological adjustment strategies for ozone trend assessment and highlights practical issues such as covariate selection and serial dependence.

Residualization treats meteorology as a predictor of concentration Y_t as in 2.1,

$$Y_t \;=\; \beta_0 \;+\; \mathbf{X}_t^{\top}\boldsymbol{\gamma} + \varepsilon_t, \tag{5.1.}$$

where X_t collects meteorological covariates. The deweathered series is $\widetilde{Y}_t \;=\; Y_t - \mathbf{X}_t^{\top}\hat{\boldsymbol{\gamma}}$ after obtaining the OLS estimate $\hat{\boldsymbol{\gamma}}$. Changes in $\widetilde{Y}_t$ over time (or between different time periods) are attributed to variations not explained by the meteorology. Equivalently, this corresponds to comparing fitted means under a fixed *reference meteorology* $\mathbf{X}^{\star}$. Li et al. (2018) modeled summer surface O_3 anomalies with flexible meteorological terms (including interactions) and attributed the residual linear trend to anthropogenic emission changes during 2013-2017. Zhai et al. (2019) performed an analogous adjustment for deseasonalized $PM_{2.5}$ anomalies and interpreted the residual trend as an emissions-related contribution to China's $PM_{2.5}$ decline over 2013-2018.

Camalier et al. (2007) proposed a related meteorological adjustment approach by fitting a generalized additive model (GAM) for the daily ozone metric Y_t (e.g., MDA8 O_3). Let doy_t and $k(t)$ denote the day-of-year and the calendar year associated with time index t, respectively, and $\mu_t = \mathbb{E}(Y_t \mid \mathbf{X}_t, \mathrm{doy}_t, k(t))$ denote the conditional mean. With a known monotone link function $g(\cdot)$ (e.g., the log link in the EPA implementation), the model

specifies the additive linear predictor

$$g(\mu_t) = \beta_0 + s_{\text{season}}(\text{doy}_t) + \sum_{j=1}^{p} s_j(X_{j,t}) + E_{k(t)} \tag{5.2.}$$

where, in notation consistent with Eq 2.3., $s_j(\cdot)$ are smooth functions of meteorological variables and $E_{k(t)}$ denotes the calendar year effect related to t. After estimating the parameters and the functions, a deweathered series $\widehat{Y}_k^{\text{adj}} = g^{-1}\left(\widehat{\beta}_0 + \widehat{E}_k\right)$, was proposed by removing the doy_t and the meteorological effects. It is apparent that this approach attempts to mimic the linear deweathering approach. However, it is unclear what the resulting $\widehat{Y}_k^{\text{adj}}$ represents statistically, since the inverse link $g^{-1}(\cdot)$ is nonlinear, which hampers its interpretability.

5.2. Meteorological Adjustment Methods

The deweathering methods above do not directly yield overall distributional summaries, such as means or quantiles. Meteorologically adjusted versions of these summaries are the focus of this subsection. Rather than subtracting a fitted meteorological component at each time point, these approaches explicitly define a weather-adjusted estimand (e.g., a season-specific mean under a common meteorological baseline distribution) and provide uncertainty quantification for the resulting measures of concentration, which are indicative of the underlying emission.

Liang et al. (2015) proposed meteorologically adjusted means and quantiles of pollution concentration distributions in the context of air quality assessment in Beijing. Let Y_{ijt} denote the concentration in year i, **a temporal stratum** j (e.g., a month or season), and hour t, X_{ijt} collect continuous meteorological covariates and W_{ijt} denote discrete wind-direction. The adjustment begins with a nonparametric model

$$Y_{ijt} = m_{ij}(X_{ijt}, W_{ijt}) + \varepsilon_{ijt}, \tag{5.3.}$$

where m_{ij} is the year-i, stratum-j conditional mean response to meteorology, and ε_{ijt} captures random variation. Year-specific **emission** shifts at fixed (X, W) are absorbed into m_{ij}. Let f_{ij} denote the density of (X, W) in year-i. The meteorologically adjusted stratum mean in year-i is

$$\mu_{ij} = \sum_{w} \int m_{ij}(x, w)\, f_{\cdot j}(x, w)\, dx, \tag{5.4.}$$

where $f_{\cdot j}(x, w) = n_{\cdot j}^{-1} \sum_{i=1}^{n_{\cdot j}} f_{ij}(x, w)$ is a reference distribution of (X, W) constructed across $n_{\cdot j}$ years *within stratum* j. For a fixed j, all μ_{ij} are standardized to the same $f_{\cdot j}$, making cross-year means $\{\mu_{ij}\}$ comparable across different years. Specifically, the difference between years i and i', $\mu_{ij} - \mu_{i'j}$, would reflect changes in the underlying emission embedded in the response surface m_{ij} and $m_{i'j}$. The same goal cannot be achieved if the single year meteorological density f_{ij} is used in 5.4 as the measure would be confounded by the weather conditions in the i-th year.

The same adjustment can be made for quantiles. Let

$$F_{ij}(y \mid x, w) = \Pr(Y_{ijt} \le y \mid X_{ijt} = x,\ W_{ijt} = w)$$

be the conditional concentration distribution in year-i and stratum-j. A meteorologically-adjusted distribution function is $F_{ij}^{\mathrm{ref}}(y) = \sum_w \int F_{ij}(y \mid x, w)\, f_{\cdot j}(x, w)\, dx$, and the adjusted τ-quantile for a $\tau \in (0,1)$ is $q_{ij}^{\mathrm{ref}}(\tau) = \inf\{y : F_{ij}^{\mathrm{ref}}(y) \geq \tau\}$.

Let $a = 1, \ldots, n_{\cdot j}$ index the years in which stratum j is observed, and n_{aj} be the number of observations in year a of stratum j. The pooled reference sample is $\{(X_{ajt}, W_{ajt})\}_{t=1}^{n_{aj}}$ for $a = 1, \cdots, n_{\cdot j}$. Empirically, Liang et al. (2015) fitted $\widehat{m}_{ij}$ nonparametrically via multivariate kernel regression and computed an estimate of 5.4. by the plug-in principle

$$\widehat{\mu}_{ij} = \left(\sum_{a=1}^{n_{\cdot j}} n_{aj} \right)^{-1} \sum_{a=1}^{n_{\cdot j}} \sum_{t=1}^{n_{aj}} \sum_{w} \widehat{m}_{ij}\big(X_{ajt}, w\big)\, \mathbf{1}\{W_{ajt} = w\},$$

which evaluates $\widehat{m}_{ij}$ over all stratum-j's meteorological state (X_{ajt}, W_{ajt}) of a year to obtain the counter-factual average of the year (inner two summations), and then averages over the $n_{\cdot j}$ years. A strength of the meteorological adjustment is that it directly targets interpretable distributional measures over comparable aggregates. Under standard smoothness and bandwidth conditions, $\widehat{\mu}_{ij}$ is consistent and supports uncertainty quantification via either asymptotic variance formulas or dependence-respecting resampling (e.g., the block bootstrap), and the estimation of the adjusted $F_{ij}^{\mathrm{ref}}(y)$ and thus $q_{ij}^{\mathrm{ref}}(\tau)$ is analogous (Liang et al. 2015).

Subsequent work extended this temporal-adjustment framework to a spatiotemporal setting. To support meaningful comparisons across both time and space, Zhang et al. (2017) allowed the meteorological response surface to vary by site location s and standardized each site-year-stratum mean to a common baseline meteorological distribution. Specifically, they defined the spatially and temporally standardized mean

$$\mu_{ij}(s) = \sum_{w} \int m_{ij}(x, w; s)\, f_{\cdot j}(x, w)\, dx,$$

where the baseline meteorological density $f_{\cdot j}(x, w)$ is the equally weighted mixture of the stratum-j densities across years and meteorological sites,

$$f_{\cdot j}(x, w) = \frac{1}{A\,|\mathcal{M}|} \sum_{a=1}^{A} \sum_{s \in \mathcal{M}} f_{aj}(x, w; s),$$

with $f_{aj}(x, w; s)$ denoting the joint density of (X, W) in year a and stratum-j at meteorological site s, A is the total number of years of available weather data, and $\mathcal{M}$ is the set of meteorological sites in the region. Zhang et al. (2020) further established large-sample guarantees for this spatiotemporal adjustment method, clarifying how bandwidth choice and within-stratum homogeneity affect consistency and asymptotic normality for standardized regional means.

As an important extension, the adjusted mean and quantiles in Liang et al. (2015) and Zhang et al. (2017) can be generalized to obtain an adjusted curve with respect to certain variables. Chen et al. (2018) defined the *adjusted curves* **by adjusting only part of the meteorological covariate while leaving a subset of the covariate fixed**. Write the continuous meteorological covariates as $X_{ijt} = \big(X_{ijt}^{(k)}, X_{ijt}^{(-k)}\big)$, where $X^{(k)}$ is our target variable. Under the stratum-j baseline distribution $f_{\cdot j}$, the adjusted curve function with respect to $X^{(k)}$ is defined by

$$\mu_{ij}(x^{(k)}) = \sum_{w} \int m_{ij}\big((x^{(k)}, x^{(-k)}), w\big)\, f_{\cdot j}\big(x^{(k)}, x^{(-k)}, w\big)\, dx^{(-k)}. \qquad 5.5.$$

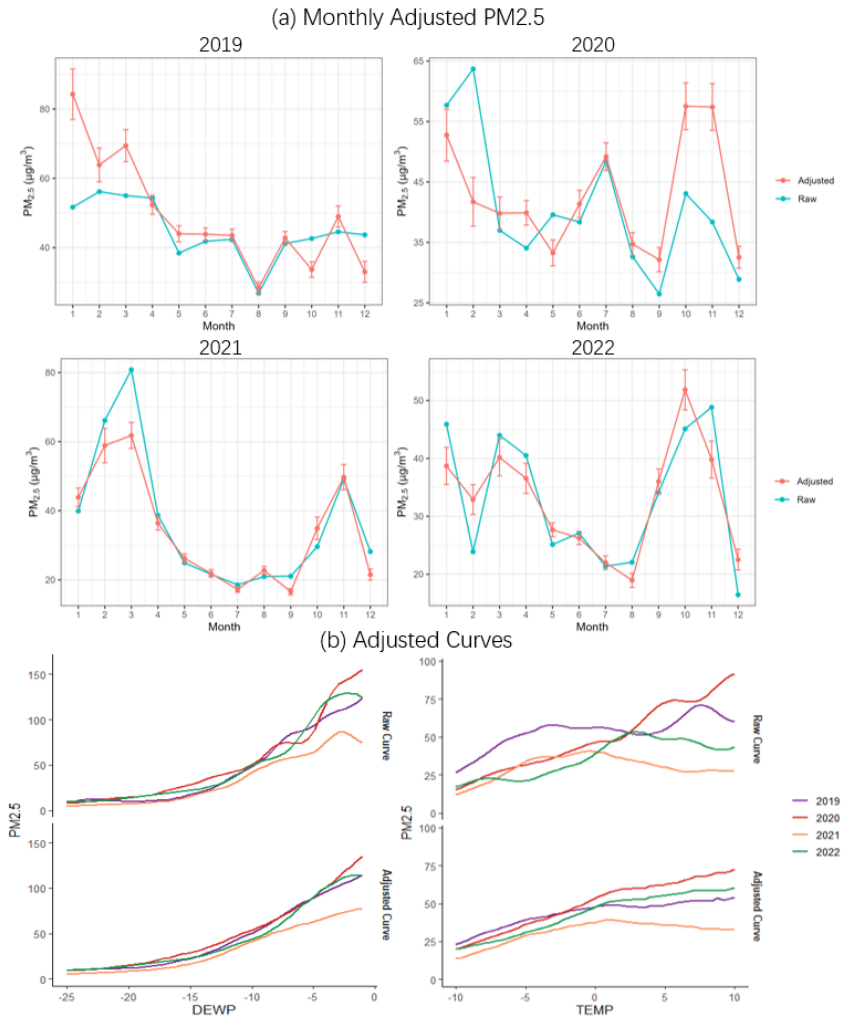

Figure 2

Meteorological adjustment at the Beijing Aotizhongxin site (2019–2022). (a) Monthly mean PM2.5: raw (cyan) and meteorologically adjusted (red); error bars denote approximate 95% bootstrap confidence intervals for the adjusted means. (b) Raw and meteorologically adjusted PM2.5 response curves versus dew point and temperature for the winter season.

where the integral is taken over $x^{(-k)}$, treating $x^{(k)}$ as fixed. This construction is analogous to a partial dependence function under a reference covariate distribution (Friedman 2001). Empirically, $\mu_{ij}(X^{(k)})$ can be estimated by averaging $\widehat{m}_{ij}(X_{ajt}, W_{ajt})$ over the pooled baseline sample restricted to a neighborhood of $X^{(k)}$: $A\big(X^{(k)}, b\big) = \big\{x = (\cdots, x^{(k)}, \cdots) : X^{(k)} - b \le x^{(k)} \le X^{(k)} + b\big\}$, where $b > 0$ is a smoothing parameter.

Figure 2 demonstrates the kernel-based meteorological adjustment based on air quality data from the Beijing Aotizhongxin site over 2019-2022. From Panel (a), the adjusted and raw monthly means differ substantially in winter months, indicating that month-to-month

variability in the observed series can be strongly modulated by contemporaneous meteorology. In early 2019 and 2020, for example, raw $PM_{2.5}$ in January-February increases, whereas the adjusted means decline, consistent with unfavorable winter weather elevating observed concentrations relative to a standardized meteorological baseline. Panel (b) compares the raw and adjusted curves against dew point and temperature in four winters. The raw curves are simply the marginal kernel regression estimation of PM2.5 with respect to the variable. The raw curves appear largely variable and not monotonic across years. After standardizing the remaining weather variables to a common winter baseline, the adjusted curves become smoother (with reduced curvature) and more stable across years.

A closely related approach is the *meteorological normalization* (MN) proposed by Grange et al. (2018). In particular, after fitting a predictive model for concentrations, it constructs a deweathered time series by holding the time/trend index fixed and averaging predictions over a reference distribution of resampled covariates, typically via Monte Carlo resampling. Let z_t denote the time (or trend) representation to be held fixed (e.g., a running index), and let X_t collect the covariates to be resampled (often dominated by meteorological and circulation variables). For any fitted predictor $\widehat{m}(\cdot)$, MN defines a normalized series by Monte Carlo averaging,

$$\widehat{\mu}^{\mathrm{ref}}(t) = \frac{1}{B}\sum_{b=1}^{B} \widehat{m}\big(X_t^{(b)}, z_t\big), \qquad \{X_t^{(b)}\}_{t=1}^{n} \sim \widehat{F}^{\mathrm{ref}},$$

where $\widehat{F}_X^{\mathrm{ref}}$ is an empirical reference distribution. In practice, $\{X_t^{(b)}\}$ is often generated by drawing a random permutation (i.e., a without-replacement reallocation) of the reference covariate vectors for each Monte Carlo replicate. To support interpretability, reproducibility, and cross-study comparison, MN studies should explicitly report (i) the learner $\widehat{m}$, (ii) the reference set defining $\widehat{F}_X^{\mathrm{ref}}$, and (iii) the time representation z_t. By construction, MN yields a *time-resolved* deweathered trajectory $\{\widehat{\mu}^{\mathrm{ref}}(t)\}_{t=1}^{n}$ rather than a single adjusted aggregate for each stratum. Conceptually, MN targets the same estimand as the kernel-based adjustment proposed in Chen et al. (2018): it fixes the time index (i.e., t) and marginalizes over the distribution of the other covariates, so the MN series can be viewed as an adjusted curve with respect to time.

The estimator $\widehat{m}$ may be specified parametrically for interpretability, or more flexibly as semiparametric or nonparametric model (e.g., random forests, gradient boosting) to capture nonlinearities and interactions, at the cost of increased sensitivity to modeling and normalization design choices; see Qi et al. (2024) for a comparative study. As an application, Shi et al. (2021) applied the random forest based MN via the rmweather package (Grange & Carslaw 2019) to deweather urban NO_2, O_3, and $PM_{2.5}$ series during the COVID-19 lockdowns and quantify the intervention-related changes. Additional benchmarking works have evaluated various deweathering methods against controlled chemistry-transport model experiments and highlighted the sensitivity of inferred trends to model class, hyperparameter tuning, and resampling design (e.g. Qiu et al. (2022), Zheng et al. (2023).)

A key design difference from stratum-based meteorological adjustment is the construction of the baseline. In Liang et al. (2015) and Chen et al. (2018), standardization is performed *within* each stratum j (or a restricted neighborhood) using a stratum-specific reference distribution $f_{\cdot j}$, so that comparisons are made on a common meteorological footing within the same period. In contrast, MN is often implemented using a single (global) reference set pooled over the entire study period, yielding a time series standardized to

an overall meteorological baseline. This choice increases the effective reference sample size and targets a smooth deweathered time series, but it can force the model to predict under meteorological combinations that rarely (or never) occur at time t, leading to unreliable extrapolation. Accordingly, stratified MN (e.g., season- or window-based resampling) is sometimes used to improve its performance (Vu et al. 2019).

Although terminology varies in the literature (*deweathering*, *weather correction*, *meteorological adjustment*, or *meteorological normalization*), these approaches share a common goal in removing meteorological confounding to enable emission-relevant comparisons. Despite differences in targets and implementation, they can all be placed within the unified framework of meteorological adjustment in Liang et al. (2015), where one first fits a pollution concentration-meteorology relationship via the conditional mean (regression) model (possibly including seasonality, year effect, or other time features), and then defines the quantity as a *standardized* measure obtained by integrating the fitted conditional mean over a fixed reference meteorological distribution. In practice, this standardization is implemented with an empirical reference distribution, so the integral reduces to a simple plug-in estimator obtained by evaluating the fitted regression of pollution levels on meteorological covariates at the pooled reference-meteorology samples. Different methods mainly differ in (i) the temporal indexing and resolution of the estimation (stratum j versus native time step t), (ii) how the reference weather is constructed (stratum-specific versus global), and (iii) the model class used to estimate the conditional mean model (parametric versus non-parametric).

6. Assessment Methods based on Pollution Episodes

The meteorological adjustment outlined in Section 5.2 yields temporally and/or spatially comparable measures of the underlying emissions. However, it is done quite passively, as it uses all observations rather than selectively focusing on periods that are especially informative about local emissions. This section presents a more targeted strategy that identifies calm episodes to estimate local emission intensity. The basic rationale is that emission intensity tends to evolve more slowly than meteorological conditions, so short-term fluctuations in observed pollution are often driven primarily by meteorological cycles. A typical episode consists of four stages: (i) a post-cleaning calm period, with weak winds and little precipitation, during which local emissions can dominate the accumulation process (Zhu et al. 2021); (ii) a regional transport-favored growth stage characterized by sustained winds from high-emission regions leading to a rapid rise in concentrations; (iii) an air stagnation period (Zhang et al. 2023) with weak dispersion, during which $PM_{2.5}$ often reaches its peak; and (iv) a removal stage triggered by wind and/or precipitation processes. When the removal stage ends, the next episode may begin.

Zhu et al. (2021) identified *calm episodes* as the periods after sustained northerly cleansing but before the arrival of transported pollutants and used data in the calm episodes to estimate local emission levels. This can be viewed as a data experiment for identifying local emissions. Motivated by the well-documented "south wind-polluting/north wind-cleaning" pattern in North China (Tie et al. 2017, Wang et al. 2024), they operationalized the data experiments by partitioning the wind direction to two broad categories: northerly (N) and southerly (S), and defined the cumulative wind speeds that reset to zero when the wind direction changed. Using proper threshold levels for the two broad cumulative wind directions, they selected post-cleaning calm periods that are dominated by local accumulation.

They fitted a linear panel regression model over the selected calm episodes. For each year $i = 1, \ldots, I$, let $j = 1, \ldots, n_i$ index the episodes within year i, and let $t = 0, 1, \ldots, T_{ij}$ denote the hour within episode j, where T_{ij} is the duration (in hours) of episode j. Define the concentration increment relative to the beginning of episode j as $\Delta Y_{ijt} = Y_{ijt} - Y_{ij0}$. The model is specified as

$$\Delta Y_{ijt} = \alpha_i + \boldsymbol{\beta}_i^\top \Delta \mathbf{M}_{ijt} + \boldsymbol{\gamma}_i^\top \mathbf{C}_{ij} + \sum_{k=1}^{T_{ij}} \eta_{ik} \mathbf{1}(t = k) + \varepsilon_{ijt},$$

where $\Delta \mathbf{M}_{ijt}$ represents meteorological variations, $\mathbf{C}_{ij}$ captures episode-level controls. Under the assumption that regional transport is weak during calm episodes, the sequence $\{\eta_{ik}\}$ characterizes the within-episode pollutant accumulation trajectory in year i, and its average over the first T hours is interpreted as an intensity measure of local emissions. To enable interannual comparison, they adopted the meteorological adjustment outlined in Section 5.2 to obtain the meteorologically adjusted growth rate estimates.

Building on the calm-period approach, Luo et al. (2022) extended it to cover the entire pollution episodes via an automated, meteorology-guided episode selection algorithm. In particular, after a calm period, their procedure terminated a *pollution episode* at the subsequent cleaning event that brings $PM_{2.5}$ back to a low (clean) level. The selected episodes are summarized by pollution loading (cumulative $PM_{2.5}$ over the episode) and modeled as functions of within-episode and pre- and post-episode meteorological variables using linear regression and random forests. Similarly, Luo et al. (2022) applied a meteorological adjustment approach as discussed in Section 5.2 for interannual comparability. Let $\widehat{m}_i(U, Dur)$ denote the fitted conditional mean pollution loading in year i given episode covariates U and duration Dur. The meteorologically adjusted mean loading is estimated by

$$\widehat{PL}_i^{\,*} = \frac{1}{A} \sum_{a=1}^{A} \frac{1}{n_a} \sum_{j=1}^{n_a} \widehat{m}_i(U_{aj}, Dur_{aj}) \,,$$

where $\{(U_{aj}, Dur_{aj})\}_{j=1}^{n_a}$ are the covariates from the j-th episode in year a, n_a is the number of episodes in year a and A denotes the number of years in the baseline pool.

Figure 3 illustrates the calm episode selection and the estimated local emission growth rate during the calm periods in Beijing. Panel (a) displays a typical episode initiated after a persistent northerly cleaning, with $PM_{2.5}$ falling below 35 $\mu g/m^3$ and then rising rapidly under sustained southerly airflow. Panel (b) shows an alternative cleaning mechanism driven by wet deposition: $PM_{2.5}$ drops below 35 $\mu g/m^3$ during precipitation, and the episode starts after rainfall ends and concludes after the arrival of southerly winds. Panel (c) summarizes the (original and meteorologically adjusted) mean growth curves over the first five hours following the episode onset, highlighting nearly monotonic growth with widening confidence bands. In most situations, the growth pattern was largely linear in the early hours, suggesting that the linear panel model used by Zhu et al. (2021) was reasonable.

7. Conclusion

Air-quality assessment is essential for public health protection and environmental management, and it demands statistical methods that can provide interpretable estimates and calibrate uncertainty quantification under dependence, heavy tails, and strong meteoro-

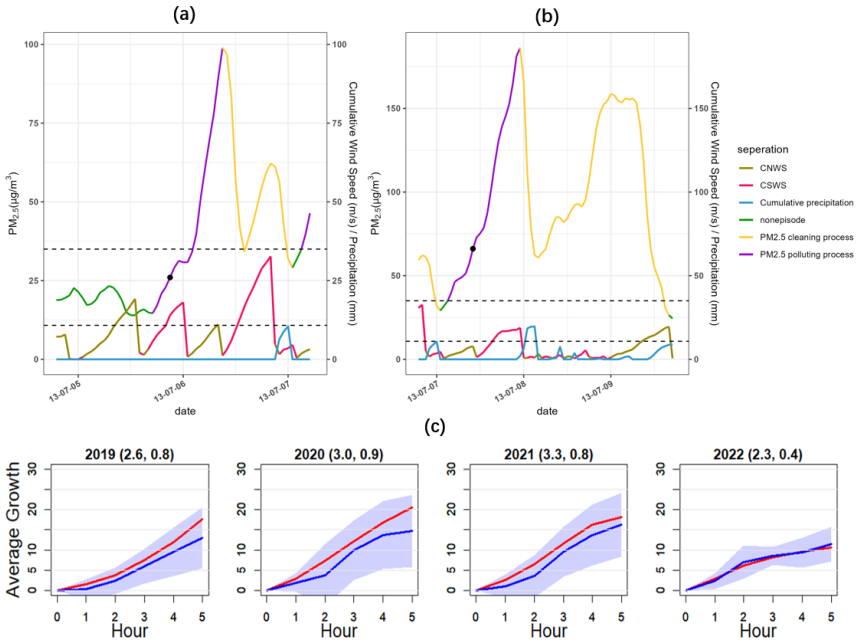

Figure 3

Calm episode selection (Panels (a)-(b)) and the estimated growth rates of local emission during the calm episodes. Panels (a)–(b): two episodes at the Beijing Aotizhongxin site (summer 2013) and the associated cumulative northerly wind speed (CNWS), cumulative southerly wind speed (CSWS), and cumulative precipitation; cleaning and polluting segments are highlighted. The dashed line marks the clean threshold (35 μg/m^3), and the CNWS threshold identifies persistent northerly cleaning. Panel (c): original and meteorologically adjusted mean $PM_{2.5}$ growth curves over the first five hours after episode onset (with 95% CIs for the adjusted curve) at the Beijing Aotizhongxin site (winter 2019-2022); numbers in parentheses report the adjusted 5-hour average growth rates and their standard errors.

logical variability. This review synthesizes widely used methodologies with relatively well-developed statistical foundations, spanning trend and regime characterization, extremes and tail risk, spatiotemporal prediction with uncertainty quantification, and meteorology-standardized comparisons for emission-related inference. We also highlight episodic assessment as a complementary lens that treats high-pollution periods as dynamic units.

The statistical foundations reviewed here provide a principled toolkit for translating complex monitoring records into interpretable, comparable, and uncertainty-quantified summaries. Future progress will benefit from tighter integration across methodological families, particularly by embedding meteorological standardization and tail- or episode-based summaries within scalable spatiotemporal models to produce coherent risk surfaces with calibrated uncertainty. Transfer learning also holds considerable promise for data fusion and for reducing estimation uncertainty, but principled strategies for transferring information across locations and time periods remain underdeveloped. Addressing such transfer will require careful treatment of heterogeneity, including domain adaptation and sample selection from increasingly large and heterogeneous monitoring data sets. At the same

time, emission-oriented conclusions will likely rely more heavily on explicit identification strategies, sensitivity diagnostics, and closer coupling with atmospheric modeling and data assimilation as data sources continue to expand in scale and resolution.

LITERATURE CITED

Abbasi-Kangevari M, Malekpour MR, Masinaei M, Saeedi Moghaddam S, Ghamari SH, et al. 2023. Effect of air pollution on disease burden, mortality, and life expectancy in North Africa and the Middle East: a systematic analysis for the Global Burden of Disease study 2019. *The Lancet Planetary Health* 7(5):e358–e369

Al-Dhurafi NA, Masseran N, Zamzuri ZH, Razali AM. 2018. Modeling unhealthy air pollution index using a peaks-over-threshold method. *Environmental Engineering Science* 35(2):101–110

Aminikhanghahi S, Cook DJ. 2017. A survey of methods for time series change point detection. *Knowledge and Information Systems* 51(2):339–367

Appel KW, Napelenok SL, Foley KM, Pye HOT, Hogrefe C, et al. 2021. The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: System updates and evaluation. *Geoscientific Model Development* 14:2867–2897

Barthwal A, Acharya D. 2022. Performance analysis of sensing-based extreme value models for urban air pollution peaks. *Modeling Earth Systems and Environment* 8:4149–4163

Bey I, Jacob DJ, Yantosca RM, Logan JA, Field BD, et al. 2001. Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research: Atmospheres* 106(D19):23073–23095

Calculli C, Fassò A, Finazzi F, Pollice A, Turnone A. 2015. Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in apulia, Italy. *Environmetrics* 26(6):406–417

Camalier L, Cox WM, Dolwick PD. 2007. The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric Environment* 41(33):7127–7137

Cameletti M, Ignaccolo R, Bande S. 2011. Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* 22(8):985–996

Carslaw DC, Ropkins K. 2012. openair — an R package for air quality data analysis. *Environmental Modelling & Software* 27–28:52–61

Casson E, Coles S. 1999. Spatial regression models for extremes. *Extremes* 1(4):449–468

Chaudhuri S, Dutta D. 2014. Mann–kendall trend of pollutants, temperature, and humidity over an urban station of India with forecast verification using different ARIMA models. *Environmental Monitoring and Assessment* 186(7):4719–4742

Chen L, Guo B, Huang J, He J, Wang H, et al. 2018. Assessing air-quality in Beijing–Tianjin–Hebei region: The method and mixed tales of $PM_{2.5}$ and O_3. *Atmospheric Environment* 193:290–301

Coles S. 2001. An Introduction to Statistical Modeling of Extreme Values. London: Springer

Collaud Coen M, Andrews EJ, Bigi A, Martucci G, Romanens G, et al. 2020. Effects of the prewhitening method, the time granularity, and the time segmentation on the Mann–Kendall trend detection and the associated Sen's slope. *Atmospheric Measurement Techniques* 13(12):6945–6964

Crippa M, Solazzo E, Huang G, Guizzardi D, Koffi EN, et al. 2020. High resolution temporal profiles in the emissions database for global atmospheric research. *Scientific Data* 7(1):121

Davison AC, Padoan SA, Ribatet M. 2012. Statistical modeling of spatial extremes. *Statistical Science* 27(2):161–186

Davison AC, Smith RL. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)* 52(3):393–442

de Jesus AL, Thompson H, Knibbs LD, Kowalski M, Cyrys J, et al. 2020. Long-term trends in $PM_{2.5}$ mass and particle number concentrations in urban air: The impacts of mitigation measures and extreme events due to changing climates. *Environmental Pollution* 263:114500

Eastoe EF, Tawn JA. 2009. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58(1):25–45

Emami F, Masiol M, Hopke PK. 2018. Air pollution at Rochester, NY: Long-term trends and multivariate analysis of upwind SO_2 source impacts. *Science of the Total Environment* 612:1506–1515

Erçelebi SG, Toros H. 2009. Extreme value analysis of Istanbul air pollution data. *Clean – Soil, Air, Water* 37(2):122–131

Eskridge RE, Ku JY, Rao ST, Porter PS, Zurbenko IG. 1997. Separating different scales of motion in time series of meteorological variables. *Bulletin of the American Meteorological Society* 78(7):1473–1483

Fassò A, Cameletti M. 2009. The EM algorithm in a distributed computing environment for modelling environmental space–time data. *Environmental Modelling & Software* 24(9):1027–1035

Fassò A, Cameletti M, Nicolis O. 2007. Air quality monitoring using heterogeneous networks. *Environmetrics* 18(3):245–264

Fassò A, Finazzi F, Ndongo FB. 2016. European population exposure to airborne pollutants based on a multivariate spatio-temporal model. *Journal of Agricultural, Biological, and Environmental Statistics* 21(3):492–511

Fassò A, Maranzano P, Otto P. 2022. Spatiotemporal variable selection and air quality impact assessment of COVID-19 lockdown. *Spatial Statistics* 49:100549

Fattah A, Morshed SR, Al Kafy A, Rahaman ZA, Ibrahim MH, Rahman MT. 2023. Wavelet coherence analysis of $PM_{2.5}$ variability in response to meteorological changes in south asian cities. *Atmospheric Pollution Research* 14(11):101889

Ferro CAT, Segers J. 2003. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(2):545–556

Finazzi F, Napier Y, Scott M, Hills A, Cameletti M. 2019. A statistical emulator for multivariate model outputs with missing values. *Atmospheric Environment* 199:415–422

Fisher RA, Tippett LHC. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24(2):180–190

Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29(5):1189–1232

Garbatov Y, Georgiev PG, Fuchedzhieva I. 2022. Extreme value analysis of NO_x air pollution in the winter seaport of Varna. *Atmosphere* 13(11):1921

Garcia-Aristizabal A, Bucchignani E, Palazzi E, D'Onofrio D, Gasparini P, Marzocchi W. 2015. Analysis of non-stationary climate-related extreme events considering climate change scenarios: an application for multi-hazard assessment in the Dar es Salaam region, tanzania. *Natural Hazards* 75(1):289–320

Gouldsbrough L, Hossaini R, Eastoe EF, Young PJ. 2022. A temperature dependent extreme value analysis of UK surface ozone, 1980–2019. *Atmospheric Environment* 273:118975

Grange SK, Carslaw DC. 2019. Using meteorological normalisation to detect interventions in air quality time series. *Science of The Total Environment* 653:578–588

Grange SK, Carslaw DC, Lewis AC, Boleti E, Hueglin C. 2018. Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmospheric Chemistry and Physics* 18:6223–6239

Grell GA, Peckham SE, Schmitz R, McKeen SA, Frost G, et al. 2005. Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment* 39(37):6957–6975

Gyarmati-Szabó J, Bogachev LV, Chen H. 2017. Nonstationary POT modelling of air pollution concentrations: Statistical analysis of the traffic and meteorological impact. *Environmetrics* 28(5):e2449

Hamed KH, Rao AR. 1998. A modified Mann–Kendall trend test for autocorrelated data. *Journal of Hydrology* 204(1–4):182–196

Härdle W. 1992. Applied Nonparametric Regression, vol. 19 of *Econometric Society Monographs*. Cambridge: Cambridge University Press

Hastie TJ, Tibshirani RJ. 1986. Generalized additive models. *Statistical Science* 1(3):297–318

Heffernan JE, Tawn JA. 2004. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(3):497–546

Hirsch RM, Slack JR, Smith RA. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18(1):107–121

Houtekamer PL, Zhang F. 2016. Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review* 144(12):4489–4532

Hsing T, Hüsler J, Leadbetter MR. 1988. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* 78:97–112

Huang HC, Cressie N. 1996. Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics & Data Analysis* 22(2):159–175

Kendall MG. 1948. Rank Correlation Methods. London: Charles Griffin

Killick R, Eckley IA. 2014. Changepoint: An R package for changepoint analysis. *Journal of Statistical Software* 58(3):1–19

Kulkarni A, von Storch H. 1992. Monte carlo experiments on the effect of serial correlation on the Mann–Kendall test of trend. *Meteorologische Zeitschrift* 4(2):82–85

Künsch HR. 1989. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17(3):1217–1241

Leadbetter MR, Lindgren G, Rootzén H. 1983. Extremes and Related Properties of Random Sequences and Processes. Springer Series in Statistics. New York: Springer

Li K, Jacob DJ, Liao H, Shen L, Zhang Q, Bates KH. 2018. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proceedings of the National Academy of Sciences of the United States of America* 116(2):422–427

Li K, Jacob DJ, Shen L, Lu X, de Smedt I, Liao H. 2020. Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences. *Atmospheric Chemistry and Physics* 20:11423–11433

Liang X, Zou T, Guo B, Li S, Zhang H, et al. 2015. Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471(2182):20150257

Luo S, Zhu Y, Chen SX. 2022. Episode based air quality assessment. *Atmospheric Environment* 285:119242

Ma L, Graham DJ, Stettler MEJ. 2021. Has the Ultra Low Emission Zone in London improved air quality? *Environmental Research Letters* 16(12):124001

Mann HB. 1945. Nonparametric tests against trend. *Econometrica* 13(3):245–259

Maranzano P, Otto P, Fassò A. 2023. Adaptive LASSO estimation for functional hidden dynamic geostatistical models. *Stochastic Environmental Research and Risk Assessment* 37(12):3615–3637

Martins LD, Wikuats CFH, Capucim MN, de Almeida DS, da Costa SC, et al. 2017. Extreme value analysis of air pollution data and their comparison between two large urban regions of South America. *Weather and Climate Extremes* 18:44–54

Masseran N, Mohd Safari MA. 2021. Mixed POT-BM approach for modeling unhealthy air pollution events. *International Journal of Environmental Research and Public Health* 18(13):6754

Masseran N, Safari MAM. 2020. Risk assessment of extreme air pollution based on partial duration series: Idf approach. *Stochastic Environmental Research and Risk Assessment* 34:545–559

Newey WK, West KD. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708

Orsini N, Bellocco R, Greenland S. 2006. Generalized least squares for trend estimation of summarized dose–response data. *The Stata Journal* 6(1):40–57

Pearce JL, Beringer J, Nicholls N, Hyndman RJ, Tapper NJ. 2011. Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*

45(6):1328–1336

Pickands J. 1975. Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1):119–131

Pozzer A, Anenberg SC, Dey S, Haines A, Lelieveld J, Chowdhury S. 2023. Mortality attributable to ambient air pollution: A review of global estimates. *GeoHealth* 7(1):e2022GH000711

Qayyum F, Tariq S, ul Haq Z, Mehmood U, Zeydan Ö. 2022. Air pollution trends measured from MODIS and TROPOMI: AOD and CO over Pakistan. *Journal of Atmospheric Chemistry* 79:199–217

Qi L, Zheng H, Ding D, Wang S. 2024. A comparison of meteorological normalization of $PM_{2.5}$ by multiple linear regression, general additive model, and random forest methods. *Atmospheric Environment* 338:120854

Qiu M, Zigler CM, Selin NE. 2022. Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions. *Atmospheric Chemistry and Physics* 22(16):10551–10566

Rao ST, Zurbenko IG. 1994. Detecting and tracking changes in ozone air quality. *Journal of the Air & Waste Management Association* 44(9):1089–1092

Rao ST, Zurbenko IG, Neagu RM, Porter PS, Ku JY, Henry RF. 1997. Space and time scales in ambient ozone data. *Bulletin of the American Meteorological Society* 78(10):2153–2166

Reinsel GC, Tiao GC, Meng XL, Choi D, Cheang WK, et al. 1998. Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research: Atmospheres* 103(D14):17149–17161

Scarrott C, MacDonald A. 2012. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT – Statistical Journal* 10(1):33–60

Sen PK. 1968. Estimates of the regression coefficient based on kendall's tau. *Journal of the American Statistical Association* 63(324):1379–1389

Shi Z, Song C, Liu B, Lu G, Xu J, et al. 2021. Abrupt but smaller than expected changes in surface air quality attributable to covid-19 lockdowns. *Science Advances* 7:eabd6696

Skachko S, Menard R, Errera Q, Christophe Y, Chabrillat S. 2016. EnKF and 4D-Var data assimilation with chemical transport model BASCOE (version 05.06). *Geoscientific Model Development* 9(8):2893–2908

Smith RL. 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science* 4(4):367–393

Su FC, Jia C, Batterman SA. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. *Atmospheric Environment* 62:97–106

Theil H. 1950. A rank-invariant method of linear and polynomial regression analysis. parts i–III. *Proceedings of the Royal Netherlands Academy of Sciences, Series A* 53:386–392; 521–525; 1397–1412

Thompson ML, Reynolds JH, Cox LH, Guttorp P, Sampson PD. 2001. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment* 35(3):617–630

Tie X, Huang RJ, Cao J, Zhang Q, Cheng Y, et al. 2017. Severe pollution in China amplified by atmospheric moisture. *Scientific Reports* 7(1):15760

Truong C, Oudre L, Vayatis N. 2020. Selective review of offline change point detection methods. *Signal Processing* 167:107299

U.S. Environmental Protection Agency. 2017. Emissions inventory guidance for implementation of ozone and particulate matter national ambient air quality standards (naaqs) and regional haze regulations. Tech. Rep. EPA-454/B-17-002, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Research Triangle Park, NC. Accessed 2026-02-15

U.S. Environmental Protection Agency. 2025. National ambient air quality standards (NAAQS) table. Last updated November 4, 2025; accessed January 4, 2026

Vu TV, Shi Z, Cheng J, Zhang Q, He K, et al. 2019. Assessing the impact of clean air action on air quality trends in beijing using a machine learning technique. *Atmospheric Chemistry and Physics* 19:11303–11314

Walker SE, Solberg S, Schneider P, Guerreiro C. 2023. The airgam 2022r1 air quality trend and prediction model. *Geoscientific Model Development* 16(22):6673–6694

Wan Y, Xu M, Huang H, Chen SX. 2021. A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. *Environmetrics* 32(1):e2648

Wang K, Ling C, Chen Y, Zhang Z. 2023. Spatio-temporal joint modelling on moderate and extreme air pollution in Spain. *Environmental and Ecological Statistics* 30(4):601–624

Wang L, An X, Li Y, Jing K, Wang Q, et al. 2024. Hourly variation characteristics of PM2.5 and main components in Beijing based on wind direction. *Atmospheric Environment* 327:120493

Wang P, Guo H, Hu J, Kota SH, Ying Q, Zhang H. 2019. Responses of $PM_{2.5}$ and O_3 concentrations to changes of meteorology and emissions in China. *Science of the Total Environment* 662:297–306

Wang Y, Finazzi F, Fassò A. 2021. D-STEM v2: A software for modelling functional spatio-temporal data. *Journal of Statistical Software* 99(10):1–29

Wikle CK, Cressie N. 1999. A dimension-reduced approach to space-time Kalman filtering. *Biometrika* 86(4):815–829

Wood SN. 2017. Generalized additive models: An introduction with R. Boca Raton, FL: Chapman and Hall/CRC, 2nd ed.

Xiao Q, Zheng Y, Geng G, Chen C, Huang X, et al. 2021. Separating emission and meteorological contributions to long-term pm2.5 trends over eastern china during 2000–2018. *Atmospheric Chemistry and Physics* 21(12):9475–9496

Xing J, Zheng S, Ding D, Kelly JT, Wang S, et al. 2020. Deep learning for prediction of the air quality response to emission changes. *Environmental Science & Technology* 54(14):8589–8600

Yue S, Pilon P. 2004. A comparison of the power of the t test, Mann–Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal* 49(1):21–37

Yue S, Pilon PJ, Phinney B, Cavadias G. 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* 16(9):1807–1829

Yusoff MF, Latif MT, Juneng L, Khan MF, Ahamad F, et al. 2019. Spatio-temporal assessment of nocturnal surface ozone in malaysia. *Atmospheric Environment* 217:116955

Zeb N, Khokhar MF, Pozzer A, Khan SA. 2019. Exploring the temporal trends and seasonal behaviour of tropospheric trace gases over Pakistan by exploiting satellite observations. *Atmospheric Environment* 198:279–290

Zhai S, Jacob DJ, Wang X, Shen L, Li K, et al. 2019. Fine particulate matter ($PM_{2.5}$) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology. *Atmospheric Chemistry and Physics* 19:11031–11041

Zhang Q, Zheng Y, Tong D, Shao M, Wang S, et al. 2019. Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. *Proceedings of the National Academy of Sciences of the United States of America* 116(49):24463–24469

Zhang S, Chen SX, Guo B, Wang H, Lin W. 2020. Regional air-quality assessment that adjusts for meteorological confounding. *Sci Sin Math* 50(4):527–558

Zhang S, Guo B, Dong A, He J, Xu Z, Chen SX. 2017. Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473(2205):20170457

Zhang Y, Chen SX, Bao L. 2023. Air pollution estimation under air stagnation—a case study of beijing. *Environmetrics* 34

Zheng H, Kong S, Zhai S, Sun X, Cheng Y, et al. 2023. An intercomparison of weather normalization of $PM_{2.5}$ concentration using traditional statistical methods, machine learning, and chemistry transport models. *npj Climate and Atmospheric Science* 6(1):214

Zhu Y, Liang Y, Chen SX. 2021. Assessing local emission for air pollution via data experiments. *Atmospheric Environment* 252:118323