Environmental Pollution 407 (2026) 128671

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Semi-parametric methods for long-term trend detection in air quality: The North-Eastern Italian Alps case

Claudia Collarin [a,b,*], Paolo Girardi [a], Mauro Masiol [a], Ilaria Prosdocimi [a]

[a] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, Venice, 30172, VE, Italy

[b] School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Edinburgh, EH9 3FD, United Kingdom

ARTICLE INFO

Keywords:
Air quality trends
Generalized additive models
Box–Cox transformation
Left-censored data
Tobit likelihood function

ABSTRACT

Atmospheric pollution has received increasing attention in recent years, particularly in regions such as Northern Italy's Po Valley, where the presence of extended industrial settings, high population density and orographic factors complicate efforts to maintain pollutant concentrations within the mandated standards. This work focuses on assessing air pollution trends in the North-East of Italy, using daily measurements of NO, NO_2, O_3, PM_{10}, and $PM_{2.5}$ collected between 2009 and 2023 from 40 monitoring stations located in foothill and mountainous areas. By means of an integrated regression modelling framework based on Generalized Additive Models that combines Tobit likelihood with Box–Cox transformation, we extract trends on long-term variations accounting for seasonal patterns and weather predictors.

Model performance varied by pollutant, with higher accuracy for gases (R^2 up to 0.86 for O_3) than for particulate matter (R^2 of approximately 0.66). Variability, evaluated by means of Shapley values, was driven mainly by autoregressive persistence (0.53–0.57) and seasonality (0.15–0.38), with wind effects being relevant for PM (contributing approximately 0.15). Trends showed strong declines in NO (up to −50%) and PM (around −30%), while O_3 concentrations increased by roughly 20%. Meteorological anomalies contributed only marginally to the overall variance. The high variability observed in the detected long-term changes underlined the high heterogeneity of air pollution evolution across the study area. This indicated that short-term factors played a dominant role, whereas long-term changes were smaller in magnitude and exert a weaker influence on local air pollution.

1. Introduction

In recent decades, strict European Union (EU) regulatory frameworks, such as the Ambient Air Quality Directives (European Union, 2004, 2008) and targeted sector-specific emission controls, have successfully driven down key pollutant concentrations across Europe (e.g., Bo et al., 2020; Sicard et al., 2021, 2023) and Italy (Masiol et al., 2017; D'Elia et al., 2021; Pivato et al., 2023). These policies resulted in substantial reductions in primary criteria pollutants, including nitrogen oxides ($NO_x = NO + NO_2$) and particulate matter (PM_{10} and $PM_{2.5}$, defined by an aerodynamic diameter less than 10 μm and 2.5 μm, respectively). However, these broad statistical declines obscure localized vulnerabilities: As highlighted by Beloconi and Vounatsou (2023), a vast majority of the European population remains exposed to concentrations that violate the updated WHO guidelines (World Health Organization, 2021) and the limits imposed by the new Directive (EU) 2024/2881 (European Union, 2024). This ongoing exposure underscores the need for realistic and regionally tailored air quality (AQ) management plans.

Currently, AQ trend analysis often relies on linear models such as ordinary, multiple, or generalized least-squares regressions, and other monotonic methods like Mann–Kendall-Sen's slope approach (Jaiswal et al., 2018; Qiu et al., 2022). The primary advantages of these straightforward approaches are their interpretability, low computational burden, and direct reporting of annual concentration change rates. However, literature also highlights their limitations, largely driven by differences in data availability (e.g., missing data, irregular sampling, etc.), structural changes in monitoring networks, seasonality and other periodic drivers. Furthermore, intrinsic data properties such as autocorrelation, non-normality, local meteorological conditions, and microclimatic differences pose significant challenges (e.g., Grange et al., 2018; Collaud Coen et al., 2020; Giovannini et al., 2020; Chen et al.,

* Corresponding author at: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, Venice, 30172, VE, Italy.

E-mail address: claudia.collarin@ed.ac.uk (C. Collarin).

https://doi.org/10.1016/j.envpol.2026.128671
Received 17 February 2026; Received in revised form 29 June 2026; Accepted 29 June 2026
Available online 16 July 2026

2022; Song et al., 2024). If left unaddressed, these confounding factors introduce substantial bias into statistical assessments of long-term trend estimates (Sokhi et al., 2021). Consequently rigid parametric approaches have increasingly been replaced by semi-parametric and non-parametric methods. These advanced techniques successfully accommodate non-linear trends (Mudelsee, 2019; Qiu et al., 2022) while effectively isolating long-term trends from seasonal and inter-annual variations (Dominici et al., 2002; He et al., 2006; Peng et al., 2006).

In this context, Generalized Additive Models (GAMs, Hastie and Tibshirani, 1986; Wood, 2017) have emerged as a robust framework to overcome these limitations. GAMs are particularly attractive because they combine the interpretability of generalized linear models with the flexibility of smooth functions, allowing linear predictors to be expressed as additive, non-linear effects of time and other covariates (Ravindra et al., 2019). Accordingly, GAMs are increasingly applied to long-term AQ studies because they provide a flexible framework for dealing with complex and non-linear relationships between air pollutant concentrations and their environmental drivers. Crucially, GAMs allow for the decoupling of long-term temporal trends from seasonal patterns and short-term meteorological variability, thereby improving the identification of underlying changes in air quality. Additionally, auxiliary predictors can be integrated into the model to account for the combined influence of emissions, meteorological conditions, and site-specific factors. Their application to AQ data has proven to be effective in isolating long-term trends, adjusting for confounding meteorological influences, and capturing complex seasonality (e.g., Carslaw et al., 2007; Pearce et al., 2011; Valotto and Varin, 2016; Ravindra et al., 2019).

Nonetheless, environmental time series present severe distributional challenges that complicate modelling with off-the-shelf GAM approaches. Air pollution data are typically right-skewed, frequently exhibiting heavy upper tails driven by episodic high-pollution events (Bartoletti and Loperfido, 2010; He et al., 2024), and often contain a non-negligible proportion of observations close to or below analytical detection limits (Houseman and Virji, 2017). A widespread practice to mitigate skewness and stabilize variance is to apply a logarithmic transformation (Salcedo et al., 1999; Martínez-Flórez et al., 2014). However, the log-transform constitutes a rigid and restrictive modelling choice that implicitly enforces multiplicative effects and a fixed variance–mean relationship; these assumptions may not hold consistently across different pollutants, monitoring sites, or temporal regimes. Moreover, the presence of left-censored observations further complicates statistical inference, as standard log-based approaches lack a mathematically rigorous or statistically coherent treatment for concentrations falling below detection thresholds.

To address these limitations, this study proposes a novel modelling strategy that integrates GAMs with a data-driven Box–Cox transformation of pollutant measurements, explicitly accounting for left-censored observations. This framework retains the flexibility of smooth temporal components, while allowing the response distribution to adapt to the empirical characteristics of the data. By estimating the optimal transformation parameter jointly with the model, we ensure mathematical coherence between the scale of the response variable and the additive structure of the smooth terms.

We apply this approach to the mountainous area of North-Eastern Italy. This region is characterized by narrow valleys, steep elevation gradients, and a mosaic of land uses ranging from densely populated valley cities and major transit corridors to sparsely inhabited high-elevation alpine zones. In this complex landscape, orographic effects, frequent thermal inversions, and local valley wind systems critically influence pollutant dispersion and accumulation (Whiteman, 2000; Largeron and Staquet, 2016; Masiol et al., 2024), while long-range transport from the Po Valley and central Europe frequently contributes to background concentrations at high-altitude sites (Diémoz et al., 2019). Using this framework, we evaluate the long-term temporal evolution of the primary pollutants regulated by EU policies (including NO, NO_2, O_3, PM_{10}, and $PM_{2.5}$) while overcoming the statistical challenges inherent to air pollution time series. This will aid in the quantification of current changes and inform future strategies for AQ improvement.

2. Data sources

2.1. Air quality data

Ambient pollutant concentrations were retrieved from the AQ monitoring networks managed by five Environmental Protection Agencies of the first-level administrative divisions of Italy: *Agenzia Regionale per la Protezione Ambientale* of Lombardy (ARPA Lombardia), Veneto (ARPAV) and Friuli Venezia-Giulia (ARPA FVG), alongside the *Agenzia provinciale per l'ambiente e la tutela del clima* of the Autonomous provinces of Bolzano (APPA BZ) and the *Agenzia provinciale per la protezione dell'ambiente* of Trento (APPA TN). Both APPA BZ and APPA TN operate under the National System for Environmental Protection (*Sistema Nazionale per la Protezione dell'Ambiente*, SNPA). Distributed across the mountainous and piedmont areas of the eastern Italian Alps, the final data set comprises 40 air monitoring stations spanning a wide range of environments, from large urban centres to remote rural locations (Figure S1b; Table S4). To ensure the study captures long-term regional patterns instead of localized industrial effects, stations located in industrial hot spots were systematically excluded. The remaining monitoring sites are classified into four distinct typologies: rural background (RB), suburban background (SB), urban background (UB), and traffic hotspots (UT).

Data quality is guaranteed by internal protocols, fully complying with the standards set by the European Commission: UNI EN 14211:2012 for NO and NO_2, UNI EN 14625:2012 for O_3, UNI EN 12341:2014 for PM_{10} and $PM_{2.5}$. For each pollutant, the dataset comprises over 10^5 measurements (Table S3), with gaseous species (NO, NO_2, O_3) recorded at hourly resolution, and particulate matter measured at hourly or daily frequencies. All stations included in this study are required to have validated records spanning at least the period from 1 January 2016 to 31 December 2023. Where available, high-quality historical series extend as far back as 2009; these earlier data were retained to maximize the temporal depth of the trend analysis framework.

2.2. Data processing and treatment of left-censored observations

This study examines the daily mean concentrations of NO, NO_2, O_3, PM_{10}, and $PM_{2.5}$. Given that measurement precision and instrumentation vary across regional agencies, particular attention is paid to the systematic handling of limits of detection (LODs) when aggregating hourly data into daily metrics. For example, ARPAV provides explicit LOD thresholds for its monitoring network: these are 2 $\mu g\,m^{-3}$ for NO and 4 $\mu g\,m^{-3}$ for all the other pollutants. For agencies where explicit LODs are unavailable, the LOD is inferred as the lowest non-zero observed value for each site-pollutant combination.

Values falling at or below their respective LOD are classified as left-censored, indicating that the true concentration is only known to be within the interval $[0, \mathrm{LOD}]$. This censoring logic extends to the calculation of daily means: if a single hourly observation y_i within a 24 h sequence falls below the LOD, the resulting daily mean $\bar{y} = \sum_{i=1}^{24} y_i/24$ is considered left-censored, representing an upper bound for the actual daily mean. Figure S1a illustrates this approach using NO_2 data from an ARPAV station. To facilitate the modelling strategy described in Section 3.3, a binary indicator is created for each daily mean to denote its censoring status.

2.3. Temporal predictors and meteorological anomalies

To isolate long-term trends from short-term variations, the statistical model incorporates both temporal and meteorological predictors. Seasonality is accounted for through three time-based variables (Table S1): day of the week (dow), day of the year (doy), and a continuous time variable (t). Meteorological variables were derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land dataset (Muñoz, 2019) at a 0.1×0.1 degree (approx. 9×9 km) spatial resolution. The variables selected for this study are known to directly or indirectly influence AQ. They include 2-m air temperature (*t2m*, K), total precipitation (*tp*, mm), surface net short-wave (solar) radiation (*ssr*, J m^{-2}), alongside eastward and northward wind speed components (*u10, v10*, m s^{-1}). Table S2 summarizes these variables along with their potential impacts on AQ. To maintain temporal consistency with the daily pollutant observations, the original hourly meteorological records are aggregated into 24-hour averages.

Because meteorological covariates exhibit strong seasonal signals, daily meteorological anomalies relative to the long-term mean over the study period are included in the model instead of raw values. This approach reduces collinearity between doy effects and weather covariates, thereby improving model stability and convergence. For each weather covariate W_{doy} at a given station, a GAM is employed to estimate its mean. Specifically, temperature and wind speed are assumed to follow a Gaussian distribution, $W_{doy} \sim \mathcal{N}(\mu_{doy}, \sigma^2)$, while total precipitation and surface net solar radiation are assumed to follow a Gamma distribution, $W_{doy} \sim Gamma(\mu_{doy}, \sigma^2)$. The mean μ_{doy} is estimated as a smooth function of doy, $\mu_{doy} = f(doy)$, while σ^2 is kept constant over time. The anomalies are defined as the resulting deviance residuals and are denoted by the prefix a hereafter (e.g. $a.ssr$ represents the solar radiation anomaly). This procedure ensures that the anomalies are approximately normally distributed around zero.

3. Statistical modelling framework

3.1. Generalized additive models

The GAM framework assumes that the response variable Y_t follows an exponential family distribution $Y_t \sim \text{EF}\left(\mu_t, \sigma\right)$ with mean μ_t and scale parameter σ. The linear predictor η_t is related to the mean $\mu_t = \mathbb{E}[Y_t]$ through a link function g:

$$\eta_t = g\left(\mu_t\right) = \mathbf{A}_t \boldsymbol{\beta}_0 + \sum_j f_j(\mathbf{x}_{tj}) \tag{1}$$

where $\mathbf{A}_t \boldsymbol{\beta}_0$ represents the parametric components and f_j terms denote smooth functions of the covariates $\mathbf{x}_j$. Each smooth term is expressed as a linear combination of basis functions, $f(\mathbf{x}) = \sum_{k=1}^{K} \beta_k b_k(\mathbf{x})$, where the basis dimension K is chosen to be sufficiently large to avoid over-smoothing. The vector containing all model coefficients $\boldsymbol{\beta}$ is estimated by maximizing the penalized log-likelihood: $\hat{\boldsymbol{\beta}} = \text{argmax}_{\boldsymbol{\beta}} \left\{ \ell(\boldsymbol{\beta}) - \frac{1}{2} \sum_{j=1}^{M} \lambda_j \boldsymbol{\beta}^\top \mathbf{S}^j \boldsymbol{\beta} \right\}$. Each penalty matrix $\mathbf{S}^j$ defines the concept of smoothness for the corresponding effect, while the smoothing parameter λ_j controls the trade-off between the goodness of fit and smoothness. The joint estimation of model coefficients and smoothing parameters is performed using the Restricted Maximum Likelihood (REML) framework as implemented in the `mgcv` package for R (Wood et al., 2016).

3.2. The Box–Cox transformation

To accommodate the skewed distributions characteristic of ambient air pollution records, the conditional distribution of the response variable must often be transformed before fitting the regression model. Rather than relying on a rigid a priori transformation, we adopt a flexible Box–Cox framework (Box and Cox, 1964), which is designed to simultaneously achieve approximate normality and homoscedasticity by applying a transformation to the response. For non-negative responses, it is defined as a function of the parameter $\zeta \in \mathbb{R}$:

$$h(y, \zeta) = \begin{cases} \frac{y^\zeta - 1}{\zeta} & \zeta \neq 0 \\ \log y & \zeta = 0. \end{cases} \tag{2}$$

Specific values of ζ correspond to standard transformations: $\zeta = 1/2$ represents the square root, $\zeta = 0$ the logarithm, and $\zeta = -1$ the reciprocal. The resulting model structure is formulated as $h(Y_t, \zeta) = \mu_t + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

3.3. Combining Box–Cox transformation and observation censoring with GAMs

To ensure the robustness of the results and minimize model uncertainty, we employ an integrated approach combining Generalized Additive Models (GAMs) with a data-driven Box–Cox transformation and a Tobit likelihood function. This joint framework addresses three primary sources of statistical uncertainty: (i) non-linear dependencies between concentrations and predictors are captured via spline smoothers; (ii) the Box–Cox transformation stabilizes variance while correcting for the inherent skewness of pollutant distributions; and (iii) the Tobit likelihood accounts for the left-censoring induced by LODs.

The inclusion of a data-driven Box–Cox transformation is particularly vital to capture the varied nature of skewness and the complex interplay of anthropogenic emissions, orography, and atmospheric chemistry in the AQ data. Furthermore, the model explicitly accounts for the left-censoring induced by the instrument LODs. The study data are characterized by different censoring patterns (Table S3), with rural and suburban background stations showing the highest proportion of left-censored observations. These differences are illustrated in Figure S2, which compares NO concentrations across three rural background stations. Despite sharing the same classification, the stations show differences in precision, concentration ranges, and censoring thresholds, all of which are addressed within the integrated GAM-Tobit-Box–Cox framework.

Due to the variability in station-specific instrument precision, and the strong dependence of the transformation parameter ζ on local conditions, models are fitted separately for each station s and pollutant p. For notational simplicity, the ps subscripts are omitted, though each model is understood to represent a specific pollutant–station pair. Let $Z^*_{t;h}$ represent the true latent concentration at hour h on day t, and let $Y^*_t = \frac{1}{24} \sum_{h=1}^{24} Z^*_{t;h}$ be the true latent daily mean. The Box–Cox transformed response $h(Y^*_t, \zeta)$ is assumed to follow a Gaussian distribution $\mathcal{N}(\mu_t, \sigma^2)$, where the mean μ_t is decomposed as:

$$\mu_t = \beta_0 + \beta_t t + f_1(t) + \text{Seasonal}_t + \text{Autoregressive}_{t-1} + \text{Meteorology}_t \tag{3}$$

with

$$\begin{aligned} &\text{Seasonal}_t = f_2(dow_t) + f_3(doy_t) \qquad \text{Autoregressive}_{t-1} = f_4\left(Y_{t-1}\right) \\ &\text{Meteorology}_t = f_5\left(a.ssr_t\right) + f_6\left(a.t2m_t\right) + f_7\left(a.tp_t\right) + f_8\left(a.u10_t, a.v10_t\right). \end{aligned} \tag{4}$$

The Trend component consists of two orthogonal terms: a linear ($\beta_t t$) and a non-linear smooth term ($f_1(t)$, with basis dimension $K_1 = 10$) represented by orthogonalized B-splines (Eilers and Marx, 1996). This explicit decoupling allows the overall monotonic direction to be separated from non-linear fluctuations due to environmental changes which are not explained by weather and seasonal variables. The Seasonal component captures periodic fluctuations using a cyclic cubic regression spline for dow ($f_2(w_t), K_2 = 7$) and a thin plate regression spline (Wood, 2003) for doy ($f_3(doy_t)$, $K_3 = 20$). The Autoregressive term ($f_4\left(Y_{t-1}\right)$, $K_4 = 10$) accounts for short-term temporal dependency using the observed concentration from the previous day. Finally, the Meteorology component uses thin plate regression splines for anomalies of solar radiation ($f_5\left(a.ssr_t\right)$, $K_5 = 10$), temperature ($f_6\left(a.t2m_t\right)$, $K_6 = 10$), and precipitation ($f_7\left(a.tp_t\right)$, $K_7 = 10$), while the wind speed's anomaly components are incorporated through a tensor product of cubic regression splines ($f_8\left(a.u10_t, a.v10_t\right)$, $K_8 = (10, 10)$). Basis dimensions (K) have been verified via effective degrees of freedom checks (see Tables S7-S11); for the autoregressive lag of NO, K_4 has been constrained to prevent overfitting and maintain a physically plausible, continuous relationship (Harrell, 2015).

To account for measurement limitations, this structure is embedded within a censored Gaussian (Tobit) framework (Tobin, 1958). The observed daily mean Y_t and the censoring indicator δ_t are defined as:

$$(Y_t, \delta_t) = \begin{cases} (Y^*_t, 1) & \text{if all } Z^*_{t;h} > \text{LOD} \quad \text{(uncensored)} \\ \left(\text{LOD}^{(Y)}_t, 0\right) & \text{if any } Z^*_{t;h} \leq \text{LOD} \quad \text{(censored),} \end{cases} \tag{5}$$

where $\mathrm{LOD}_t^{(Y)} = \frac{1}{24}\sum_{h=1}^{24}\left[\delta_h Z_{t;h}^* + (1-\delta_h)\mathrm{LOD}_{ps}\right]$ represents the daily detection limit, calculated by substituting the station-specific LOD for any latent hourly concentrations falling below the threshold.

3.4. Model interpretation and predictor importance

In model (3), estimating $\mu_t = \mathbb{E}[h(Y_t,\zeta)]$ implies that the covariate effects are additive on the transformed scale, rather than on the original scale. Because the Box–Cox parameter ζ is station-specific, fitted concentrations cannot be directly compared across the network in their transformed state. We therefore employ two applied statistical modelling approaches to visualize and quantify predictor influences on a comparable scale.

The first approach uses Accumulated Local Effects plots (ALE, Apley and Zhu, 2020). ALE plots are preferred over traditional partial dependence plots as they provide unbiased visualizations of covariate effects, even when predictors are correlated. In this framework, the local effect of X_j is averaged over the conditional distribution of the other predictors and then accumulated through integration. To ensure comparability across stations with different baseline concentrations and transformation parameters, ALE effects are normalized by using the station-specific mean pollutant concentration recorded in the first year of monitoring. Thus, these plots represent the isolated effect of each variable as a percentage change relative to the initial year.

The second method assesses the relative contribution of each predictor using a derivation of Shapley values for GAMs (Khorrami Chokami and Rabitti, 2024). This method provides a normalized measure of the proportion of total variability in the linear predictor attributable to each smooth component. This allows for a direct comparison of the relative roles that seasonality, weather anomalies, and long-term trends play in driving pollutant concentrations across the study area.

4. Results

4.1. Response distribution and model fit

The results indicate substantial heterogeneity in the estimated response distributions across different pollutants and monitoring stations. This variability is directly captured by the estimated station-specific Box–Cox parameters ζ_{ps} (Figure S4), which vary systematically by pollutant type. For PM_{10}, $PM_{2.5}$, and NO_2, the estimated ζ_{ps} values lie predominantly within the range $(0, 0.5)$, reflecting moderately right-skewed distributions. Conversely, NO concentrations show markedly lower ζ_{ps} estimates that cluster near zero, indicating a highly skewed profile that aligns closely with a classic logarithmic scale. Finally, O_3 displays the highest ζ_{ps} estimates, with most values ranging from 0.5 to 1, which is consistent with near-normal distributions. These differences highlight the ability of the model to adapt to pollutant-specific distributional characteristics.

Across all pollutants, the model shows a strong correlation between the transformed observed and predicted concentrations (Fig. 1a–e, Table S6), indicating that, by using the Tobit formulation, it effectively recovers the unobserved latent concentrations that lie below the LOD. While the fitted concentrations might appear to underestimate observations close to the lower bound for specific stations (e.g., Fig. 1d), this behaviour is a mathematically expected consequence of integrating the left-censoring mechanism into the likelihood. The clear benefit of this explicit formulation is best illustrated in the temporal domain in Fig. 1f.

4.2. Predictor importance

To evaluate predictor importance, median Shapley values are employed to decompose the relative variance contribution of each model component (Table 1). The resulting metrics identify the seasonal profile, $f_3(doy_t)$, and the autoregressive lag, $f_4(Y_{t-1})$, as the dominant contributors to pollutant variability, which highlights the strong influence of recurring emission patterns, atmospheric chemistry, and accumulation processes. Wind dynamics, captured by the anomaly component $f_8(a.u10_t, a.v10_t)$, play a substantial role for PM_{10} and $PM_{2.5}$ concentrations, indicating the importance of dispersion, resuspension, and transport mechanisms for particulate matter. The remaining meteorological anomalies explain only a small fraction of the total variance, indicating that short-term departures from average weather conditions have a limited impact compared to seasonality and persistence in pollution concentrations. In contrast, the long-term trend component, $\beta_t t + f_1(t)$, explains a much smaller proportion of total variance, but remains crucial to capture the gradual, monotonic trajectories over the multi-year study period.

Table 1
Median of Shapley values.

	NO	NO_2	O_3	PM_{10}	$PM_{2.5}$
$\beta_t t + f_1(t)$	0.037	0.057	0.003	0.025	0.028
$f_2(w)$	0.026	0.031	0.001	0.008	0.003
$f_3(doy)$	0.325	0.304	0.380	0.152	0.158
$f_4(Y_{t-1})$	0.571	0.546	0.571	0.545	0.527
$f_5(a.ssr)$	−0.001	−0.011	0.013	0.012	0.002
$f_6(a.t2m)$	0.001	0.001	0.003	0.060	0.071
$f_7(a.tp)$	0.010	0.013	0.001	0.012	0.006
$f_8(a.u10, a.v10)$	0.017	0.051	0.013	0.153	0.147

4.3. Long-term trends

After accounting for meteorological and seasonal confounding, the trend component isolates long-term, multi-annual variations on the transformed scale. When interpreting these trajectories, attention must be paid to the uneven distribution of monitoring sites between different station types, which requires a contextualized assessment of the aggregated patterns. Furthermore, because of the inverse Box–Cox transformation, the normalized ALE profiles are no longer strictly linear when mapped back to the original measurement scale, although their monotonicity is fully preserved. This monotonic behaviour is clearly reflected in Fig. 2, where O_3 concentrations exhibit an overall increasing trend, with relative increases reaching up to 20% compared to the initial monitoring year, which serves as the 100% reference baseline. In contrast, most of the remaining pollutants show negative trends, suggesting a general improvement in AQ in the study area. The most pronounced decreases occur for NO, with reductions of up to 50% at the end of the study period. The estimated trend for particulate matter is more moderate, with decreases reaching a maximum of approximately 30%. Figure S3 and S5 in the Supplementary Material show the value of the regression coefficients for the linear term and the ALE plot of the non-linear component of t.

4.4. Seasonality and meteorological anomalies

The doy effect ($f_3(doy_t)$, Figure S6) captures seasonal patterns driven by a combination of boundary-layer dynamics and human activities. For NO, NO_2, PM_{10}, and $PM_{2.5}$, the model exhibit pronounced seasonal cycles, with elevated concentrations during the cold months. These patterns can be explained by the combined effect of several covarying factors (e.g., Masiol et al., 2017): (i) lower wintertime mixing layer heights, which limit pollutant dispersion, and an increased frequency of persistent thermal inversions; (ii) temperature-driven gas–particle partitioning of semi-volatile compounds, such as ammonium nitrate and secondary organic aerosols; (iii) increased emissions during the cold season, mainly from additional domestic heating, including biomass burning and typically operating from mid-October to mid-April; and (iv) reduced winter solar radiation and actinic fluxes, leading to lower photochemistry and oxidative capacity, which slows pollutant

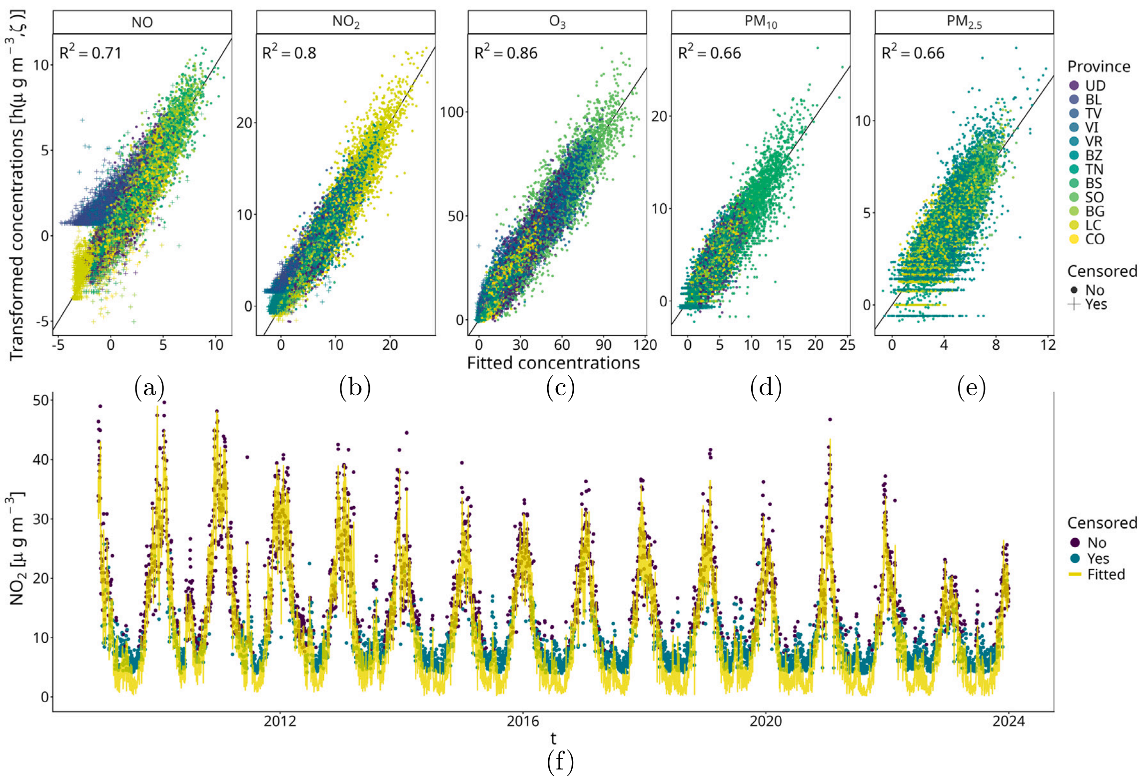

Fig. 1. Estimated concentrations on the Box–Cox transformed scale (a-e) and estimated NO_2 concentrations on the original scale for station BL_SB_1 (f).

removal. However, this pattern is much weaker, absent, or even reversed at rural background sites, which are less directly affected by local emissions and more strongly controlled by boundary-layer dynamics, particularly at low-elevation locations. In these rural environments, enhanced vertical mixing during the warm season can favour the accumulation and regional transport of secondary pollutants. Notably, PM_{10} and $PM_{2.5}$ generally exhibit an opposite seasonal cycle, with higher concentrations in summer, a trend consistent with enhanced formation of secondary biogenic aerosols driven by increased biogenic volatile organic compounds (VOCs) emissions during the vegetative period. Conversely, O_3 displays an inverse seasonal pattern, peaking during the summer. This behaviour is consistent with the seasonality of solar radiation and temperature, which intensify photochemical reactions involving NO_x and volatile organic compounds that drive O_3 formation and accumulation during the warm season (e.g., Monks et al., 2015; Seinfeld and Pandis, 2016).

While meteorological anomalies contribute marginally to the total variance, specific patterns are identified through ALE plots. Positive anomalies in surface solar radiation are associated with higher O_3 concentrations (Figure S7), reflecting the solar radiation's role in driving photochemical O_3 production. Similarly, anomalies in 2-m air temperature are associated with higher concentrations of PM_{10}, $PM_{2.5}$ and O_3 (Figure S8). Conversely precipitation anomalies exhibit a negative effect, most pronounced for particulate matter (PM_{10} and $PM_{2.5}$) and O_3 concentrations (Figure S9), a pattern consistent with the role of wet scavenging and precipitation-driven removal processes. Notably, the northward wind component ($a.v10$) is positively associated with particulate matter (Figure S11), likely reflecting the transport of persistently polluted air masses from the Po Valley (Whiteman, 2000; Largeron and Staquet, 2016; Diémoz et al., 2019), located south of the study area, towards the Alpine region.

Shapley values clarify why the explained variance is lower for PM_{10} and $PM_{2.5}$ (R^2 approximately 0.66, Fig. 1d,e) than for gaseous pollutants such as O_3 (R^2 = 0.86, Fig. 1c). These results indicate that the day-of-year effect has substantially weaker predictive power for PM (around 0.15) than for gaseous pollutants (more than 0.30). This suggests that PM concentrations are less driven by regular seasonal cycles and more influenced by a combination of local orographic features and meteorological conditions. Accordingly, meteorological anomalies exhibit a slightly stronger impact on PM estimates, likely because they interact with the specific environmental characteristics of each site of Alpine foothills. Finally, a key limitation is the absence of specific source-apportionment predictors, such as biomass burning indicators or traffic volumes, or high-resolution local emission data, which are crucial for explaining the remaining variance in particulate matter concentrations.

5. Discussion

The proposed statistical approach identifies long-term trends for each pollutant–monitoring station pair and quantifies the impact of various predictors on pollutant concentrations. A first clear result is that seasonality (doy) and short-term persistence (lagged term) represent the primary drivers of pollutant variability. This result is consistent with the established understanding that air-pollution time series are largely governed by recurring seasonal patterns and day-to-day dynamics, whereas long-term signals are typically smaller in magnitude and contribute only a limited fraction of the overall variance (Dominici et al., 2002; Peng et al., 2006; Zhang et al., 2025).

The minor variance contribution of the long-term component should not be interpreted as evidence of weak long-term change; rather, it underscores that policy- and technology-driven trends manifest as gradual shifts that can be subtle relative to meteorologically driven fluctuations and persistence. In this regard, the adopted decomposition of the long-term trend into a linear component plus an orthogonalized smooth is advantageous because it yields a directly interpretable monotonic direction while retaining the ability to capture departures from strict linearity without confounding the linear signal. The estimated

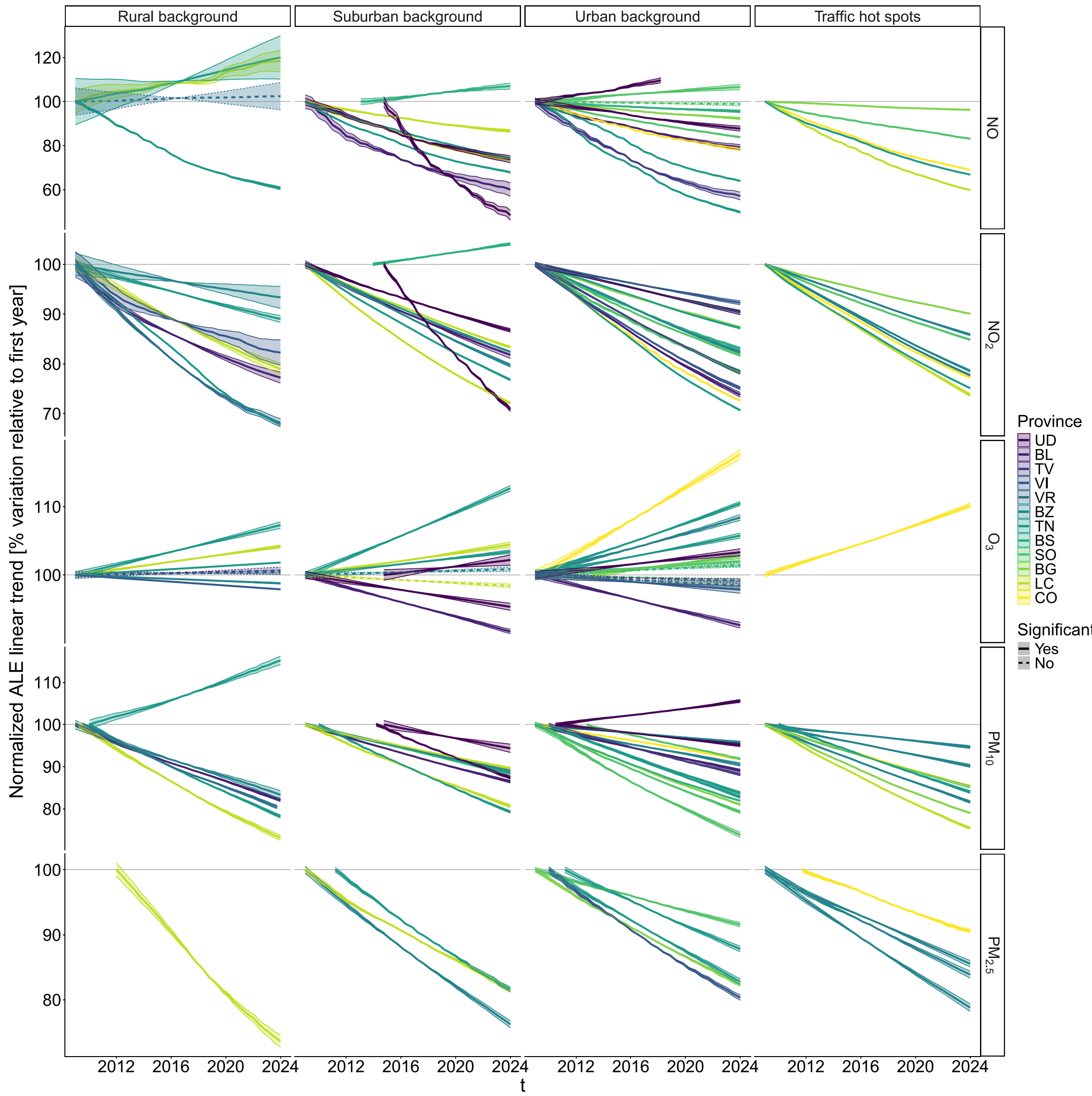

Fig. 2. Estimated Accumulated Local Effects (ALE) for the linear component of trends in NO, NO_2, O_3, PM_{10}, and $PM_{2.5}$. Trends are expressed as percentage variations relative to the mean concentration observed during the first year of monitoring (baseline = 100%). Smooths are grouped by station classification and coloured by province; note that *y*-axis scales are pollutant-specific (row-wise). Dashed lines indicate trends where the associated linear coefficient is not statistically significant ($p > 0.05$).

long-term changes align with the broader European and Italian evidence of improved AQ in recent decades (Colette et al., 2011; Guerreiro et al., 2014; Bo et al., 2020; Sicard et al., 2021, 2023; Masiol et al., 2017; D'Elia et al., 2021; Di Bernardino et al., 2022; Pivato et al., 2023). The strongest decreases were detected for NO and NO_2, likely in connection with EU emission-control policies targeting transport and combustion sources (European Union, 2004, 2008). In contrast, PM exhibited more moderate declines, a plausible outcome given its multi-source nature and the relevance of both primary emissions and secondary formation processes. In mountain valleys, the frequent occurrence of stable boundary layers and thermal inversions can further sustain elevated PM concentrations even when emissions decrease, limiting the extent to which long-term reductions in emissions translate into proportional reductions in concentrations (Whiteman, 2000; Largeron and Staquet, 2016; Masiol et al., 2024). The high heterogeneity observed between stations reinforces that mountainous regions are not a uniform receptor environment owing to local factors.

O_3 emerges as the main exception to the general AQ improvement across the study area, showing a tendency towards slightly increasing concentrations in several stations even in the presence of decreasing NO_x. The contrasting trends of O_3 and NO_x have been reported in several European contexts (e.g., Yan et al., 2019; Boleti et al., 2020; Sicard et al., 2021, 2023; Real et al., 2024). Overall, these studies show that O_3 trends are often flat or even positive, suggesting that reductions in precursor emissions do not necessarily translate to lower O_3 concentrations, especially in heavily polluted environments. However, most of these studies have focused on major European urban areas or lowland/plain environments, whereas the current study is centred on mountain sites, where meteorology, topography, vertical mixing, and valley–slope circulations can substantially modulate the relationship between precursor emissions and observed O_3 trends. Hence, the O_3 response should also be interpreted in relation to the complex topography of the Alpine region, which leads to the uneven distribution of primary sources and distinctive meteorological conditions.

O_3 concentrations depend on photochemistry and precursor regimes (Monks et al., 2015; Seinfeld and Pandis, 2016), and their response to emission reductions varies with the local NO_x–VOC balance, NO titration, background O_3 transport, and meteorological conditions. Consequently, its response to precursor reductions might be non-linear. Local emission inventories report that NO_x in the study area are emitted by primary sources, mostly road traffic, residential combustion and industrial processes (ARPA Friuli Venezia Giulia, 2021; ARPA Lombardia, 2023; ARPA Veneto and Regione del Veneto, 2026). NO rapidly consumes O_3 through titration in fresh plumes to form NO_2. Therefore, emission reductions in sites heavily affected by primary sources (e.g., traffic hotspots, urbanized areas, or road corridors) may weaken local O_3 titration, partly compensating for, or even reversing, the expected decrease in O_3. This mechanism is expected to be relevant in NO_x-rich or VOC-limited environments, where reducing NO_x alone may not immediately reduce O_3 but could eventually lead to increased O_3 concentrations.

In the Alpine area, the interpretation is further complicated by terrain-driven meteorology. Mountain terrain creates distinctive atmospheric circulation patterns driven mainly by local thermal forcing, including thermally induced winds (Whiteman, 2000). Winter and nocturnal inversions and poor dispersion may also occur frequently (Largeron and Staquet, 2016) leading to pollutant accumulation in valley floors, while daytime slope and valley winds can lift polluted boundary-layer air towards elevated slopes and the lower free troposphere. Such topographic venting modifies both the spatial distribution of NO_x and the chemical environment in which O_3 is produced.

In mountain regions, long-range transport and coupling with free-tropospheric air masses can further influence O_3 concentrations at higher elevations (Diémoz et al., 2019), potentially increasing the relevance of regional background trends relative to local emission changes. These findings suggest that additional analyses linking O_3 behaviour with precursor changes and transport indicators would be valuable, particularly in view of forthcoming tighter AQ targets.

From a policy perspective, the results suggest that reductions in NO_x and (to a lesser extent) particulate matter are detectable even in the foothill and mountainous sectors of Northern Italy, supporting the effectiveness of emission-control measures in urban and non-urban environments. At the same time, the tendency for O_3 to remain stable or increase indicates that compliance with future more stringent standards may be challenging if strategies focus primarily on primary pollutants. This is particularly relevant given the upcoming tightening of EU air-quality requirements and the alignment with updated WHO guidelines (European Union, 2024; World Health Organization, 2021). In this context, site-specific evidence from mountain monitoring stations is essential for designing realistic mitigation plans that reflect local atmospheric dynamics and the role of regional transport, including areas that may be perceived as relatively pristine.

The modelling choices were motivated by the well-known distributional features of pollutant concentrations, including non-negativity, right skewness, left-censoring, and heavy tails associated with episodic events (Bartoletti and Loperfido, 2010; He et al., 2024). These were considered alongside the practical issue that monitoring data often include concentrations at or below analytical detection limits (Houseman and Virji, 2017). While log-transformations are commonly adopted in air-quality trend studies (Salcedo et al., 1999; Martínez-Flórez et al., 2014), a fixed transformation can impose an overly restrictive variance-mean relationship and may not be equally appropriate for different pollutants and monitoring sites. Instead, the model is fitted separately for each station and pollutant, estimating the Box–Cox transformation parameter jointly with the GAM. While fitting models separately for each station was motivated by computational feasibility and the need to allow for station-specific Box–Cox parameters, this represents an adaptive compromise: the marked variation of ζ_{ps} across pollutants and stations suggests that a single transformation would be suboptimal in such a heterogeneous setting. This is also evidenced by a comparison in the predictive ability for models based on the fixed logarithmic scale transformation shown in Table S12. However, this approach implicitly assumes independence between locations and neglects spatial correlation, potentially limiting the ability to disentangle local effects from regional-scale transport and shared atmospheric drivers.

6. Conclusions

The proposed approach tackles a number of challenges posed by the statistical modelling of long daily air-quality series: this is achieved by combining GAMs with a data-driven Box–Cox transformation and a Tobit likelihood. Communicating and interpreting the results of this framework presents a challenge because the station- and pollutant-specific transformations complicate direct comparisons across the monitoring network. Therefore, it is necessary to further process the model's output to enable a general comparison of the results. In this study, ALE plots are employed to map the marginal covariate effects directly back onto the original measurement scale, thereby facilitating the comparison of long-term trends. Complementing this visual framework, Shapley values provide a numerical synthesis of predictor contributions, enabling an objective ranking of the primary drivers of air quality variability. From a computational point of view, a station-wise modelling approach facilitates a more feasible estimation task given the dataset's scale, which exceeds 10^5 total measurements for each pollutant. While complex hierarchical or spatial models could facilitate a joint analysis across sites, potentially improving the separation of local effects from large-scale transport, such frameworks are computationally demanding. Furthermore, they require restrictive assumptions, such as a uniform Box–Cox transformation across all locations, which would undermine the model's capacity to adapt to site-specific distributional heterogeneity.

The conservative definition of daily censoring (a day flagged as censored when any hourly value is below the LOD) preserves measurement-limit information but may overstate censoring for days affected only in a small number of hours. However, preliminary comparisons (available in the supplementary material) indicate that this choice has a small but positive effect on the predictive ability of the estimation procedure. Future work could further improve the proposed modelling framework by accounting for the varying reliability of censored observations, for example, by incorporating information on measurement uncertainty near detection limits or by weighting censored concentrations according to instrument performance and quality-control metadata.

CRediT authorship contribution statement

Claudia Collarin: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Paolo Girardi:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Mauro Masiol:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Ilaria Prosdocimi:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Dr. Marianna D'Amico for facilitating the access to the pollution data. The authors also thank prof. Simon Wood for implementing the censored Box–Cox Gaussian family in the mgcv R package. The authors acknowledge the editor and the anonymous reviewers for their feedback and constructive suggestions, which substantially improved the clarity and quality of this manuscript. The contribution of CC, PG and IP to this study was supported by DoE 2023–2027 (MUR, AIS.DIP.ECCELLENZA2023_27.FF project). The contribution of MM to this study was funded by the European Union - NextGenerationEU, within the framework of the iNEST - Interconnected Nord-Est Innovation Ecosystem (iNEST ECS_00000043 – CUP H43C22000540006). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.envpol.2026.128671.

Data availability

Data and code to reproduce results contained in this article are available at Collarin et al. (2026).

References

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. J. R. Stat. Soc. Ser. B Stat. Methodol. 82 (4), 1059–1086.

ARPA Friuli Venezia Giulia, 2021. INEMAR friuli venezia giulia 2021 - inventario regionale delle emissioni in Atmosfera, Edizione 2021. URL: https://www.arpa.fvg.it/temi/temi/aria/sezioni-principali/catasto-emissioni/catasto-emissioni-2021/. Database online. (Last accessed 5 May 2026).

ARPA Lombardia, 2023. INEMAR Lombardia - inventario regionale delle emissioni in atmosfera, edizione 2023. URL: https://www.arpalombardia.it/temi-ambientali/aria/inventario/inventario-delle-emissioni/. Sistema INEMAR - Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. (Last accessed 5 May 2026).

ARPA Veneto, Regione del Veneto, 2026. INEMAR veneto 2023 - inventario regionale delle emissioni in atmosfera in veneto. URL: https://www.arpa.veneto.it/dati-ambientali/open-data/atmosfera/dati-comunali-emissioni-inemar-veneto. Edizione 2023 (gennaio 2026). ARPAV – Dipartimento Regionale Qualità dell'Ambiente – UQA e Regione del Veneto – Area Tutela e Sicurezza del Territorio, Direzione Ambiente e Transizione Ecologica – UOQATA. (Last accessed 5 May 2026).

Bartoletti, S., Loperfido, N., 2010. Modelling air pollution data by the skew-normal distribution. Stoch. Environ. Res. Risk Assess. 24 (4), 513–517.

Beloconi, A., Vounatsou, P., 2023. Revised EU and WHO air quality thresholds: Where does europe stand? Atmos. Environ. 314, 120110.

Bo, M., Charvolin-Volta, P., Clerico, M., Nguyen, C.V., Pognant, F., Soulhac, L., Salizzoni, P., 2020. Urban air quality and meteorology on opposite sides of the Alps: The Lyon and Torino case studies. Urban Clim. 34, 100698.

Boleti, E., Hueglin, C., Grange, S.K., Prévôt, A.S., Takahama, S., 2020. Temporal and spatial analysis of ozone concentrations in europe based on timescale decomposition and a multi-clustering approach. Atmospheric Chem. Phys. 20 (14), 9051–9066.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2), 211–243.

Carslaw, D.C., Beevers, S.D., Tate, J.E., 2007. Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach. Atmos. Environ. 41 (26), 5289–5299.

Chen, Y., Masiol, M., Squizzato, S., Chalupa, D.C., zda Zíková, N., Pokorná, P., Rich, D.Q., Hopke, P.K., 2022. Long-term trends of ultrafine and fine particle number concentrations in New York state: Apportioning between emissions and dispersion. Environ. Pollut. 310, 119797.

Colette, A., Granier, C., Hodnebrog, Ø., Jakobs, H., Maurizi, A., Nyiri, A., Bessagnet, B., d'Angiola, A., D'Isidoro, M., Gauss, M., et al., 2011. Air quality trends in europe over the past decade: a first multi-model assessment. Atmos. Chem. Phys. 11 (22), 11657–11678.

Collarin, C., Girardi, P., Masiol, M., Prosdocimi, I., 2026. Code for "Semi-Parametric Methods for Long-Term Trend Detection in Air Quality: The North-Eastern Italian Alps Case". Zenodo, http://dx.doi.org/10.5281/zenodo.18657516.

Collaud Coen, M., Andrews, E., Bigi, A., Martucci, G., Romanens, G., Vogt, F.P.A., Vuilleumier, L., 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. Atmos. Meas. Tech. 13 (12), 6945–6964.

D'Elia, I., Briganti, G., Vitali, L., Piersanti, A., Righini, G., D'Isidoro, M., Cappelletti, A., Mircea, M., Adani, M., Zanini, G., Ciancarella, L., 2021. Measured and modelled air quality trends in Italy over the period 2003–2010. Atmos. Chem. Phys. 21 (13), 10825–10849.

Di Bernardino, A., Iannarelli, A.M., Diémoz, H., Casadio, S., Cacciani, M., Siani, A.M., 2022. Analysis of two-decade meteorological and air quality trends in Rome (Italy). Theor. Appl. Climatol. 149 (1), 291–307.

Diémoz, H., Barnaba, F., Magri, T., Pession, G., Dionisi, D., Pittavino, S., Tombolato, I.K., Campanelli, M., Della Ceca, L.S., Hervo, M., et al., 2019. Transport of po valley aerosol pollution to the northwestern Alps–part 1: Phenomenology. Atmos. Chem. Phys. 19 (5), 3065–3095.

Dominici, F., McDermott, A., Zeger, S.L., Samet, J.M., 2002. On the use of generalized additive models in time-series studies of air pollution and health. Am. J. Epidemiol. 156 (3), 193–203.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statist. Sci. 11 (2), 89–121.

European Union, 2004. Directive 2004/107/EC of the European parliament and of the council of 15 december 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. Off. J. Eur. Union L 23, 26.1.2005, 3–16.

European Union, 2008. Directive 2008/50/EC of the European parliament and of the council of 21 may 2008 on ambient air quality and cleaner air for Europe. Off. J. Eur. Union 152, 1–44.

European Union, 2024. EU directive (EU) 2024/2881 of the European parliament and of the council of 23 october 2024 on ambient air quality and cleaner air for Europe (recast). Off. J. Eur. Union 2024/2881, 20.11.2024.

Giovannini, L., Ferrero, E., Karl, T., Rotach, M.W., Staquet, C., Trini Castelli, S., Zardi, D., 2020. Atmospheric pollutant dispersion over complex terrain: Challenges and needs for improving air quality measurements and modeling. Atmosphere 11 (6), 646.

Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E., Hueglin, C., 2018. Random forest meteorological normalisation models for Swiss PM 10 trend analysis. Atmos. Chem. Phys. 18 (9), 6223–6239.

Guerreiro, C.B., Foltescu, V., De Leeuw, F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384.

Harrell, Jr., F.E., 2015. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis, second ed. In: Springer Series in Statistics, Springer Cham.

Hastie, T., Tibshirani, R., 1986. Generalized additive models. Statist. Sci. 1 (3), 297–310.

He, S., Mazumdar, S., Arena, V.C., 2006. A comparative study of the use of GAM and GLM in air pollution research. Environmetrics: Off. J. Int. Environmetrics Soc. 17 (1), 81–93.

He, H., Schäfer, B., Beck, C., 2024. Spatial analysis of tails of air pollution PDFs in Europe. Environ. Data Sci. 3, e30.

Houseman, E.A., Virji, M.A., 2017. A Bayesian approach for summarizing and modeling time-series exposure data with left censoring. Ann. Work. Expo. Health 61 (7), 773–783.

Jaiswal, A., Samuel, C., Kadabgaon, V., 2018. Statistical trend analysis and forecast modeling of air pollutants. Glob. J. Environ. Sci. Manag. 4 (4), 427–438.

Khorrami Chokami, A., Rabitti, G., 2024. An exact game-theoretic variable importance index for generalized additive models. J. Comput. Graph. Statist. 33 (4), 1276–1285.

Largeron, Y., Staquet, C., 2016. Persistent inversion dynamics and wintertime PM10 air pollution in Alpine valleys. Atmos. Environ. 135, 92–108.

Martínez-Flórez, G., Bolfarine, H., Gómez, H.W., 2014. The log-power-normal distribution with application to air pollution. Environmetrics 25 (1), 44–56.

Masiol, M., Formenton, G., Visin, F., Bonetto, A., Rovea, M., Ficotto, S., Danesin, E., Toffanin, T., Maggiulli, A., Battistel, M., et al., 2024. Sources of wintertime PM2.5 at a major city in an Alpine valley: The role of atmospheric dispersion and inversion dynamics. Atmos. Environ. 330, 120556.

Masiol, M., Squizzato, S., Formenton, G., Harrison, R.M., Agostinelli, C., 2017. Air quality across a European hotspot: Spatial gradients, seasonality, diurnal cycles and trends in the Veneto region, NE Italy. Sci. Total Environ. 576, 210–224.

Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. 15 (15), 8889–8973.

Mudelsee, M., 2019. Trend analysis of climate time series: A review of methods. Earth-Sci. Rev. 190, 310–322.

Muñoz, S.J., 2019. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), http://dx.doi.org/10.24381/cds.e2161bac.

Pearce, J.L., Beringer, J., Nicholls, N., Hyndman, R.J., Tapper, N.J., 2011. Quantifying the influence of local meteorology on air quality using generalized additive models. Atmos. Environ. 45 (6), 1328–1336.

Peng, R.D., Dominici, F., Louis, T.A., 2006. Model choice in time series studies of air pollution and mortality. J. R. Stat. Soc. Ser. A: Stat. Soc. 169 (2), 179–203.

Pivato, A., Pegoraro, L., Masiol, M., Bortolazzo, E., Bonato, T., Formenton, G., Cappai, G., Beggio, G., Giancristofaro, R.A., 2023. Long time series analysis of air quality data in the Veneto region (northern Italy) to support environmental policies. Atmos. Environ. 298, 119610.

Qiu, M., Zigler, C., Selin, N.E., 2022. Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions. Atmos. Chem. Phys. 22 (16), 10551–10566.

Ravindra, K., Rattan, P., Mor, S., Aggarwal, A.N., 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. Environ. Int. 132, 104987.

Real, E., Megaritis, A., Colette, A., Valastro, G., Messina, P., 2024. Atlas of ozone chemical regimes in Europe. Atmos. Environ. 320, 120323.

Salcedo, R., Ferraz, M.A., Alves, C., Martins, F., 1999. Time-series analysis of air pollution data. Atmos. Environ. 33 (15), 2361–2372.

Seinfeld, J.H., Pandis, S.N., 2016. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, third ed. John Wiley & Sons, Hoboken, NJ, p. 1152.

Sicard, P., Agathokleous, E., Anenberg, S.C., De Marco, A., Paoletti, E., Calatayud, V., 2023. Trends in urban air pollution over the last two decades: A global perspective. Sci. Total Environ. 858, 160064.

Sicard, P., Agathokleous, E., De Marco, A., Paoletti, E., Calatayud, V., 2021. Urban population exposure to air pollution in europe over the last decades. Environ. Sci. Eur. 33 (1), 28.

Sokhi, R.S., Moussiopoulos, N., Baklanov, A., Bartzis, J., Coll, I., Finardi, S., Friedrich, R., Geels, C., Grönholm, T., Halenka, T., et al., 2021. Advances in air quality research–current and emerging challenges. Atmos. Chem. Phys. Discuss. 2021, 1–133.

Song, Y., Wu, H., Dai, Q., Liu, X., Zhang, Y., Feng, Y., 2024. Differentiating periodic drivers of air quality changes: A two-step decomposition approach integrating machine learning and wavelet analysis. J. Geophys. Res.: Atmos. 129 (7), e2023JD039658.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.

Valotto, G., Varin, C., 2016. Characterization of hourly NOx atmospheric concentrations near the venice international airport with additive semi-parametric statistical models. Atmos. Res. 167, 216–223.

Whiteman, C.D., 2000. Mountain Meteorology: Fundamentals and Applications. Oxford University Press, New York.

Wood, S.N., 2003. Thin plate regression splines. J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (1), 95–114.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R, second ed. Chapman and Hall/CRC, New York.

Wood, S.N., Pya, N., Säfken, B., 2016. Smoothing parameter and model selection for general smooth models. J. Amer. Statist. Assoc. 111 (516), 1548–1563.

World Health Organization, 2021. WHO Global Air Quality Guidelines: Particulate Matter (PM2. 5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization, Geneva.

Yan, Y., Lin, J., Pozzer, A., Kong, S., Lelieveld, J., 2019. Trend reversal from high-to-low and from rural-to-urban ozone concentrations over Europe. Atmos. Environ. 213, 25–36.

Zhang, R., Awang, N., Feng, J., Ma, X., 2025. Spatio-temporal long short-term memory neural network with seasonal-trend decomposition for ambient air pollutant forecasting. Earth Sci. Inform. 18 (1), 112.