# A Latent Curve Model to estimate the evolution of urban air pollution

Bottazzi Schenone Mariaelena[1] and Grimaccia Elena[2]

[1] Department of Statistical Sciences, Sapienza University, Rome (Italy),
mariaelena.bottazzischenone@uniroma1.it
[2] ISTAT - Italian National Institute of Statistics, Rome (Italy),
elgrimac@istat.it

**Abstract.** The goal of this study is to evaluate the evolution of the latent construct of air pollution, accounting for the interplay of time-varying covariates, specifically temperature variations, and geographical factors such as latitude and longitude. This is achieved employing a Latent Curve Model that quantifies and interprets the changes of the Air Pollution Index in Europe across the years 2019 to 2023. This analysis is crucial for discerning long-term trends in air pollution, providing valuable insights into potential environmental shifts that may impact public health and guide future policy considerations. Results show an increase of European air pollution level in the last five years. This trend is closely linked with increasing temperatures, underscoring the urgent need to address the intertwined challenges of air quality and climate change.

**Keywords:** Latent Curve Model, air pollution, climate change, European urban areas.

## 1 Introduction

Monitoring air pollution is crucial as it plays an important role in safeguarding public health and the environment. In recent years, the impact of climate change has become evident, stressing even more the challenges associated with air quality [1]. Rising global temperatures, altered precipitation patterns, and changing weather conditions contribute to the intensification of air pollution. Using a statistical technique to model air pollution, we gain insights into the dynamic interplay between climate change and air quality, enabling the development of effective strategies to mitigate the adverse effects on the human health and the ecosystem. Indeed, comprehensive air quality studies not only enhance our understanding of environmental changes but also empower policy makers to implement targeted interventions to address the evolving challenges posed by climate-induced alterations in air composition. In this context, many different air pollution models have been proposed for short-term forecast, long-term projection, air-pollutant exposure and risk assessments, and most of them includes spatial and spatio-temporal effects. For instance, Taghavi et al. [2] use Functional Principal Component Analysis in the spatio-temporal land-use regression modeling of Particulate Matter 2.5, while Ahmadi Basiri et al. [3] employ Functional Kriging for spatio-temporal modeling of Nitrogen Dioxide in Middle East. Multiscale geographical and temporal weighted Regression Methods are used by Yue et al. [4] and Li et al. [5] to model the determinants of Particulate Matter 2.5 in China, considering localized spatio-temporal effects. In this paper, Latent Curve Models (LCMs) are employed to analyse the dynamic trends in the multidimensional latent construct of air pollution. These flexible models can describe the evolution of air pollution over time, taking into account temperature changes as well as spatial effects.

This paper is structured as follows. Section 2 illustrates the data employed in the empirical study, while in Section 3 LCMs are theoretically presented. Section 4 is devoted to the air pollution analysis for 106 European urban areas between 2019 and 2023. Finally, in Section 5, concluding remarks are drawn and the key findings and insights derived from the empirical study are summarized.

## 2 Data

The data on which the study is conducted come from the Wordwide Air Quality database (https://aqicn.org/), and refer to the concentrations of the six main air pollutants, as identified by the Environmental Protection Agency: Ground-level ozone ($O_3$), Particle pollution (also known as Particulate Matter (PM), including $PM2.5$ and $PM10$), Carbon monoxide ($CO$), Sulfur dioxide ($SO_2$) and Nitrogen dioxide ($NO_2$) for 106 European urban areas for the years 2019 to 2023. For each urban area and year, the pollutants concentrations are computed as the average over the 365 days of the daily median values. These variables have been firstly employed to construct an Air Pollution Index (API) through Structural Equation Models (SEM), as in Bottazzi Schenone et al. [10]. This index ranges in $[0-1]$, where 1 corresponds to the highest pollution level, and 0 to the lowest, after a min-max normalization procedure. To quantify the impact of climatic changes on the Air Pollution Index, also temperature, humidity, wind gust, wind speed and pressure variables are considered for each urban area, from the same database. Latitude and longitude are also taken into account to control for spatial effects [6].

## 3 Methodology

Latent Curve Models are widely employed methodologies in longitudinal data analysis to examine the dynamic relationships among multiple variables over time. These models integrate latent variables, which represent unobservable constructs, with observed variables to describe underlying trends, patterns, and interdependencies that characterize the phenomenon under study. In our case, we will employ a latent curve to explore the evolution of the Air Pollution Index in Europe in the last five years.

Consider a population of $N$ units observed at $T$ times. A Latent Curve Model can be mathematically formalized as follows [7] [8]. In a general notation, let $y_{it}$

be the observations of the response variable for unit $i$ ($i = 1, \ldots, N$) at time $t$ ($t = 1, \ldots, T$), $\alpha_i$ the intercept associated to the $i$-th unit ($\boldsymbol{\alpha} = \alpha_1, \ldots, \alpha_N$), $\beta_i$ the slope associated to the $i$-th unit ($\boldsymbol{\beta} = \beta_1, \ldots, \beta_N$) and $\lambda_t$ the time trend variable. Consider also $J$ time-varying covariates, with $w_{jit}$ denoting the $j$-th covariate for unit $i$ at time $t$ and $\gamma_{jt}$ representing the corresponding coefficient. Let $\epsilon_{it}$ be the errors associated with the $i$-th unit at time $t$. The equation determining the trend of the response variable $y_{it}$ is:

$$y_{it} = \alpha_i + \lambda_t \beta_i + \sum_{j=1}^{J} \gamma_{jt} w_{jit} + \epsilon_{it}. \tag{1}$$

A common convention for the temporal coefficients is to set, for each $t = 1, \ldots, T$, $\lambda_t = t - 1$ for linear trends. Other typical assumptions in regression analyses are that errors are centred ($\mathbb{E}(\epsilon_{it}) = 0 \ \forall i = 1, \ldots, N, \ t = 1, \ldots, T$) and normally distributed, and that $\alpha_i$ and $\beta_i$ are uncorrelated with $\epsilon_{it}$ $\forall i = 1, \ldots, N, \ t = 1, \ldots, T$. The equations describing the intercept and the slope are:

$$\begin{cases} \alpha_i &= \mu_{\boldsymbol{\alpha}} + \sum_{q=1}^{Q} \delta_{\alpha q} x_{qi} + \xi_{\alpha i} \\ \beta_i &= \mu_{\boldsymbol{\beta}} + \sum_{q=1}^{Q} \delta_{\beta q} x_{qi} + \xi_{\beta i}, \end{cases} \tag{2}$$

where $\mu_{\boldsymbol{\alpha}}$ and $\mu_{\boldsymbol{\beta}}$ are the mean intercept and slope, respectively. A total of $Q$ time-invariant covariates is considered in the model. The $q$-th time invariant covariate for the $i$-th unit is $x_{qi}$, with coefficients $\delta_{\boldsymbol{\alpha} q}$ and $\delta_{\boldsymbol{\beta} q}$. The residuals $\xi_{\alpha i}$ and $\xi_{\beta i}$ have zero mean are normally distributed and are uncorrelated with the $\epsilon_{it}$. Estimates are obtained within the Maximum Likelihood framework [9].

## 4 Application

In this Section, a LCM is applied to air pollution, measured by the Air Pollution Index, considering 106 European urban areas from 2019 to 2023. The air pollution situation in Europe according to API is depicted in Fig. 1 for the year 2023.

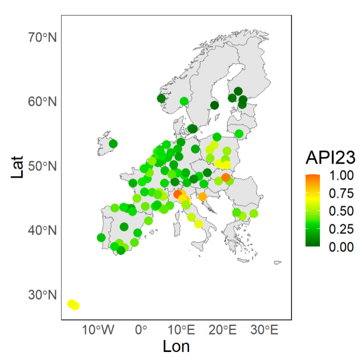


**Fig. 1.** Air Pollution Index values for European urban areas in 2023.

According to API, the most polluted areas, marked in orange and yellow, are located in Italy, Hungary, and Croatia. Notably, the two yellow points at the bottom-left of the figure are Tenerife and Las Palmas: these two urban areas were among the less polluted ones until 2021, while their air quality conditions have deeply worsened in the latest two years, due to the increasing population (see also Tab. 1).

Tab. 1 presents a ranking of European urban areas based on their API scores over the considered five-year period. Top ranks refer to the most polluted areas.

**Table 1.** Most (top ranks) and less (bottom ranks) polluted urban areas in Europe, according to the API ranking (2019-2023).

| **Rank** | **2019** | **2020** | **2021** | **2022** | **2023** |
|---|---|---|---|---|---|
| 1 | Katowice | Milan | Modena | Milan | Milan |
| 2 | Milan | Parma | Katowice | Modena | Miskolc |
| 3 | Miskolc | Modena | Parma | Zagreb | Zagreb |
| 4 | Plovdiv | Miskolc | Miskolc | Brescia | Modena |
| 5 | Brescia | Plovdiv | Cracow | Parma | Brescia |
| ... | ... | ... | ... | ... | ... |
| 102 | Tenerife | Stockholm | Dublin | Stockholm | Malmo |
| 103 | Tallinn | Helsinki | Tenerife | Tampere | Turku |
| 104 | Tampere | Bergen | Santander | Turku | Helsinki |
| 105 | Las Palmas | Tallinn | Zurich | Helsinki | Tampere |
| 106 | Bergen | Dublin | Turku | Tallinn | Stockholm |

Milan appears to have the highest rank for three years (2020,2022,2023), indicating a persistent air pollution issue. It is also possible to note a shift in the rankings of certain areas such as Tenerife, Las Palmas and Dublin, which were in the last positions until 2021, suggesting a recent change in their air quality conditions. Less polluted zones remain in Finland and Sweden across all the time period considered, while we can notice an increased number of Italian urban areas in the top ranking positions.

A LCM is applied considering the APIs scores as response variables ($N = 106$, $T = 5$), the meteorological variables as time-varying covariates ($J = 5$) and latitude and longitude as time-invariant covariates ($Q = 2$). From the complete model with all the covariates, we proceed with a backward selection, excluding at each step the non-statistically significant covariates. As a result, we selected the model shown in Fig. 2 ($J = 1, Q = 2$). The normality of model's residuals, computed as difference between the API score and the estimated one according to (1) for each year, has been assessed through the Shapiro-Wilks test.

The intercept coefficient $\mu_{\boldsymbol{\alpha}}$ corresponds to the average European API in 2019 and it is equal to 0.592, with a standard deviation of 0.015. The slope coefficient $\mu_{\boldsymbol{\beta}}$ captures the linear trend in the Air Pollution Index over the last five years, and it is equal to 0.011, with a standard deviation of 0.003. This positive coefficient highlights a small, yet statistically significant increase in the European average API between 2019 and 2023. The first model's finding is, indeed, the significant growth of air pollution during the period in analysis. Furthermore, the positive relationship between $\boldsymbol{\beta}$ and temperature over time suggests a substantial influence of urban temperatures on escalating pollution levels. These findings not only confirm but also quantify what is known in literature [11] [12], offering valuable insights specific for the European context.

The model provides also the coefficients associated to temperature on API over time. The coefficients' estimates are positive, indicating that urban areas' temperature has a statistically significant effect on the increasing air pollution level. Moreover, the fact that they have strongly increased in the last two years warns about the impact that increasing temperatures due to climate changes may have on air pollution.

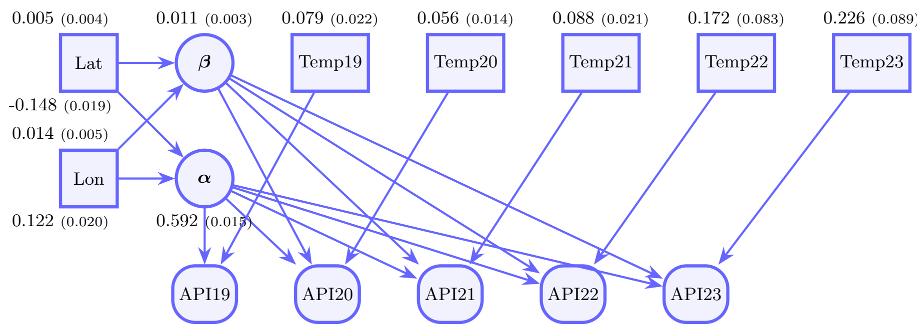


**Fig. 2.** Path diagram of the API LCM, estimates' standard errors are reported in parentheses.

Finally, the model reveals disparities within the European area. The positive longitude coefficient and negative latitude coefficient associated to the intercept show that Northern and Western Europe urban areas generally experience lower pollution levels. This observation is consistent with extant environmental research. Notably, the latitude's non statistically significant effect on the slope indicates that the pollution level divergence between Northern and Southern European urban areas has not widened over time: it may even suggest a slight convergence. Conversely, the significant positive coefficient for longitude points to a growing divergence, indicating that Eastern European urban areas are becoming increasingly more polluted than their Western counterparts.

## 5 Concluding remarks

In this study, we present a Latent Curve Model tailored to explore the dynamics of multiple variables over time and space, focusing on the relationship between air pollution and temperature. By adjusting for spatial covariates, our approach mitigates potential biases in the estimates that might arise if these factors were overlooked.

Through this work, we achieve several key findings. Primarily, the analysis reveals a concerning escalation in air pollution across Europe over the last five years, underscoring the urgency of addressing this environmental challenge. Additionally, the proposed model reveals a significant association between climatic changes and pollution levels, indicating a concurrent rise in temperature and contaminants' concentration. The model's capability to assess spatial covariates further enriches our understanding, showing an almost stable difference over time among Northern and Southern European areas and a slight increase of air pollution in Eastern Europe.

Future research could be devoted to identify and integrate additional covariates into the model, also considering non linear interactions among them. Such inclusions would not only enhance model's comprehensiveness and accuracy, but also improve its predictive capability. Furthermore, by adopting a more granular temporal resolution such as monthly intervals instead of yearly, it would be possible to gain more time-detailed insights into the phenomenon under study, addressing for instance seasonality effects.

## References

1. Pinho Gomes, A. C., Roaf, E., Fuller, G., Fowler, D., Lewis, A., ApSimon, H., Noakes, C., Johnstone, P., Holgate, S.: Air pollution and climate change. The Lancet. Planetary health 7(9), e727–e728 (2023). https://doi.org/10.1016/S2542-5196(23)00189-4
2. Taghavi, M., Ghanizadeh, G., Ghasemi, M., Fassò, A., Hoek, G., Hushmandi, K., Raei, M.: Application of Functional Principal Component Analysis in the Spatiotemporal Land-Use Regression Modeling of PM2.5. Atmosphere 14(6), 926 (2023). https://doi.org/10.3390/atmos14060926
3. Ahmadi Basiri, E., Taghavi-Shahri, S.M., Mahaki, B., Amini, H.: Functional Kriging for Spatiotemporal Modeling of Nitrogen Dioxide in a Middle Eastern Megacity. Atmosphere 13(7), 1095 (2022). https://doi.org/10.3390/atmos13071095
4. Yue, H., Duan, L., Lu, M., Huang, H., Zhang, X., Liu, H.: Modeling the Determinants of PM2.5 in China Considering the Localized Spatiotemporal Effects: A Multiscale Geographically Weighted Regression Method. Atmosphere 13(4), 627 (2022). https://doi.org/10.3390/atmos13040627
5. Li, Q., Li, X., Li, H.: Factors Influencing PM2.5 Concentrations in the Beijing–Tianjin–Hebei Urban Agglomeration Using a Geographical and Temporal Weighted Regression Model. Atmosphere 13(3), 407 (2022). https://doi.org/10.3390/atmos13030407
6. Urdangarin, A., Goicoa, T., Ugarte, M. D.: Evaluating recent methods to overcome spatial confounding. Revista Matemática Complutense 36, 333–360 (2022). https://doi.org/10.1007/s13163-022-00449-8
7. Bollen, K.A., Curran, P.J.: Latent Curve Models: A Structural Equation Perspective. John Wiley & Sons, Inc., New York (2006).
8. Allison, P. D.: Fixed effects regression models. SAGE Publications, Inc., New York (2009).
9. Meredith, W., & Tisak, J.: Latent curve analysis. Psychometrika 55 (1), 107-122 (1990). https://doi.org/10.1007/BF02294746
10. Bottazzi Schenone, M., Grimaccia, E., Vichi, M.: Computational assessment of k-means clustering on a Structural Equation Model based index. In: Book of Short Papers SIS 2023 - Scientific meeting of the Italian Statistical Society, pp. 978-983. Pearson, Milano (2023).
11. Jacob, D.J., Winner, D.A.: Effect of climate change on air quality. Atmospheric Environment 43(1), 51-63 (2009). https://doi.org/10.1016/j.atmosenv.2008.09.051
12. The Royal Society: Effects of net-zero policies and climate change on air quality. DES7687 (2021). ISBN: 978-1-78252-558-5