

ENVIRONMENTAL RISK ASSESSMENT VIA NONHOMOGENEOUS HIDDEN SEMI-MARKOV MODELS WITH PENALIZED VECTOR AUTOREGRESSION

BY MARCO MINGIONE^{1,a}, PIERFRANCESCO ALAIMO DI LORO^{2,b},
 FRANCESCO LAGONA^{3,c} AND ANTONELLO MARUOTTI^{4,5,2,d}

¹*Department of Sports, Human and Health Sciences, University of Rome “Foro Italico”, marco.mingione@uniroma4.it*

²*Department Law, Economics, Politics and Modern Languages, LUMSA University, p.alaimodiloro@lumsa.it*

³*Department of Political Sciences, Roma Tre University, francesco.lagona@uniroma3.it*

⁴*Department of Public Health and Epidemiology, Khalifa University, antonello.maruotti@ku.ac.ae*

⁵*Center for Biotechnology, Khalifa University*

Motivated by the study of pollution trends in the city of Bergen, we introduce a flexible statistical framework for modeling multivariate air pollution data via a nonhomogeneous hidden semi-Markov vector autoregression. The hidden process captures unobserved environmental conditions, while the vector autoregressive structure accounts for temporal autocorrelation and cross-pollutant dependencies. The model further allows time-varying environmental conditions to influence both the average levels of pollutant concentrations and the duration of different transient states. Parameters are estimated via maximum likelihood using a tailored expectation-maximization (EM) algorithm, integrated with state-specific ℓ_1 regularization to control overfitting and automatically select relevant temporal lags. The proposal is tested on simulated data under different scenarios and then applied to daily concentrations of nitrogens and particulate matter recorded in an urban area. Environmental risk is assessed by a Shapley value-based decomposition that attributes marginal risk contributions. This approach offers a comprehensive framework for multivariate environmental risk modeling, enabling better identification of high-pollution episodes and informing policy interventions.

1. Introduction. Identifying the sources and patterns of air pollution is essential for planning policy interventions meant to preserve the environment and public health (Shan et al. (2024)). Traditional methods have largely focused on single-pollutant models (Martinez-Zarzoso and Maruotti (2013), Liang et al. (2021), Maruotti and Alaimo Di Loro (2023)); however air pollutants do not occur in isolation (Finazzi, Scott and Fassò (2013), Merlo et al. (2022), Cao (2024), Zhu et al. (2024)) but are emitted and dispersed jointly. They often share common sources and co-occur under specific environmental conditions, which influence how they interact through atmospheric chemistry, particularly during episodes of atmospheric instability when simultaneous surges in pollutant concentrations occur. This makes the environmental risk associated with air pollution a fundamentally multivariate phenomenon that exhibits complex mixtures of behaviors (Boaz, Lawson and Pearce (2019), Baragaño et al. (2022)). As such, neglecting their joint dynamics can hinder a comprehensive understanding of pollution patterns.

To address these issues, air pollution modeling must move beyond univariate approaches and adopt multivariate statistical frameworks that reflect the true complexity of pollution mixtures (Maruotti et al. (2017), Bouveyron et al. (2022)). We develop a general and flexible modeling approach that reflects the empirical features observed in air quality measurements—features that are not only specific to air pollution trends but also broadly relevant across other applications involving multivariate time series. We assume that pollutant concentrations can be generated under multiple unobserved environmental conditions that can favor or limit the

accumulation of air pollutants. These states can be inferred from the data via a time-varying mixture model, where the hidden state dynamics are governed by a semi-Markov process (Barbu and Limnios (2009), Yu (2015), Ruiz-Suarez, Leos-Barajas and Morales (2022)). Unlike nonhomogeneous Markov models (Hoskovec et al. (2023)), the semi-Markov formulation allows for sojourn times that are not necessarily geometrically distributed, enabling a more realistic representation of the persistence and duration of pollution episodes. Moreover, we allow for the inclusion of covariates' effects (e.g., meteorological conditions) in the sojourn distribution, introducing an additional layer of flexibility in how long different environmental states might last (Lagona and Mingione (2025), Koslik (2025)). Explicitly modeling the duration of each latent state and the parametric effects of covariates on transition probabilities is essential for uncovering the mechanisms that drive shifts between pollution regimes. Given the hidden state, we assume that the observed multivariate process can be described by a Gaussian vector autoregressive (VAR) model with state-dependent parameters (Hadj-Amar, Jewson and Vannucci (2024)). This formulation allows accounting for temporal autocorrelation, which is a prominent feature of atmospheric pollutant series due to accumulation and transport mechanisms, and the instantaneous cross-dependence, allowing it to vary across different environmental conditions. For instance, pollutants' accumulation and correlations can strengthen during high-pollution episodes and weaken under cleaner conditions. This enables the model to reflect short-term dependencies and Granger-causal relationships among pollutants, capturing both feedback dynamics and temporal clustering in their evolution. Though the model is very flexible, the computational burden is not cumbersome. Parameters are estimated by an expectation-maximization (EM) algorithm with a regularized M-step. Regularization is obtained by integrating the M-Step with a ℓ_1 penalty on the autoregressive coefficients, considering a state-dependent penalty as in Städler and Mukherjee (2013). In the absence of appropriate asymptotic results, estimation uncertainty is quantified by parametric bootstrap.

Drawing on the extensive literature on financial risk assessment, we adapt the multiple-CoVaR (MCoVaR) and multiple-CoES (MCoES) methodologies to the environmental domain, emphasizing the importance of capturing the multidimensional nature of environmental risk. To this end, we adopt the multivariate risk framework introduced by Adrian and Brunnermeier (2016) and extended to dynamic mixtures in Bernardi, Maruotti and Petrella (2017). In particular, we implement a *standardized* Shapley value-based risk attribution methodology to ensure an equitable decomposition of each pollutant's risk.

The model is motivated by an analysis of daily pollutant concentration levels of NO, NO₂, PM₁, PM_{2.5} and PM₁₀ at Danmarksplass, the busiest traffic intersection in Bergen. It has garnered media attention due to poor air quality during cold, dry periods with minimal wind. Approximately 40,000 vehicles traverse this intersection daily and the surrounding contains a significant proportion of old wooden houses heated by wood-burning stoves (Bergen Kommune (2023)). The results reveal notable features of the multivariate time series of pollutants, particularly the emergence of two distinct states characterized by low and high concentration levels. Beyond differences in average pollutant levels, these states also exhibit markedly different correlation structures and temporal dynamics. Specifically, we notice that, in the absence of favorable meteorological conditions, the high-pollution state may persist indefinitely. In terms of risk contribution, the analysis identifies the nitrogens (NO and NO₂) and PM_{2.5} and PM₁₀ as mutually reinforcing in terms of risk. By contrast, PM₁ exhibits a more autonomous behavior, with a seasonally varying relationship to PM_{2.5}. This seasonal asymmetry may reflect the shifting composition of pollution sources across the year.

2. Data description. We focus on daily air pollutant concentrations of particulate matter across three size categories (PM₁₀, PM_{2.5} and PM₁) as well as nitrogen oxides (NO and

NO_2), recorded from January 1, 2020, to December 31, 2022. All concentrations are measured in micrograms per cubic meter ($\mu\text{g}/\text{m}^3$). By definition, particulate matter concentrations of increasing size are intrinsically dependent. Specifically, $\text{PM}_{10} = \text{PM}_{2.5} + \text{PM}_{10}^*$ and $\text{PM}_{2.5} = \text{PM}_1 + \text{PM}_{2.5}^*$, where PM_{10}^* and $\text{PM}_{2.5}^*$ represent residual particulate matter concentrations corresponding to particles sized between 2.5–10 and 1–2.5, respectively. This structure reflects a natural hierarchy in particulate matter composition: PM_1 includes the smallest particles, and larger particle categories are constructed by sequentially adding residual mass. Consequently, any valid probabilistic modeling of the PM variables must respect the ordering constraint $\text{PM}_{10} \geq \text{PM}_{2.5} \geq \text{PM}_1$. To address and disentangle this inherent dependence structure, we work with the residual PM components. We retain PM_1 in its original form and define the residual variables as $\text{PM}_{10}^* = \text{PM}_{10} - \text{PM}_{2.5}$ and finally $\text{PM}_{2.5}^* = \text{PM}_{2.5} - \text{PM}_1$.

Additionally, the raw concentration data are log-transformed to accommodate the multivariate Gaussian assumption of the emission density within the VAR model discussed in Section 3.1. The use of log-concentrations is well established in the environmental literature (Liao et al. (2021)), supported both by physical justifications for log-normality (Ott (1990), Andersson (2021)) and by the interpretability of log-scale relationships as elasticities. From this point onward, unless otherwise specified, the terms PM_{10}^* , $\text{PM}_{2.5}^*$, PM_1 , NO and NO_2 refer to the log-transformed pollutant concentrations. Figure 1 presents the observed time series of daily pollutant concentrations. The data exhibit a broadly stationary pattern with yearly seasonality, while closely occurring spikes suggest a degree of temporal dependence. These patterns are further emphasized by the partial autocorrelation functions reported in the Supplementary Material (Mingione et al. (2026)), together with both the marginal and joint (pairwise) distributions of the log-transformed concentrations, supporting the Gaussian assumptions and the existence of possible heterogeneity.

To enrich the analysis, these pollution data are merged with meteorological data that are collected hourly from the nearest high-quality weather station located at Florida, approximately one kilometer from Danmarksplass. These meteorological variables have been aggregated on the daily scale to match the pollutant resolutions and can be used as covariates for

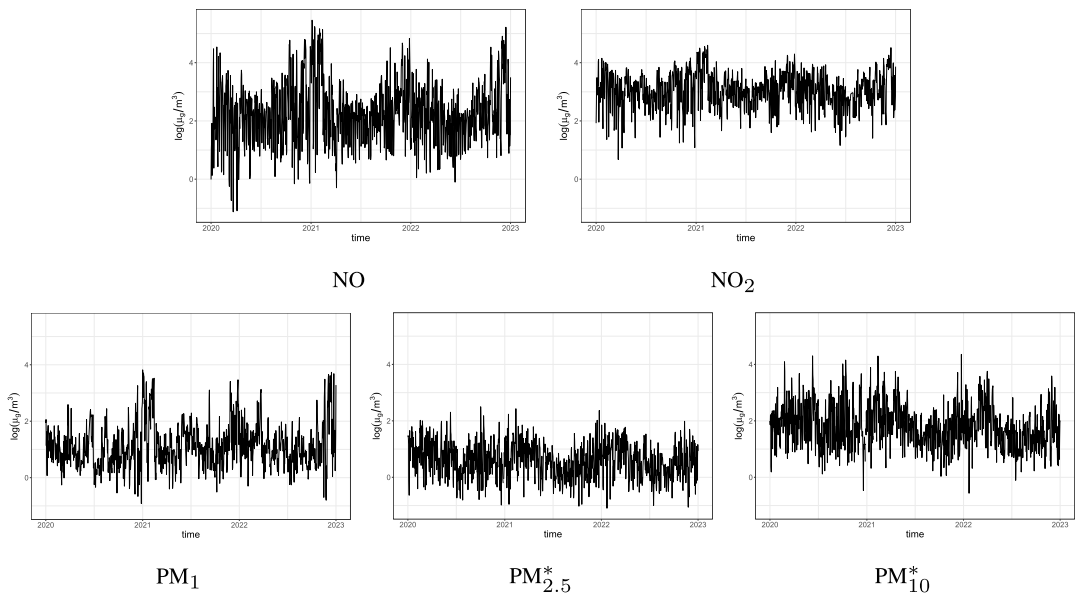


FIG. 1. Observed time series (on the log-scale) of the five considered pollutants over the whole study period: (a) NO , (b) NO_2 , (c) PM_1 , (d) $\text{PM}_{2.5}^*$, (e) PM_{10}^* .

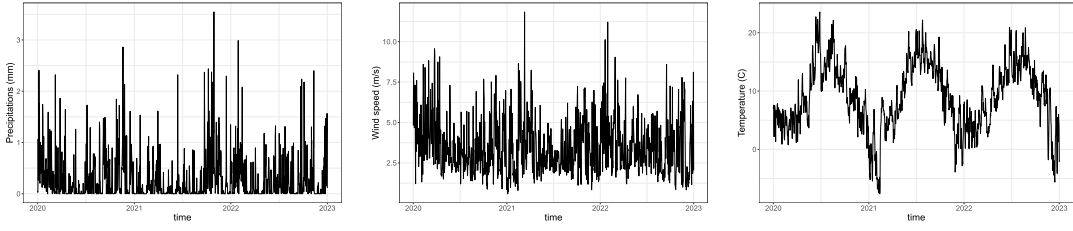


FIG. 2. Observed time series of the available covariates for the whole study period: (a) total precipitation (millimeters), (b) wind speed (m/s) and (c) temperature (C°).

modeling the pollutant concentrations dynamics. In particular, we consider the daily total precipitation (in millimeters), daily average wind speed (in meters per second), and daily average temperature (in degrees Celsius); see Figure 2. Bergen’s geographic location—the North Atlantic Ocean and situated along the Byfjorden—gives rise to an oceanic climate characterized by mild summers, cool and wet winters, and generally high humidity. The proximity to the ocean exerts a thermal buffering effect, moderating seasonal temperature variations: average temperatures hover around 9°C, with winters being relatively mild and summers cooler than those experienced in more inland regions. Therefore, we expect the daily average temperature to be a reasonable proxy for the seasonal patterns visible in the average level of pollutant concentrations’. Furthermore, the city’s coastal position allows moist Atlantic air to funnel inland via the fjord, resulting in frequent precipitation and moderate wind conditions year-round. During the observation period, the longest recorded spell without precipitation lasted 11 consecutive days, although, on average, rainfall occurred at intervals no longer than three days. As rain and wind are considered two of the major drivers in abruptly reducing the concentration of air pollutants (Ouyang et al. (2015), Zhang et al. (2018)), we expect to see similar patterns in the transitions from high-pollution states to lower ones. Finally, we also enrich the available data by building up a weekend effect covariate from the recording date. This effect might be useful to adjust for the inherent weekly seasonality of pollution data related to the typical patterns of human activities (Tavella et al. (2023)).

3. Methods. Multivariate time series data can be expressed in the form of a $T \times p$ array of observations, say $\mathbf{Y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_T^\top]$ with $\mathbf{y}_t = [y_{tj}]_{j=1}^p$, where each single entry y_{tj} denotes the value of the j th outcome at time t . These observations are viewed as a realization of a stochastic process $\{\mathbf{Y}_t : t = 1, \dots, T\}$, whose joint distribution is more tractable when conditioned on hidden states. We model this unobserved heterogeneity through a time dependent hidden process $\mathbf{u} = \{\mathbf{u}_t : t = 1, \dots, T\}$, where $\mathbf{u}_t = [u_{t1}, \dots, u_{tK}]$ is a multinomial random variable with one trial and K states. The process is said to be in state k at time t if $u_{tk} = 1$. The law of $\mathbf{Y}$ can then be specified by a hierarchical model that combines a parametric conditional distribution of the observations, given the hidden states $f(\mathbf{Y}|\mathbf{u}; \boldsymbol{\theta})$ (the observation process), with a parametric model for the latent chain $p(\mathbf{u}; \boldsymbol{\eta})$ (the hidden process). The corresponding marginal distribution of the observed data can be obtained through marginalization of the hidden process as

$$(1) \quad f(\mathbf{Y}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{\mathbf{u}} f(\mathbf{Y}|\mathbf{u}; \boldsymbol{\theta}) \cdot p(\mathbf{u}; \boldsymbol{\eta}).$$

3.1. The observation process. Hierarchical time-series models involving hidden states can be simplified by assuming that the observation vectors $\mathbf{y}_t, t = 1, \dots, T$, are conditionally independent given the corresponding hidden state $\mathbf{u}_t$. Under this assumption, the conditional distribution of the full data conveniently factorizes as $f(\mathbf{Y}|\mathbf{u}) = \prod_{t=1}^T \prod_{k=1}^K f(\mathbf{y}_t; \boldsymbol{\theta}_k)^{u_{tk}}$,

where θ_k denotes the vector of parameters associated with state k . This formulation implies that each multivariate observation is independent of past and future observations, conditional on the hidden state at time t . Consequently, the observed process is fully characterized by specifying a model for the conditional distribution $\mathbf{y}_t | \mathbf{u}_t$. A popular choice is the multivariate Gaussian distribution $\mathbf{y}_t | \mathbf{u}_{tk} = 1 \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $t = 1, \dots, T$, where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the state-dependent mean vector and covariance matrix, respectively. This choice can capture the within-state dependence structure among the p components of $\mathbf{y}_t$.

However, in many real-world applications—particularly in environmental contexts—the assumption that the state-specific mean vector $\boldsymbol{\mu}_k$ remains constant over time can be overly restrictive. Empirical evidence often indicates that multivariate time series exhibit autoregressive dependencies, with current outcomes directly influenced by past observations. As a result, assuming conditional independence of observations given only the hidden state may be overly simplistic. To accommodate this additional temporal structure, we relax the conditional independence assumption and assume that each observation $\mathbf{y}_t$ is conditionally independent of the rest of the process given both the hidden state $\mathbf{u}_t$ and a finite history of past observations up to lag H . In this setting, the distribution of the observed process factorizes as:

$$f(\mathbf{Y} | \mathbf{u}) = \prod_{t=1}^T \prod_{k=1}^K f_k(\mathbf{y}_t | \mathbf{y}_{(t-H):(t-1)}; \theta_k)^{u_{tk}},$$

where $\mathbf{y}_{(t-H):(t-1)} = [\mathbf{y}_{(t-H)}, \dots, \mathbf{y}_{(t-1)}]$. More specifically, we model the conditional distributions using a state-dependent vector autoregressive model of order H (VAR(H)),

$$(2) \quad f_k(\mathbf{y}_t | \mathbf{y}_{(t-H):(t-1)}; \theta_k) = \mathcal{N}\left(\mathbf{b}_k + \sum_{h=1}^H \mathbf{A}_{kh} \mathbf{y}_{t-h}, \boldsymbol{\Sigma}_k\right),$$

where $\mathbf{b}_k$ is a $p \times 1$ baseline mean vector and $\mathbf{A}_{kh}$, $h = 1, \dots, H$ are $p \times p$ autoregressive coefficient matrices that capture temporal dependence across components and lags within each state.

When exogenous covariates are available in the form of a $T \times J$ matrix $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_T^\top]$, we incorporate their influence into the baseline level through a linear regression term $\mathbf{b}_k(\mathbf{x}_t) = \mathbf{b}_{k0} + \mathbf{B}_k \mathbf{x}_t^\top$, where $\mathbf{B}_k$ is a $p \times J$ matrix of regression coefficients. In our case study, only meteorological factors have been included, but further applications might also include policy interventions or demographic information.

Within this framework, the conditional mean of the observation process is dynamically updated based on both historical observations and contemporaneous exogenous variables. We notice that the VAR process within each state k is stable whenever the eigenvalues of the corresponding companion matrix lie strictly within the unit circle. This condition guarantees that if the process were to remain in state k indefinitely, it would exhibit stationary behavior rather than diverging. In particular, if the VAR process of all states $k = 1, \dots, K$ is stable, then the resulting *mixture of VAR* processes $\{\mathbf{Y}_t : t = 1, \dots, T\}$ also inherits stability properties (Fong et al. (2007)).

3.2. The hidden process. We assume the hidden process to be the realization of a K -states multinomial chain, which is a discrete-time stochastic process $\mathbf{u} = \{\mathbf{u}_t : t = 1, \dots, T\}$. Its joint distribution can be expressed in the following conditional factorized form:

$$(3) \quad p(\mathbf{u}) = p(\mathbf{u}_1) \cdot \prod_{t=2}^T p(\mathbf{u}_t | \mathbf{u}_{t-1}, \dots, \mathbf{u}_1),$$

where $p(\mathbf{u}_1) = \prod_{k=1}^K \pi_k^{u_{1k}}$ is the initial distribution at time $t = 1$ and $p(\mathbf{u}_t | \mathbf{u}_{t-1}, \dots, \mathbf{u}_1)$ denotes the conditional distribution at times $t = 2, \dots, T$, given the full history of the process. The time-dependence properties of the chain is characterized by specific assumptions on the conditional distributions building up equation (3). In a *semi-Markov chain*, they are assumed to be conditionally independent on the full history of the process, given the time of the last transition. This is conveniently summarized by the state of the chain at time $t - 1$, say k , and the current dwell time d_{t-1} , which is the time spent in k since the last transition. The behavior is then captured by the transition probability from any state $k = 1, \dots, K$ to state $j = 1, \dots, K$ after having sojourned d times in state k , which we denote as $\gamma_{kj}(d) = P(u_{tj} = 1 | u_{(t-1)k} = 1, d_{t-1} = d)$. We say that $\mathbf{u}$ is a semi-Markov chain if the conditional distributions can be expressed as

$$(4) \quad p(\mathbf{u}_t | \mathbf{u}_{t-1}, \dots, \mathbf{u}_1) = \prod_{k=1}^K \prod_{j=1}^K \gamma_{kj}(d_{t-1})^{u_{tkj}}, \quad t = 2, 3, \dots, T,$$

where $u_{tkj} = u_{(t-1)k} \cdot u_{tj}$. The specification of the statistical model for a (homogeneous) semi-Markov chain is traditionally completed by assuming a parametric model for the sojourn distribution, say $p_k(d)$, and the conditional transition probability $K \times K$ matrix $\Omega = [\omega_{kj}]$, where $\omega_{kj} = P(u_{tj} = 1 | u_{tk} = 0, u_{(t-1)k} = 1)$. Popular choices for the sojourn distribution are the shifted Poisson, the shifted negative binomial, and the logarithmic (see, e.g., [O'Connell and Højsgaard \(2011\)](#)), but any distribution on the positive integers is admissible. The extension to inhomogeneous semi-Markov chains is achieved by allowing the dwell-time distribution to depend on a covariate profile $\mathbf{z}^\top$, say $p_k(d; \mathbf{z}^\top)$. However, this formulation restricts the role of covariates to the onset of a stay, after which the dwell-time distribution remains fixed in determining its length ([Ricciotti et al. \(2025\)](#)). This is an overly restrictive limitation in studies where the stay of the system in a specific latent regime is instantaneously influenced by time-varying weather conditions, as in the air pollution study motivating this work.

To obtain a more convenient parametric specification of the transition probabilities in equation (4), we follow an alternative approach recently suggested by [Lagona and Mingione \(2025\)](#), who induce a model on $p_k(d)$ by means of the probability of leaving state k after a dwell time d ,

$$q_k(d) = P(u_{tk} = 0 | u_{(t-1)k} = 1, d_{t-1} = d), \quad k = 1, \dots, K.$$

Under this setting, the transition probabilities are obtained as

$$(5) \quad \gamma_{kj}(d_{t-1}) = \begin{cases} 1 - q_k(d_{t-1}) & j = k, \\ q_k(d_{t-1}) \cdot \omega_{kj} & j \neq k, \end{cases} \quad k, j = 1, \dots, K.$$

In (5), $q_k(d)$ plays the role of a discrete hazard function, as it indicates the probability of switching state given that the chain has “survived” d times in a state k . Note that the two approaches define equivalent quantities, as there is a one-to-one correspondence between hazards and probability distributions of sojourn times:

$$(6) \quad p_k(d) = q_k(d) \prod_{\delta=1}^{d-1} (1 - q_k(\delta)), \quad k = 1, \dots, K.$$

If the hazard is constant, say $q_k(d) = q_k$, then (6) reduces to a geometric distribution. If this holds for any state $k = 1, \dots, K$, then $\gamma_{kj}(d) = \gamma_{kj}$ do not depend on the dwell time and $\mathbf{u}$ reduces to a Markov chain with transition probabilities γ_{kj} . Otherwise, the Markovianity is

lost as the transition probabilities are endogenously updated by the dwell time of the current state.

Drawing from the survival analysis literature, the hazards can be expressed in a generalized linear model fashion as $g(q_k(d_{t-1})) = \beta_{k0} + \beta_{k1}d_{t-1}$, where g is a suitable link function (e.g., the complementary log-log) and the β s are regression coefficients. This expression yields two main advantages. First, the constant hazard model is included as a particular case (when $\beta_{1k} = 0$), facilitating tests against the null hypothesis of a geometric dwell time. These tests are more complicated under the traditional approach, because popular sojourn time distributions do not typically include the geometric distribution as a particular case. Second, it can be extended to include a profile of time-varying covariates $\mathbf{z}_t$, not necessarily equal to the profile $\mathbf{x}_t$ that has been exploited at the observation level, namely,

$$(7) \quad g(q_k(d_{t-1}, \mathbf{z}_t)) = \beta_{k0} + \beta_{k1}d_{t-1} + \mathbf{z}_t^\top \boldsymbol{\beta}_{k2}, \quad k = 1, \dots, K.$$

Under (7), the hazard is endogenously updated by the current dwell time and, exogenously, by time-varying covariates. By plugging this dynamic hazard regression into equation (6), we obtain a sojourn time distribution that can be modulated by covariates at every time, and not just at baseline.

The joint distribution of $\mathbf{u}$ is thus fully specified by a parametric model $p(\mathbf{u}|\boldsymbol{\eta})$ that depends on the parameter vector $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\omega})$, where the regression coefficients $\boldsymbol{\beta}$ of the hazard regressions (7) drive the sojourn distribution of each state, whereas the conditional switching probabilities $\boldsymbol{\omega}$ steer the chain to a specific state at each transition.

4. Likelihood-based inference. For a given K , the joint distribution of the proposed hierarchical model is known up to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k : k = 1, \dots, K\}$ are the observation process parameters and $\boldsymbol{\eta} = \{\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\Omega}\}$ are the hidden process ones. Parameter estimation given the observed data for a fixed number of hidden states K and maximum autoregressive lag H can be pursued by optimizing the marginal likelihood corresponding to equation (1),

$$L(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{\mathbf{u}} \prod_{t=1}^T \prod_{k=1}^K f_k(\mathbf{y}_t | \mathbf{y}_{(t-H):(t-1)}; \boldsymbol{\theta}_k)^{u_{tk}} p(\mathbf{u}_1; \boldsymbol{\pi}) \prod_{t=2}^T \prod_{k=1}^K \prod_{j=1}^K \gamma_{kj}(d_{t-1}; \omega_{kj}, \boldsymbol{\beta}_k)^{u_{tkj}}.$$

However, its direct maximization is hindered by the overwhelming summation over all the possible paths of the latent semi-Markov chain. This difficulty can be overcome through the approximation of the hidden semi-Markov chain as a hidden-Markov one with $K \times m$ number of states, where m is the maximum dwell-time for which the parametric assumption of equation (7) holds. Dwell times larger than m are still allowed, but their distribution is approximated by a geometric tail (Zucchini, MacDonald and Langrock (2016), Lagona and Mingione (2025)). The HMM formulation gives way to the implementation of a workable EM algorithm where maximization of a weighted complete-data log-likelihood (M step) is iteratively alternated with weights updating (E step), until convergence. Convergence is assessed through log-likelihood- and/or parameter-based stopping criteria.

Specifically, at each iteration ℓ , the E-step evaluates the univariate posterior probabilities $\hat{\pi}_{tkd}^{(\ell)} = P(u_{tkd} = 1 | \mathbf{y}, \boldsymbol{\theta}^{(\ell-1)}, \boldsymbol{\eta}^{(\ell-1)})$ and the bivariate posterior probabilities $\hat{\pi}_{tkdj}^{(\ell)} = P(u_{(t-1)kd} = 1, u_{tjd} = 1 | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(\ell-1)}, \hat{\boldsymbol{\eta}}^{(\ell-1)})$ through the *forward-backward* recursions of the Baum–Welch algorithm (see Cappé, Moulines and Rydén (2005) for an excellent review).

Such weights are exploited by the subsequent M-step that updates the parameters by maximizing the weighted log-likelihood function,

$$\begin{aligned}
 Q(\boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\omega}, \boldsymbol{\beta}) = & \underbrace{\sum_{k=1}^K \hat{\pi}_{1k}^{(\ell)} \log \pi_k}_{Q(\boldsymbol{\pi})} + \underbrace{\sum_{t=2}^T \sum_{k=1}^K \sum_{j=1}^K \sum_{d,d'=1}^m \hat{\pi}_{tkdj d'}^{(\ell)} \log \gamma_{kj}(d, \mathbf{z}_t; \boldsymbol{\Omega}, \boldsymbol{\beta})}_{Q(\boldsymbol{\Omega}, \boldsymbol{\beta})} \\
 & + \underbrace{\sum_{t=1}^T \sum_{k=1}^K \hat{\pi}_{tk}^{(\ell)} \log f_k(\mathbf{y}_t | \mathbf{y}_{(t-H):(t-1)}; \boldsymbol{\theta}_k)}_{Q(\boldsymbol{\theta})},
 \end{aligned}$$

where $Q(\boldsymbol{\Omega}, \boldsymbol{\beta}) = Q(\boldsymbol{\Omega}) + Q(\boldsymbol{\beta})$ with $Q(\boldsymbol{\Omega}) = \sum_{t=2}^T \sum_{k=1}^K \sum_{j \neq k} \sum_{d,d'=1}^m \hat{\pi}_{tkdj d'} \log \omega_{kj}$ and

$$\begin{aligned}
 Q(\boldsymbol{\beta}) = & \sum_{t=2}^T \sum_{k=1}^K \sum_{h \neq k} \sum_{d,d'=1}^m \hat{\pi}_{tkdj d'} \log q_k(d, \mathbf{z}_t; \boldsymbol{\beta}) \\
 & + \sum_{t=2}^T \sum_{k=1}^K \sum_{d,d'=1}^m \hat{\pi}_{tkdk d'} \log(1 - q_k(d, \mathbf{z}_t; \boldsymbol{\beta})).
 \end{aligned}$$

The weighted log-likelihood function is the sum of functions that depend on independent sets of parameters, which can be maximized separately. Functions $Q(\boldsymbol{\pi})$ and $Q(\boldsymbol{\Omega})$ are weighted multinomial log-likelihoods for which the points of maximum are available in closed form as

$$\hat{\pi}_k^{(\ell)} = \hat{\pi}_{1k}^{(\ell)}, \hat{\omega}_{kj}^{(\ell)} = \frac{\sum_{t=2}^T \sum_{j \neq k} \sum_{d,d'=1}^m \hat{\pi}_{tkdj d'}^{(\ell)}}{\sum_{t=2}^T \sum_{k=1}^K \sum_{j \neq k} \sum_{d,d'=1}^m \hat{\pi}_{tkdj d'}^{(\ell)}}.$$

$Q(\boldsymbol{\beta})$ is a Binomial log-likelihood that can be maximized by conventional iteratively reweighted least-squares routines, leading to $\hat{\boldsymbol{\beta}}_k^{(\ell)}$.

Function $Q(\boldsymbol{\theta})$ is the weighted log-likelihood of a VAR model and can be maximized via the SUR (seemingly unrelated regression) method. For each state k and outcome j , the optimal estimate of the VAR parameters $\{\hat{\mathbf{a}}_{kj}, \hat{\mathbf{b}}_{kj0}, \hat{\mathbf{b}}_{kj}\}^{(\ell)}$ are the minimizers of the weighted sum of squares $\mathbf{e}_{kj}^\top \mathbf{W}_k^{(\ell)} \mathbf{e}_{kj}$, where $\mathbf{e}_{kj} = [e_{1kj}, \dots, e_{Tkj}]$ is the residual vector of model (2),

$$(8) \quad e_{tkj} = \left(y_{tj} - b_{kj0} - \mathbf{x}^\top \mathbf{b}_{kj} - \sum_{h=1}^H \sum_{i=1}^p a_{kji h} y_{(t-h)i} \right)$$

and $\mathbf{W}_k^{(\ell)}$ is the $T \times T$ diagonal matrix of the weights $\hat{\pi}_{tk}^{(\ell)} = \sum_d \hat{\pi}_{tkd}^{(\ell)}$. The corresponding estimate of $\hat{\Sigma}_k^{(\ell)}$ can be obtained by plugging these estimates into the residuals (8) and computing each entry as

$$\hat{\sigma}_{kij} = \frac{\sum_{t=1}^T \hat{\pi}_{tk}^{(\ell)} \cdot e_{tki} e_{tkj}}{\sum_{t=1}^T \hat{\pi}_{tk}^{(\ell)}}.$$

However, such SUR approach involves a large number of parameters and requires the selection of the lags to be included in the VAR specification. These are generally unknown and would require post hoc model selection scheme. Therefore, to enhance interpretability and avoid the arbitrary choice of the lags, we regularize the least squares by a LASSO penalty

(Basu and Michailidis (2015), Tan, Chiong and Moon (2021)), updating the VAR parameters as the solution of the penalized least squares problem,

$$(9) \quad \min \{ \mathbf{e}_{kj}^\top \mathbf{W}_k^{(\ell)} \mathbf{e}_{kj} + \lambda_k \cdot \|\mathbf{a}_{kj} + \mathbf{b}_{kj}\|_1 \},$$

using a battery of penalty coefficients λ_k that operate on the state-specific regression coefficients of the VAR. For a sufficiently large value of the maximum lag H , this LASSO regularization enforces sparsity and ensures that the irrelevant lagged variables and covariates are automatically excluded from the model, leading to a more parsimonious representation. Such LASSO selection procedure also helps mitigating the underestimation of exogeneous covariates' effects, if they are included in the model. Indeed, LASSO would favor lagged values of the main variables over other exogeneous variables if and only if they are deemed to be a better predictor of future outcomes.

The proposed estimation procedure leads to an optimal solution for a specific number K of hidden states K and a specific value of the penalty coefficients λ_k , $k = 1, \dots, K$. Their choice is nontrivial, especially under a mixture setting, where a post hoc selection over a grid of possible combinations of the two is compared through goodness-of-fit or prediction metric. In our case we consider and test through a simulation study the use of the Integrated Complete Likelihood (ICL, see Biernacki, Celeux and Govaert (2000)) to perform model selection in terms of both these terms. However, we must note that in a mixture context the size of the grid of λ_k s on which the validation process should be performed increases quadratically with the number of states K , becoming rapidly overwhelming, even for moderate K . This issue is particularly pronounced in an HSMM context, where even a single fit might be computationally demanding. One could consider adopting one single value $\lambda_k = \lambda$, but it is well known that the optimal penalty value in penalized regression usually depends on sample size (Bühlmann and van de Geer (2011)). Therefore, the penalty parameter magnitude should at least account for the effective number of units belonging to each state. In the context of finite mixture of regression, Khalili and Chen (2007) propose penalties that depend on the size of the regression coefficients and the mixture structure. Similarly, Städler and Mukherjee (2013) propose a penalty that automatically and dynamically adapts to the current state-specific effective sizes within the EM optimization routine. In the latter the dynamically varying penalty term is specified as

$$\lambda_k = \lambda_0 \cdot \sqrt{\sum_{t=1}^T \hat{\pi}_{tk}}, \quad k = 1, \dots, K.$$

λ_0 is a baseline penalty term susceptible to state-specific penalty adjustments which are proportional to the square root of the *effective sample size* of each state, conforming to the asymptotic theory on the optimal penalty developed in Städler, Bühlmann and van de Geer (2010).

The proposed EM algorithm is stopped when the relative variation in all the parameters between two consecutive iterations is lower than 10^{-4} . This yields the final parameters' estimates $(\hat{\theta}, \hat{\eta})$. In such a complex context, the derivation of uncertainty of the parameters through the inversion of the Hessian is not practical. Therefore, we quantify uncertainty through the implementation of a parametric bootstrap (Efron (2000), Visser, Raijmakers and Molenaar (2000)). This is further complicated by the LASSO penalty, for which the bootstrap would return biased confidence intervals of the true parameters (Chatterjee and Lahiri (2011), Li (2020)). Therefore, we first de-bias the final estimates by mean of a *relaxed* LASSO fit (i.e., one more round of the M-step in EM with only the selected coefficient and no penalty), obtaining $\hat{\theta}^R$. Then we simulate B datasets from the joint model with parameters $(\hat{\theta}^R, \hat{\eta})$ and re-fit the whole procedure, model selection included, on each of them. This yields $(\hat{\theta}^b, \hat{\eta}^b)_{b=1}^B$ that can be used to quantify the overall uncertainty of $(\hat{\theta}, \hat{\eta})$.

5. Simulation study. We devise a simulation study to assess and verify the validity of our proposal in three major aspects: (i) evaluate whether the ICL can effectively discriminate the best pair (K, λ_0) ; (ii) recovery of the model parameters and selection of the nonzero coefficients in the observation process; (iii) time-series segmentation.

We simulate $B = 200$ replicate datasets with $n = 1100$ observations to match the size of the real data under analysis. The outcome dimension is set to $p = 3$ and the maximum lag order equal to $H = 4$, for a varying number of mixture components $K = 2, 3, 4$. The autoregressive coefficients have been chosen such that each state corresponds to a stable VAR and the effect at lags going from 1 to 3 is decreasing within each outcome and state. The *fourth* lags do not necessarily respect this decreasing behavior, as they could represent seasonality patterns. The correlation matrices have been randomly generated from an inverse Wishart distribution centered around identity matrices of conforming size and $\nu = 3$ degrees of freedom. The dwell time parameters have been set to values that ensure different sojourn distributions. Furthermore, we include the effect of one randomly generated (mean-zero Gaussian) time-varying covariate on the observation and the hidden process. This setup simulates realistic scenarios with partially overlapped mixture components exhibiting a wide range of correlations and different behaviors of the latent components. In all scenarios we fit the models for varying number of hidden states $\tilde{K} = 2, 3, 4$ and a sequence of values starting from 0 and followed by 20 equally spaced values of $\log(\lambda_0)$ between 10^{-4} and 0.05.

We first verify whether the $\tilde{K}^*$ identified as optimal by the ICL matches the true size across all scenarios. Table 1 reports the average ICL for each combination of simulated and estimated model. Results show that the average ICL is lowest for the model with the *true* number of components, highlighting the ability of the ICL to select the correct model specification also under a penalized regression framework. In addition, comparing the ICL of each single replica, we report that the correct model (i.e., $\tilde{K}^* = K$) is selected 100% of the time across all K s. Therefore, we can focus on the performances of the well-specified models only (i.e., $K = \tilde{K}^*$), and in particular, we report in the main text the results for the $K = 2$ scenario for reason of space. The results for $K = 3, 4$ are briefly commented here but detailed in the Supplementary Material.

To evaluate the recovery of the regression and autoregressive coefficients of the observation process, other than the optimal selection of λ_0^* , we compare the bootstrapped distribution of the estimates with their true values. Table 2 reports some relevant statistics on the regression coefficients, while the performances on the autoregressive coefficients are summarized in Figure 3. The coefficients' selection performs well across all outcomes and states, with a few exceptions that are characterized by selection percentages close to 50%. In particular, the estimates of the coefficients that are wrongly selected as nonzero always present near-zero point estimates and zero-overlapping intervals. The distribution of the nonzero coefficients is slightly biased toward lower values, which is a common issue in penalized regression, but the true value is always contained in the central 95% of the estimation interval. The recovery of the variance-covariance matrices is evaluated through the matrix Kullback–Leibler (KL)

TABLE 1
Average ICL scores for model selection in the simulation study

		Estimated		
	K	2	3	4
True	2	10,854.56	11,084.04	11,342.60
	3	11,974.74	11,432.90	11,675.66
	4	11,475.18	11,314.95	10,858.25

TABLE 2
Recovery of the regression coefficient for $K = \tilde{K}^* = 2$: True value, Mean of the estimates, 95% Estimation Interval, Root Mean Squared Error (RMSE), proportion of selection

B_k	$k = 1$					$k = 2$				
	True	Mean	EI _{0.95}	RMSE	%	True	Mean	EI _{0.95}	RMSE	%
b_{k10}	3	3.11	(2.97, 3.26)	0.14	—	−1	−0.88	(−1.02, −0.75)	0.14	—
b_{k20}	1.5	1.38	(1.23, 1.57)	0.15	—	−2	−2.14	(−2.29, −2.00)	0.15	—
b_{k30}	2	2.01	(1.85, 2.14)	0.08	—	−1.5	−1.49	(−1.58, −1.38)	0.06	—
b_{k11}	0.5	0.457	(0.38, 0.55)	0.06	100	−0.2	−0.158	(−0.25, −0.08)	0.06	100
b_{k21}	0.0	0.00	(−0.04, 0.05)	0.02	35	0.4	0.36	(0.29, 0.44)	0.06	100
b_{k31}	0.0	0.00	(−0.06, 0.05)	0.02	41	−0.1	−0.06	(−0.13, 0.000)	0.06	87

discrepancy, whose average values are very low and equal to 0.013 and 0.010 for $K = 1, 2$, respectively. Similar performances are observed for the other K s (see the Supplementary Material). Table 3 shows the estimates of the dwell time coefficients. The point estimates yield averages that closely align with the true values, which consistently fall within the 95% estimation intervals. The RMSEs remain low across all cases, demonstrating that our method effectively captures the dynamics of the latent terms. The conditional transition probability matrix Ω is not estimated for $K = \tilde{K}^* = 2$, as it is deterministically defined as a row-reversed identity matrix J_K .

Regarding time-series segmentation, each observation is assigned a cluster label with the Maximum A Posteriori (MAP) rule. The average ARI and Accuracy are notably high (exceeding 87% and 95%, respectively) in all scenarios. This underscores the model’s ability to accurately assign each data point to its corresponding cluster, regardless of model complexity. Detailed results are included in the Supplementary Material.

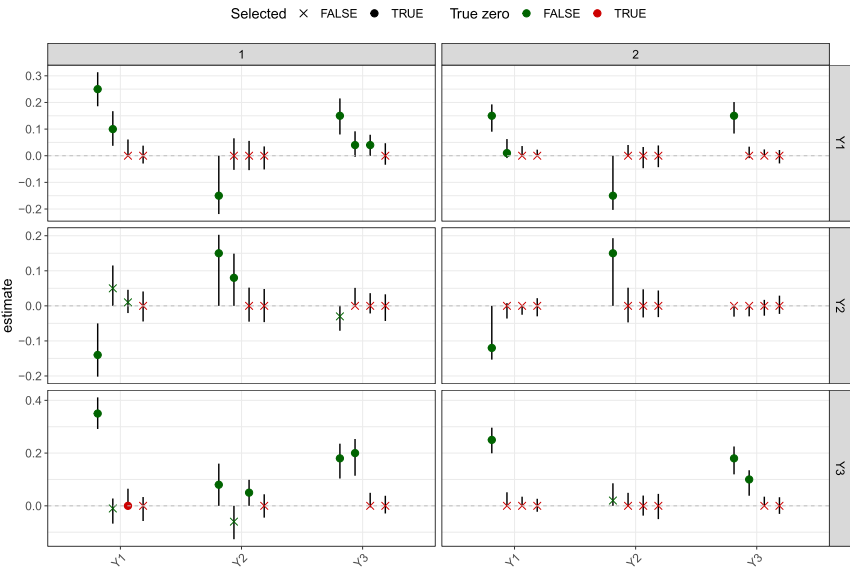


FIG. 3. Recovery of the autoregressive coefficients for $K = 2$. True values are: green if different from 0, red if not; marked as a cross if shrunk to 0 more than 50% of the times or as a dot otherwise. The black line represents the central 95% of the estimates distribution.

TABLE 3
Recovery of the dwell-time regression coefficients $K = \tilde{K}^* = 2$: True value, Mean of the estimates, 95% Estimation Interval, Root Mean Squared Error (RMSE)

η	$k = 1$				$k = 2$			
	True	Mean	El _{0.95}	RMSE	True	Mean	El _{0.95}	RMSE
β_{k0}	-1	-1.03	(-1.35, -0.76)	0.16	-2	-2.03	(-2.38, -1.71)	0.18
β_{k1}	0.15	0.16	(0.07, 0.25)	0.05	0.35	0.36	(0.28, 0.46)	0.05
β_{k2}	-0.50	-0.51	(-0.67, -0.38)	0.07	0.50	0.50	(0.36, 0.66)	0.07

6. Application. The model is applied to the data described in Section 2 for different values of $K = 2, 3, 4$ and $\lambda_0 = 0, e^{10^{-4}}, \dots, e^{0.05}$, where the exponents from 10^{-4} and 0.05 are equally spaced. The order of the VAR is set to $H = 7$ across all pollutants and states so that all reasonable lags up to those responsible for weekly seasonality could be selected. Data with different time resolution might require a different choice of H . For example, for hourly data one could set $H = 24$ to account for possible intradaily cycles. We use the average daily temperature and the weekend effect (namely, weekend = 1 if the day of the week is Saturday or Sunday, weekend = 0 otherwise) as exogenous time-varying covariates for the mean term of the observations process. We instead consider the average wind speed and total precipitation as time-varying covariates on the hidden process hazards. These are well-known environmental conditions that affect the chances of reducing pollution levels by either *diluting* their concentrations into larger areas or washing them out, that is, dragging them to the ground. As a link function for the hazard g , we use the c-loglog, which is the canonical link resulting from the specification of a Gompertz-type hazard in discrete time. The tail of the dwell time distribution is approximated by a geometric tail for $d > m$, setting $m = 28$ (see the Supplementary Material for additional details on this choice). Uncertainty is quantified through bootstrapping over $B = 300$ simulated sets. The optimal penalty λ^* and number of states K is selected via cross-validation according on the ICL. The results are summarized in Figure 4. The behavior for each K is convex, as expected, and the best score is achieved for the $K = 2$, at the 11th value of λ . Furthermore, the adherence of the data to the selected

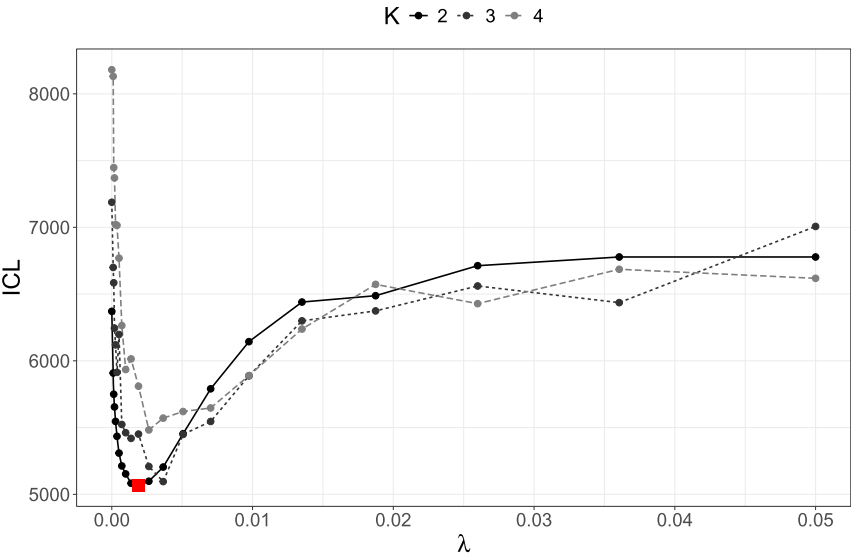
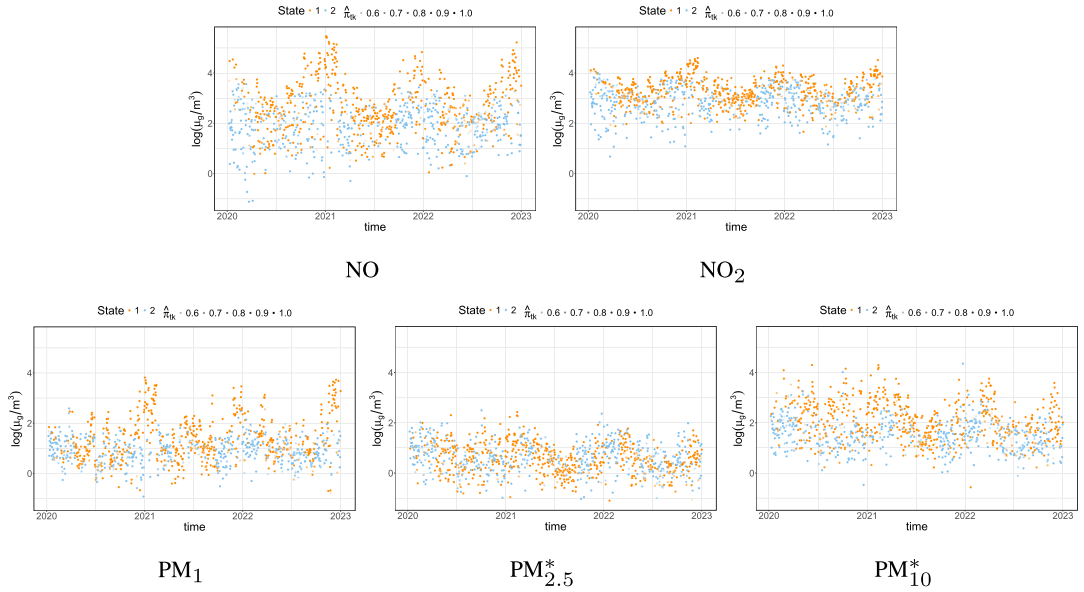


FIG. 4. Model selection: Cross-validation of λ for each K .

FIG. 5. *Estimated time-series segmentation.*

model is successfully confirmed through the inspection of forecast pseudo-residuals (Zucchini, MacDonald and Langrock (2016)), with further details provided in the Supplementary Material. Therefore, all the following results will refer to this model’s specification.

Figure 5 shows the estimated time-series segmentation over the entire study period. The estimated VAR process is stable in both regimes, with marginal means (net of the covariates effects) equal to $\mu_1 = [3.76, 3.70, 1.80, 1.00, 2.76]$ and $\mu_2 = [2.34, 3.24, 0.84, 0.67, 1.77]$, respectively. Thus, we can interpret state 1 (orange) as a “high pollution” regime and state 2 (sky blue) as a “low pollution” one. The pollutant series remains in state 1 for approximately 55% of the time, predominantly during spring and summer. State 2 occurs with a similar frequency but is mainly observed in autumn and winter. This seasonal asymmetry can be linked to the intensified use of heating systems, which elevate emissions and reduce pollutant dispersion, and to the greater reliance on motorized transport under adverse weather conditions that discourage active mobility. Table 4 reports the estimated regression coefficients on the observation process. The temperature effect matches the summer-winter seasonality highlighted above. We can also notice how, regardless of the temperature, weekends exhibit consistently lower average pollution levels—except for PM_1 —supporting the presence of different traffic patterns between weekdays and weekends that contribute to a reduction in pollution emissions. The estimates of the autoregressive coefficients are summarized in Figure 6, using a similar scheme to that utilized in the simulation study. The LASSO penalty shrinks to 0 most of the autoregressive coefficients, and it does so consistently across the bootstrap samples. This means that, as expected, the maximum lag $H = 7$ was too large and gives rise to very sparse relationships. However, it proves to be very useful as it allows detecting some significant seasonality patterns occurring within or across pollutants at lag 6 and 7.

The estimated correlation patterns within each state are reported in Figure 7a. In State 1, pollutants are generally more correlated one another. More specifically, the pairs $PM_{2.5}$ – PM_{10} and NO–NO₂ form strongly correlated pairs, each exhibiting stronger internal associations than with any other pollutants. PM_1 , although still correlated, seems to behave more independently. In State 2, corresponding to lower pollution levels, two distinct clusters of highly correlated variables emerge, which are only weakly correlated with each other: nitrogen oxides and particulate matters. This shift reflects changes in dominant emission

TABLE 4
Estimated regression coefficients on the observations process: Estimates, 95% bootstrap Confidence Intervals, percentage of selection

	<i>k</i>	NO		NO ₂		PM ₁		PM _{2.5} [*]		PM ₁₀ [*]	
		Est (CI)	%	Est (CI)	%	Est (CI)	%	Est (CI)	%	Est (CI)	%
Intercept	1	2.63 (2.31, 3.40)	–	2.47 (2.21, 2.75)	–	0.73 (0.42, 1.02)	–	0.49 (0.18, 0.59)	–	1.36 (1.06, 1.72)	–
	2	2.16 (1.70, 2.49)	–	2.93 (2.58, 3.21)	–	0.59 (0.32, 0.76)	–	0.42 (0.01, 0.76)	–	1.51 (1.08, 1.81)	–
Temperature	1	–0.06 (–0.07, –0.05)	100	–0.02 (–0.03, –0.02)	100	–0.02 (–0.03, –0.01)	100	–0.01 (–0.02, –0.01)	100	–0.01 (–0.02, –0.01)	100
	2	–0.01 (–0.03, –0.01)	100	–0.02 (–0.04, –0.02)	100	0 (0.00, 0.00)	13	0 (–0.01, 0.00)	49	–0.02 (–0.03, –0.01)	100
Weekend	1	–0.83 (–1.03, –0.73)	100	–0.28 (–0.39, –0.23)	100	0 (–0.04, 0.00)	15	–0.42 (–0.53, –0.35)	100	–0.70 (–0.85, –0.63)	100
	2	–1.05 (–1.22, –0.91)	100	–0.66 (–0.77, –0.57)	100	0 (–0.02, 0.03)	11	–0.02 (–0.23, 0.00)	89	–0.17 (–0.35, –0.08)	100



FIG. 6. Estimated autoregressive coefficients. Estimates are denoted with a green dot if selected for more than 50% of the bootstrap samples or a red cross otherwise. The black line represents the bootstrap 95% Confidence Intervals conditional on the parameter being selected.

patterns and atmospheric dynamics. At higher pollution levels, pollutant sources tend to be common and localized; at lower pollution levels, differences in source types and atmospheric behavior become more pronounced. Nitrogens primarily originate from combustion processes, such as traffic emissions, whereas PM includes both primary particles and secondary aerosols formed from various precursors. Moreover, PM levels are influenced by diffuse and less-regulated sources like residential heating, agriculture and regional transport, while nitro-

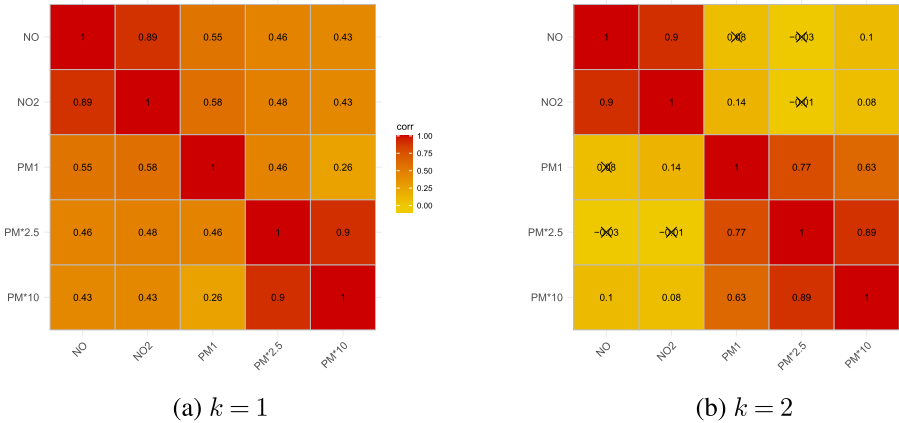


FIG. 7. Estimated correlation matrix for each state.

TABLE 5
Estimated parameters of the dwell time: Estimates and 95% bootstrap Confidence Intervals

k	β_{k0}	β_{k1}	Precipitation	Wind speed
1	-3.35 (-4.41, -2.92)	-0.12 (-0.18, -0.04)	6.65 (5.50, 8.64)	0.49 (0.36, 0.68)
2	1.61 (1.00, 2.29)	-0.02 (-0.07, 0.12)	-3.13 (-4.47, -2.14)	-0.46 (-0.71, -0.32)

gens levels are short-lived and more spatially confined. As a result, the correlation between nitrogens and PMs weakens under cleaner atmospheric conditions.

Table 5 reports the estimated dwell-time regression coefficients and corresponding 95% bootstrapped confidence intervals. State 2 follows a geometric sojourn distribution, as indicated by the nonsignificant β_{11} , and is characterized by relatively short durations (high β_{10} and longest sojourn of at most five days). Transitions from this low-pollution state to the high-pollution one become less likely with increasing total precipitation and average wind speed, suggesting that rainy and windy conditions prolong cleaner atmospheric episodes. Conversely, in the absence of rain and wind, the probability of switching from high to low pollution decreases over time ($\beta_{21} < 0$), leading to longer persistence in the polluted state (the plot of the estimated distributions is reported in the Supplementary Material). This indicates that the polluted state can behave as *quasi-absorbing*, left only when specific meteorological conditions occur. This behavior is consistent with the phenomenon of accumulation in which stagnant atmospheric conditions inhibit pollutant dispersion, leading to a progressive build-up of emissions from traffic, heating, and industrial activities. Such episodes can persist until wind or precipitation restores vertical and horizontal mixing.

6.1. *Risk measures with a focus on multivariate risk assessment.* Risk measures have historically been developed in the econometric literature to quantify the systemic risk of financial institutions. In such contexts, risk is typically represented by the realization of particularly low returns. When applied to environmental risk and air pollution, this perspective must be reversed, with risk corresponding to the realization of particularly high concentrations of air pollutants. Let $\mathcal{S} = \{\text{NO}, \dots, \text{PM}_{10}^*\}$ denote the set of $p = 5$ pollutants under analysis. For each pollutant $i \in \mathcal{S}$, we can thereby define the Value-at-Risk $VaR_i(\tau)$ as the minimum log-concentration level that belongs to the $\tau \times 100$ “worst” realizations of pollutant i . In other words, $VaR_i(\tau)$ represents the $(1 - \tau)$ -quantile of the pollutant’s log-concentration distribution, identifying the threshold beyond which exceedances correspond to extreme pollution events. The associated Expected Shortfall $ES_i(\tau)$ is then defined as the conditional expectation of the pollutant’s log-concentration given that it exceeds the corresponding $VaR_i(\tau)$. Thus, $ES_i(\tau)$ quantifies the average severity of pollution levels in the upper τ -tail of the distribution. In practice, the value of τ is typically chosen to be small (e.g., $\tau = 0.05$), reflecting interest in the most extreme pollution episodes. In the multivariate framework considered here, we consider the extension of these risk-measures to their multivariate counterparts, namely, $MCoVaR$ and $MCoES$ (Adrian and Brunnermeier (2016)). The core idea remains unchanged, but instead of considering the univariate marginal distribution of each pollutant, they are defined on the conditional distribution of pollutant i given the others $\mathcal{S} \setminus \{i\}$. The remaining pollutants are partitioned into two subsets:

- $\mathcal{H}_d \subset \mathcal{S} \setminus \{i\}$, representing pollutants in an *in-distress* situation;
- $\mathcal{H}_n = \mathcal{S} \setminus \{\mathcal{H}_d, i\}$, representing pollutants in a *non-in-distress* situation.

The $MCoVaR$ and $MCoES$ of pollutant i are defined conditionally on this partition, capturing the tail risk of the i th pollutant given that the others are simultaneously experiencing extreme (in-distress) or normal (nondistress) conditions. A detailed formal definition of these measures is provided in the Supplementary Material.

Our framework introduces an additional layer of complexity by incorporating temporal dynamics and a latent process. Consequently, the joint distribution of pollutants is not time-invariant and depends on the current latent state, inducing time-varying risk measures. Accordingly, we consider dynamic $MCoVaR_t$ and $MCoES_t$, which capture the evolution of joint tail risk in a time-adaptive manner (Bernardi, Maruotti and Petrella (2017)). These dynamic measures have been originally defined in terms of the joint distribution of the multivariate outcome at each time t . In our setting, however, we seek risk measures that are marginal with respect to the latent process (and thus detached from specific latent states) while remaining conditional on the history of the process. Hence, we derive the following expression for the one-step-ahead predictive distribution:

$$(10) \quad f(\mathbf{y}_t | \mathbf{y}_{1:(t-1)}; \boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{k=1}^K \underbrace{\left(\sum_{j=1}^K \sum_{d=1}^m \bar{\alpha}_{t-1}(j, d) \cdot \gamma_{jk}(d) \right)}_{\psi_{tk}} \cdot f_k(\mathbf{y}_t | \mathbf{y}_{(t-H):(t-1)}; \boldsymbol{\theta}_k).$$

Here $\bar{\alpha}_{t-1}(j, d)$ denotes the normalized forward probability of being in state j with sojourn d at time $t - 1$, and $\gamma_{jk}(d)$ is the transition probability from state j to state k . The weights ψ_{tk} are the one-step-ahead filtering probabilities (Cappé, Moulines and Rydén (2005)), which sum to 1 and act as mixture weights in a K -component Gaussian mixture at each time t (equation (2)). In particular, the corresponding conditional distribution of each element, given all others, is also a mixture of univariate Gaussian distributions with well-known updated mean and variance terms. The required conditional quantiles can be obtained via numerical methods, while the tail-expectations present closed-form solutions as weighted combinations of those of each mixture components.

The multivariate risk measures for a given pollutant i depend on how the other pollutants are partitioned into $\mathcal{H}_d \cup \mathcal{H}_n = \mathcal{S} \setminus i$. This results in a combinatorially large number of possible configurations, and thus an equally large number of potential values for the risk measure of pollutant i at each time t . To obtain a compact and interpretable summary, and to assess the overall contribution of each pollutant to the risk of others, we adopt the Shapley Value methodology. This approach decomposes the total risk across contributing factors by computing their marginal contributions across all possible distress and nondistress configurations. In practice, the contribution of pollutant j to the risk of pollutant i is quantified by averaging its incremental effect across all combinations of the remaining pollutants being in distress or not. This yields the Shapley Value $SH_{ti}(j)$ (see the Supplementary Material for details). Originally developed in game theory to fairly distribute gains, the Shapley Value requires comparability of payoffs across contributors. However, such comparability is not consistent within the pollutants' context at hand, as each pollutant has its own intrinsic variability and, most especially, might be recorded in a different unit of measure. Therefore, we propose a modification of the original approach to ensure comparability of $SH_{ti}(\cdot)$ across pollutants, so variations in the risk measure are unit-free and can be interpreted in relative terms. Let $\eta_{ti}(\mathcal{H}_d)$ denote a multivariate risk measure (e.g., $MCoVaR$ or $MCoES$) for pollutant i at time t , conditional on the subset $\mathcal{H}_d \subset \mathcal{S} \setminus \{i\}$ being in distress, while $\mathcal{S} \setminus \mathcal{H}_d$ is not. Our proposed standardized Shapley Value is

$$\widetilde{Sh}_{ti}(j) = \frac{1}{\sigma_{ti}} \cdot \sum_{\mathcal{H} \subset \mathcal{S} \setminus \{i, j\}} v(\mathcal{H}) \cdot (\eta_i(\mathcal{H} \cup \{j\}) - \eta_i(\mathcal{H})),$$

where σ_{ti} is the marginal standard deviation of pollutant i at time t and $v(\mathcal{H})$ denotes the Shapley weighting function. This standardized version ensures comparability of risk contributions across pollutants and over time, providing a coherent measure of each pollutant's relative impact on the system-wide pollution risk.

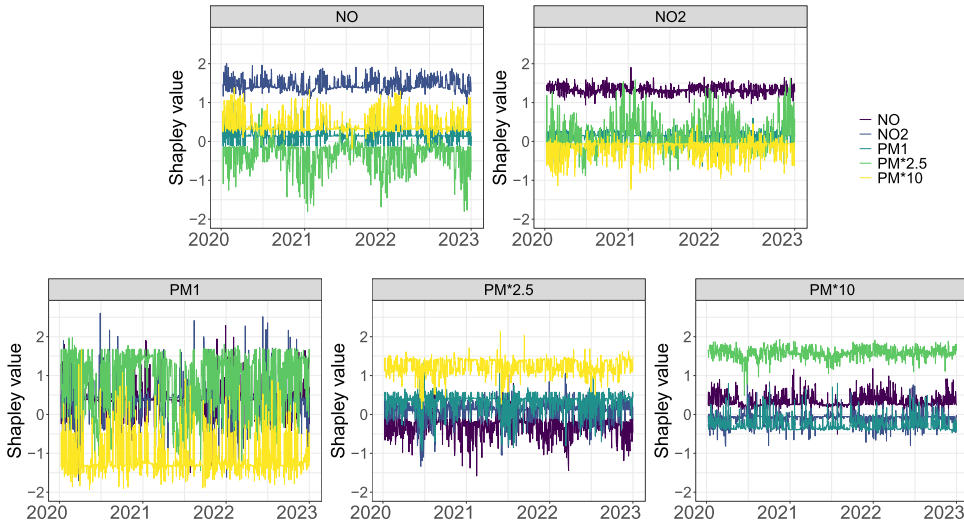


FIG. 8. Time series of the Shapley Values of each pollutant on each other for the MCoVaR.

We use this method to evaluate the time-varying risk contributions between each pair of pollutants in our empirical application. In particular, we set $\tau = 0.05$ and plug-in the final model estimates $(\hat{\theta}, \hat{\eta})$ into equation (10) to obtain the estimated one-step-ahead predictive distribution $\hat{f}_t(\cdot) = f(\cdot | \mathbf{y}_{1:(t-1)}; \hat{\theta}, \hat{\eta})$. Figure 8 reports the results for the MCoVaR, offering a nuanced view of how each pollutant contributes to the risk of others over time. We observe distinct patterns of interaction and risk attribution. NO and NO₂ display consistently high and mutually reinforcing Shapley Values. This pattern reflects the strong chemical coupling between these two gases—NO₂ is primarily formed through the oxidation of NO—and their common emission sources, particularly road traffic. The stable and prominent risk contributions of NO and NO₂ indicate that these two pollutants are not only co-occurring but also jointly drive the dynamics of air quality deterioration, particularly in urban and suburban settings. Moreover, the symmetry in their risk profiles highlights the bidirectional nature of their influence. This mutual influence is especially pronounced during peak pollution episodes, suggesting that risk mitigation strategies targeting one of the two may produce cascading benefits across the system. The temporal evolution of the Shapley Values corresponding to NO and NO₂ highlights periods of heightened systemic risk, where their dependency intensifies. In particular, we observe stronger interconnections during winter months, reflecting how meteorological conditions modulate joint pollution extremes.

In contrast, PM₁, PM_{2.5}^{*}, and PM₁₀^{*} exhibit more heterogeneous patterns, characterized by episodic peaks which could be associated with stagnation or transboundary pollution events. Overall, PM_{2.5}^{*} and PM₁₀^{*} have high mutual Shapley Values, indicating their joint behavior and potential common sources, such as regional particulate matter transport and resuspension. These relationships suggest that the monitoring of one pollutant in each pair may suffice to track the overall risk, reducing the complexity of surveillance systems without substantial loss in informational content. Conversely, PM₁ exhibits a more independent profile, with lower and more variable Shapley Values across both risk measures. In some high-pollution periods, PM₁ even shows negative contributions to the overall risk of other particulates, suggesting competitive dynamics in particulate composition. By averaging the dynamic Shapley Values over time, Figure 9 presents the resulting matrix of average inter-pollutant Shapley Values, which can be interpreted as a network of risk propagation among pollutants. The structure of this matrix reinforces the patterns observed in the dynamic analysis. Specifically,

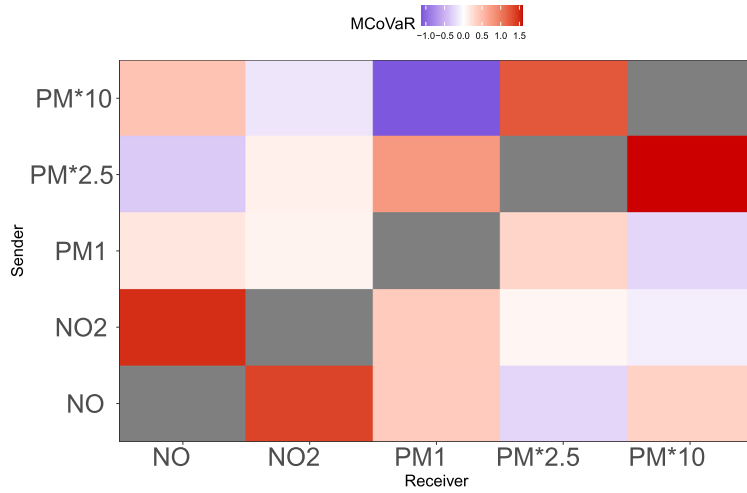


FIG. 9. Matrix of the average MCoVaR Shapley Value between each pair of pollutants.

the NO-NO₂ and PM_{2.5}^{*}-PM₁₀^{*} pairs emerge as strongly interconnected risk couples, indicating that these pollutants tend to jointly drive episodes of elevated pollution risk. Within the particulate pair, a hierarchy is observed, with PM_{2.5}^{*} acting as the primary risk contributor. On the contrary, PM₁ behaves predominantly as a risk receiver, with its tail behavior responding to fluctuations in the concentrations of other gaseous precursors (Geng et al. (2017)). The matrix also reveals some systemic negative risk interactions, specifically between PM_{2.5}^{*} and NO, between PM₁₀^{*} and NO₂ and between PM₁₀^{*} and PM₁. The inverse relationships between NO_x and large particulate matter particles are consistent with findings reported by Thunis et al. (2021), which attributed them to the enhanced oxidative capacity of the atmosphere under conditions of reduced NO_x concentrations. The inverse pattern between PM₁₀^{*} and PM₁ can instead be explained by the differing physical and chemical processes governing these particle size ranges. PM₁₀^{*} is primarily generated through mechanical resuspension and crustal material dispersion, typically under dry and windy conditions that enhance atmospheric mixing. Such conditions tend to disperse fine aerosols, leading to lower PM₁ concentrations during coarse-particle episodes (Zhai et al. (2023)).

Similar interpretations are obtained for the MCoES values, whose results are reported in the Supplementary Material. The proposed dynamic Shapley-based decomposition and interpretation provide a clear quantitative framework for prioritizing interventions. Targeting pollutants with generally positive and high outgoing Shapley influence (e.g., systemic senders) could generate system-wide reductions in multivariate pollution risk, offering an efficient basis for coordinated emission control strategies.

7. Final discussion. The study of environmental risk is inherently complex, as it encompasses a multitude of interacting natural and anthropogenic processes, each evolving under uncertainty and exhibiting time-varying dynamics. The modeling of air quality data, in particular, is a notable challenge, since relatively little is known about the way air pollutants interact within an open atmospheric system and about the endogenous and exogenous factors that drive their joint evolution in a urban area. Consequently, their modeling requires the integration of diverse statistical tools capable of capturing their faceted behavior.

We introduced a comprehensive and flexible framework for environmental risk assessment based on a nonhomogeneous hidden semi-Markov model with multivariate, autoregressive, and state-dependent emission structures. The proposed model represents a novel, data-driven approach to model pollution emissions, allowing for the simultaneous estimation of hidden

regimes, pollutant dynamics, and risk contributions in a coherent statistical setting. Simulation studies confirm the reliability of the estimation procedure, demonstrating that the model parameters are accurately recovered under various configurations. Importantly, the computational burden of the estimation process is manageable and scalable. While model complexity increases with the number of hidden states, the length of dwell times, and the dimensionality of the outcome variables, the framework remains tractable and can be implemented efficiently with modern computing resources. The model's structure also lends itself to further generalization, incorporating other kinds of penalties to improve regularization or alternative specifications of the hazard function. A general framework such as the one proposed here offers a structured platform where not only these tools can be efficiently integrated but also distinct data features can be allocated to different levels of the model hierarchy, thereby enhancing flexibility and interpretability.

The empirical application to the Bergen air quality data demonstrates the model's versatility and practical relevance. The results confirm that the air quality of the area is generally good, although temporary episodes of elevated pollutant concentrations might occur during periods of intensified human activity and unfavorable meteorological conditions. Unlike simpler approaches, the proposed framework captures the time-varying and regime-dependent dependence structures among pollutants, allowing for a richer and more realistic representation of urban air quality dynamics than static or univariate models could achieve. In particular, the model provides a principled way to predict transitions between favorable and unfavorable air quality conditions, fully quantifying their timeliness in terms of regime-specific hazard functions. This dynamic formulation enables the continuous updating of predictions regarding the expected duration of high- or low-risk regimes, leading to the major conclusion that the polluted state behaves as a *quasi-absorbing* regime that persists unless favorable meteorological conditions occur. The integration with dynamic risk measures and the Shapley value decomposition yields an interpretable and time-resolved characterization of inter-pollutant risk propagation, facilitating the identification of systemically paired groups of pollutants and highlighting the key contributors to overall air pollution risk. This offers valuable insights into the mechanisms driving urban air quality fluctuations.

Ultimately, the proposed framework provides a powerful decision-support tool for targeted environmental interventions, enabling policymakers to identify the most influential risk-driving pollutants and monitor how their impacts evolve over time. Its flexible structure and interpretability make it broadly applicable to other domains of environmental and systemic risk assessment beyond air quality monitoring.

Significance statement. We develop a hierarchical model that captures multivariate, time-varying dependencies among pollutants and links regime transitions to environmental covariates. Compared to simpler univariate or standard hidden Markov approaches, our framework allows the estimation of regime-specific correlation structures and flexible state durations. The results reveal that high-pollution regimes are nearly absorbing unless favorable meteorological conditions occur, explaining why many short-term emission reduction measures have limited long-term impact. These findings advance the quantitative understanding of urban air dynamics and support more effective pollution management strategies.

Data availability statement. The air quality data from Danmarksplads, Bergen, is provided by the National Air Quality Reference Laboratory and is publicly available at <https://luftkvalitet.nilu.no/en/historical>. The weather observations from station SN50540 Florida, Bergen, are downloaded from <https://seklima.met.no/en> and are managed by The Norwegian Meteorological Institute.

Acknowledgments. The authors wish to express their sincere gratitude to Sondre Nedreås Hølleland for his assistance with the data acquisition process and for providing insightful perspectives on pollution patterns in Bergen.

Funding. This work has been supported by MIUR, grant number 2022XRHT8R—The SMILE project: Statistical Modelling and Inference for Living the Environment. The contribution of Pierfrancesco Alaimo Di Loro to this work has been supported by PON “Ricerca e Innovazione” 2014–2020 (PON R&I FSE-REACT EU), Azione IV.6 “Contratti di ricerca su tematiche Green,” grant number 60-G-34690-1.

SUPPLEMENTARY MATERIAL

Additional analysis, simulations and methodological details (DOI: [10.1214/26-AOAS2142SUPPA](https://doi.org/10.1214/26-AOAS2142SUPPA); .pdf). Supplementary Material is available in a separate file. It includes further information about the analyzed data, additional results of the simulations and real data application, and methodological notes excluded from the main text for reasons of space.

Data and computer programs (DOI: [10.1214/26-AOAS2142SUPPB](https://doi.org/10.1214/26-AOAS2142SUPPB); .zip). Data and computer programs developed for reproducibility of the results in the paper are available as Supplementary Material in the form of a compressed folder. The folder also includes a detailed readme file for its use. This can also be downloaded from a GitHub repository https://github.com/minmar94/envrisk_hsmm_varlasso.

REFERENCES

- ADRIAN, T. and BRUNNERMEIER, M. K. (2016). CoVaR. *Amer. Econ. Rev.* **106** 1705–1741.
- ANDERSSON, A. (2021). Mechanisms for log normal concentration distributions in the environment. *Sci. Rep.* **11** 16418.
- BARAGAÑO, D., RATÍE, G., SIERRA, C., CHRASTNÝ, V., KOMÁREK, M. and GALLEGÓ, J. (2022). Multiple pollution sources unravelled by environmental forensics techniques and multivariate statistics. *J. Hazard. Mater.* **424** 127413.
- BARBU, V. S. and LIMNIO, N. (2009). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis* **191**. Springer, Berlin.
- BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. [MR3357870 https://doi.org/10.1214/15-AOS1315](https://doi.org/10.1214/15-AOS1315)
- BERGEN KOMMUNE (2023). Årsrapport luftkvalitet i Bergen 2023 Technical Report Bergen Kommune.
- BERNARDI, M., MARUOTTI, A. and PETRELLA, L. (2017). Multiple risk measures for multivariate dynamic heavy-tailed models. *J. Empir. Finance* **43** 1–32.
- BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 719–725.
- BOAZ, R. M., LAWSON, A. B. and PEARCE, J. L. (2019). Multivariate air pollution prediction modeling with partial missingness. *Environmetrics* **30** e2592. [MR4021440 https://doi.org/10.1002/env.2592](https://doi.org/10.1002/env.2592)
- BOUYEYRON, C., JACQUES, J., SCHMUTZ, A., SIMÕES, F. and BOTTINI, S. (2022). Co-clustering of multivariate functional data for the analysis of air pollution in the south of France. *Ann. Appl. Stat.* **16** 1400–1422. [MR4455886 https://doi.org/10.1214/21-aos1547](https://doi.org/10.1214/21-aos1547)
- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer, Heidelberg. [MR2807761 https://doi.org/10.1007/978-3-642-20192-9](https://doi.org/10.1007/978-3-642-20192-9)
- CAO, C. (2024). Integration of ten years of daily weather, traffic, and air pollution data from Norway’s six largest cities. *Sci. Data* **11** 744.
- CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, New York. [MR2159833](https://doi.org/10.1214/05-aoas1547)
- CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. [MR2847974 https://doi.org/10.1198/jasa.2011.tm10159](https://doi.org/10.1198/jasa.2011.tm10159)
- EFRON, B. (2000). The bootstrap and modern statistics. *J. Amer. Statist. Assoc.* **95** 1293–1296. [MR1825279 https://doi.org/10.2307/2669773](https://doi.org/10.2307/2669773)

- FINAZZI, F., SCOTT, E. M. and FASSÒ, A. (2013). A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **62** 287–308. [MR3045878](#) <https://doi.org/10.1111/rssc.12001>
- FONG, P. W., LI, W. K., YAU, C. W. and WONG, C. S. (2007). On a mixture vector autoregressive model. *Canad. J. Statist.* **35** 135–150. [MR2345379](#) <https://doi.org/10.1002/cjs.5550350112>
- GENG, G., ZHANG, Q., TONG, D., LI, M., ZHENG, Y., WANG, S. and HE, K. (2017). Chemical composition of ambient PM_{2.5} over China and relationship to precursor emissions during 2005–2012. *Atmos. Chem. Phys.* **17** 9187–9203.
- HADJ-AMAR, B., JEWSON, J. and VANNUCCI, M. (2024). Bayesian sparse vector autoregressive switching models with application to human gesture phase segmentation. *Ann. Appl. Stat.* **18** 2511–2531. [MR4782500](#) <https://doi.org/10.1214/24-aos1892>
- HOSKOVEC, L., KOSLOVSKY, M. D., KOEHLER, K., GOOD, N., PEEL, J. L., VOLCKENS, J. and WILSON, A. (2023). Infinite hidden Markov models for multiple multivariate time series with missing data. *Biometrics* **79** 2592–2604. [MR4644018](#) <https://doi.org/10.1111/biom.13715>
- KHALILI, A. and CHEN, J. (2007). Variable selection in finite mixture of regression models. *J. Amer. Statist. Assoc.* **102** 1025–1038. [MR2411662](#) <https://doi.org/10.1198/016214507000000590>
- KOSLIK, J.-O. (2025). Hidden semi-Markov models with inhomogeneous state dwell-time distributions. *Comput. Statist. Data Anal.* **209** 108171. [MR4880548](#) <https://doi.org/10.1016/j.csda.2025.108171>
- LAGONA, F. and MINGIONE, M. (2025). Nonhomogeneous hidden semi-Markov models for toroidal data. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **74** 142–166. [MR4849420](#) <https://doi.org/10.1093/jrsssc/qlae049>
- LI, S. (2020). Debiasing the debiased Lasso with bootstrap. *Electron. J. Stat.* **14** 2298–2337. [MR4112726](#) <https://doi.org/10.1214/20-EJS1713>
- LIANG, D., ZHANG, H., CHANG, X. and HUANG, H. (2021). Modeling and regionalization of China's PM_{2.5} using spatial-functional mixture models. *J. Amer. Statist. Assoc.* **116** 116–132. [MR4227679](#) <https://doi.org/10.1080/01621459.2020.1764363>
- LIAO, K., PARK, E. S., ZHANG, J., CHENG, L., JI, D., YING, Q. and YU, J. Z. (2021). A multiple linear regression model with multiplicative log-normal error term for atmospheric concentration data. *Sci. Total Environ.* **767** 144282.
- MARTINEZ-ZARZOSO, I. and MARUOTTI, A. (2013). The environmental Kuznets curve: Functional form, time-varying heterogeneity and outliers in a panel setting. *Environmetrics* **24** 461–475. [MR3137747](#) <https://doi.org/10.1002/env.2232>
- MARUOTTI, A. and ALAIMO DI LORO, P. (2023). CO₂ emissions and growth: A bivariate bidimensional mean-variance random effects model. *Environmetrics* **34** e2793. [MR4614191](#) <https://doi.org/10.1002/env.2793>
- MARUOTTI, A., BULLA, J., LAGONA, F., PICONE, M. and MARTELLA, F. (2017). Dynamic mixtures of factor analyzers to characterize multivariate air pollutant exposures. *Ann. Appl. Stat.* **11** 1617–1648. [MR3709572](#) <https://doi.org/10.1214/17-AOAS1049>
- MERLO, L., MARUOTTI, A., PETRELLA, L. and PUNZO, A. (2022). Quantile hidden semi-Markov models for multivariate time series. *Stat. Comput.* **32** 61. [MR4466973](#) <https://doi.org/10.1007/s11222-022-10130-1>
- MINGIONE, M., ALAIMO DI LORO, P., LAGONA, F. and MARUOTTI, A. (2026). Supplement to “Environmental risk assessment via nonhomogeneous hidden semi-Markov models with penalized vector autoregression.” <https://doi.org/10.1214/26-AOAS2142SUPPA>, <https://doi.org/10.1214/26-AOAS2142SUPPB>
- O’CONNELL, J. and HØJSGAARD, S. (2011). Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *J. Stat. Softw.* **39** 1–22.
- OTT, W. R. (1990). A physical explanation of the lognormality of pollutant concentrations. *J. Air Waste Manage. Assoc.* **40** 1378–1383.
- OUYANG, W., GUO, B., CAI, G., LI, Q., HAN, S., LIU, B. and LIU, X. (2015). The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing. *Sci. Total Environ.* **505** 306–314.
- RICCIOTTI, L., PICONE, M., POLLICE, A. and MARUOTTI, A. (2025). A zero-inflated hidden semi-Markov model with covariate-dependent sojourn parameters for analysing marine data in the Venice lagoon. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **74** 506–529. [MR4896628](#) <https://doi.org/10.1093/jrsssc/qlae065>
- RUIZ-SUAREZ, S., LEOS-BARAJAS, V. and MORALES, J. M. (2022). Hidden Markov and semi-Markov models: When and why are these models useful for classifying states in time series data? *J. Agric. Biol. Environ. Stat.* **27** 339–363. [MR4416787](#) <https://doi.org/10.1007/s13253-021-00483-x>
- SHAN, X., CASEY, J. A., SHEARSTON, J. A. and HENNEMAN, L. R. (2024). Methods for quantifying source-specific air pollution exposure to serve epidemiology, risk assessment, and environmental justice. *GeoHealth* **8** e2024GH001188.
- STÄDLER, N., BÜHLMANN, P. and VAN DE GEER, S. (2010). ℓ_1 -penalization for mixture regression models. *TEST* **19** 209–256. [MR2677722](#) <https://doi.org/10.1007/s11749-010-0197-z>

- STÄDLER, N. and MUKHERJEE, S. (2013). Penalized estimation in high-dimensional hidden Markov models with state-specific graphical models. *Ann. Appl. Stat.* **7** 2157–2179. [MR3161717](#) <https://doi.org/10.1214/13-AOAS662>
- TAN, L., CHIONG, K. X. and MOON, H. R. (2021). Estimation of high-dimensional seemingly unrelated regression models. *Econometric Rev.* **40** 830–851. [MR4296872](#) <https://doi.org/10.1080/07474938.2021.1889195>
- TAVELLA, R. A., GALEAO DA ROSA MORAES, N., MACIEL AICK, C. D., RAMIRES, P. F., PEREIRA, N., SOARES, A. G. and DA SILVA JÚNIOR, F. M. R. (2023). Weekend effect of air pollutants in small and medium-sized cities: The role of policies stringency to COVID-19 containment. *Atmos. Pollut. Res.* **14** 101662.
- THUNIS, P., CLAPPIER, A., BEEKMANN, M., PUTAUD, J. P., CUVELIER, C., MADRAZO, J. and DE MEIJ, A. (2021). Non-linear response of $PM_{2.5}$ to changes in NO_x and NH_3 emissions in the Po basin (Italy): Consequences for air quality plans. *Atmos. Chem. Phys.* **21** 9309–9327.
- VISSER, I., RAIJMAKERS, M. E. and MOLENAAR, P. C. (2000). Confidence intervals for hidden Markov model parameters. *Br. J. Math. Stat. Psychol.* **53** 317–327.
- YU, S.-Z. (2015). *Hidden Semi-Markov Models: Theory, Algorithms and Applications*. Morgan Kaufmann, San Mateo.
- ZHAI, S., JACOB, D. J., PENDERGRASS, D. C., COLOMBI, N. K., SHAH, V., YANG, L. H., ZHANG, Q., WANG, S., KIM, H. et al. (2023). Coarse particulate matter air quality in East Asia: Implications for fine particulate nitrate. *Atmos. Chem. Phys.* **23** 4271–4281.
- ZHANG, B., JIAO, L., XU, G., ZHAO, S., TANG, X., ZHOU, Y. and GONG, C. (2018). Influences of wind and precipitation on different-sized particulate matter concentrations (PM 2.5, PM 10, PM 2.5–10). *Meteorol. Atmos. Phys.* **130** 383–392.
- ZHU, G., WEN, Y., CAO, K., HE, S. and WANG, T. (2024). A review of common statistical methods for dealing with multiple pollutant mixtures and multiple exposures. *Front. Public Health* **12** 1377685.
- ZUCCHINI, W., MACDONALD, I. L. and LANGROCK, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*, 2nd ed. *Monographs on Statistics and Applied Probability* **150**. CRC Press, Boca Raton, FL. [MR3618333](#)