

# Bayesian Time-Series Analysis of Atmospheric NO<sub>2</sub> Concentrations

## Introduction

NO<sub>2</sub> is a short-lived atmospheric pollutant widely used as an indicator of air quality and emission-related policy outcomes.

In this study, we fit and compare three Bayesian time-series models of increasing temporal complexity to daily NO<sub>2</sub> measurements. Predictive performance is evaluated using the leave-one-out information criterion (LOOIC), Akaike-style model weights are computed to assess relative support, and the best-performing model is used to quantify the long-term NO<sub>2</sub> trend.

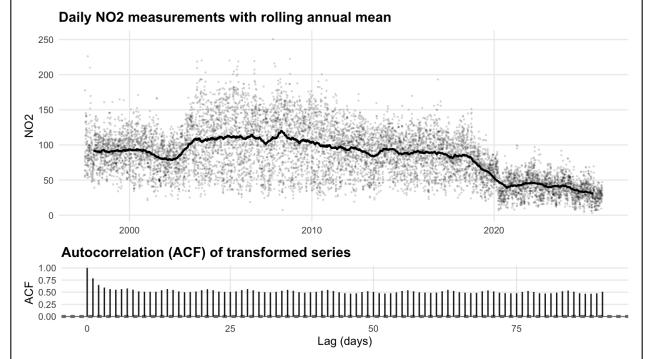
## Data

The dataset consists of daily NO<sub>2</sub> concentration measurements collected at a single European monitoring station between July 1997 and December 2025. Duplicate days were aggregated and missing values removed. Since NO<sub>2</sub> concentrations are non-negative and right-skewed, we applied the variance-stabilizing transformation  $y_t = \log(1 + \text{NO}_2 t)$ . Time was indexed as  $t = 1, \dots, N$  and standardized:  $t_t^{\text{std}} = \frac{t - \bar{t}}{\text{sd}(t)}$ . Seasonality was modeled using a Fourier basis with annual periodicity (365.25 days), consisting of sine and cosine terms.

## Models

All models describe daily NO<sub>2</sub> concentrations using a linear trend and deterministic seasonality. They differ only in how temporal dependence in deviations from this structure is modeled. The autocorrelation function (ACF) of the transformed series exhibits strong periodic behavior and substantial short-lag dependence, motivating progressively richer error structures.

All models use weakly informative priors.  $\mathcal{N}(0, \sigma^2)$  priors with moderate scale. Autoregressive parameters are modeled on an unconstrained scale and transformed using  $\tanh(\cdot)$  to ensure stationarity. Scale parameters follow exponential priors. Models are fitted using Hamiltonian Monte Carlo (`rstan`); all parameters satisfy  $\hat{R} \approx 1$  and show large effective sample sizes.



**Figure 1:** Light points: daily observations. Thick line: 365-day rolling mean. ACF computed on the model scale (standardized  $\log(1+\text{NO}_2)$ ). Persistent positive autocorrelation motivates AR errors.

tonian Monte Carlo (`rstan`); all parameters satisfy  $\hat{R} \approx 1$  and show large effective sample sizes.

### Model 1: Independent Errors

$$y_t \sim \mathcal{N}(\mu_t, \sigma), \quad (1)$$

$$\mu_t = \alpha + \beta t_t^{\text{std}} + \mathbf{x}_t^\top \mathbf{b}_{\text{seas}}. \quad (2)$$

This baseline model captures long-term change and seasonality, assuming conditional independence after removing these components.

### Model 2: AR(1) Residuals

Let  $\varepsilon_t = y_t - \mu_t$ . For  $t \geq 2$ ,

$$\varepsilon_t | \varepsilon_{t-1} \sim \mathcal{N}(\phi \varepsilon_{t-1}, \sigma_\varepsilon), \quad (3)$$

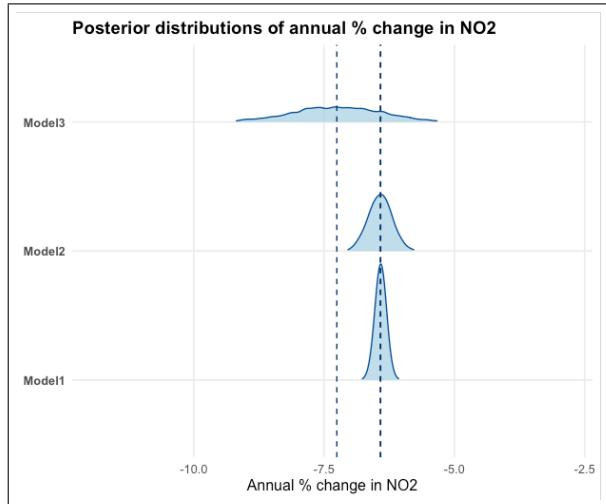
with the first observation treated as a boundary condition. This specification captures short-term persistence in daily deviations without introducing latent states.

### Model 3: Seasonal AR(1) Errors

$$r_t = \phi_S r_{t-S} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_{\text{sar}}), \quad (4)$$

$$y_t \sim \mathcal{N}(\mu_t + r_t, \sigma_{\text{obs}}). \quad (5)$$

This model allows deviations to recur across annual lags, beyond deterministic seasonality.



**Figure 2:** Beta posteriors ridgeline for all models.

## Results

### Q1: Which model best explains the NO<sub>2</sub> data?

Predictive performance was evaluated using PSIS leave-one-out cross-validation. Table 1 reports LOOIC values.

Model	LOOIC	SE
2: Trend + AR(1)	18894.6	178.1
3: Trend + seasonal AR	21143.8	177.9
1: No temporal dependence	25696.5	152.7

Model 2 substantially outperforms both alternatives, with differences of several thousand LOOIC units. This provides strong evidence that short-term temporal dependence is a key feature of daily NO<sub>2</sub> dynamics.

**Conclusion:** A conditional AR(1) model offers the best predictive description of the data.

### Q2: How to combine the ensemble?

Akaike-style weights computed from LOOIC values assign essentially all weight to Model 2, with negligible support for Models 1 and 3.

**Conclusion:** The optimal model ensemble collapses to Model 2; model averaging does not materially affect inference.

### Q3: What is the long-term trend?

Inference on the long-term trend is based on Model 2. The posterior distribution of the standardized slope  $\beta$  implies a clear downward trend.

After transformation to an annual scale,

$$\beta_{\text{year}} \in [-0.070, -0.062] \quad (90\% \text{ CI}),$$

with posterior median  $-0.066$ , which corresponds to an annual change of  $-6.8\%$  to  $-6.1\%$ .

**Conclusion:** Atmospheric NO<sub>2</sub> concentrations have declined substantially and persistently over the study period.

## Discussion

The analysis underscores the importance of explicitly modeling temporal dependence in environmental time series. A purely regression-based specification underfits the data by treating correlated observations as independent, while adding seasonal stochastic structure does not improve predictive performance. Among the models considered, the AR(1) specification achieves an effective balance between flexibility and parsimony, capturing dominant short-term dependence without unnecessary complexity.

This balance is reflected in the posterior distributions of the trend parameter. Relative to the independent-errors model, the AR(1) specification yields slightly broader and conservative uncertainty once temporal dependence is accounted for, whereas the seasonal autoregressive model produces a much more diffuse trend posterior, indicating reduced identifiability due to overflexible error structure. Along with its weaker predictive performance, this suggests that the additional stochastic seasonal component absorbs variation without improving out-of-sample fit.

**Limitations.** The analysis is based on a single monitoring station and does not account for spatial heterogeneity. Seasonal effects are assumed constant over time, and the study is observational rather than causal. Extreme events or unmodeled meteorological drivers may still challenge the Gaussian observation assumptions.

## Conclusion

Bayesian time-series modeling reveals a strong and statistically robust decline in atmospheric NO<sub>2</sub> concentrations over the past three decades. Accounting for short-term temporal dependence is essential for reliable inference, while additional seasonal autoregressive complexity is not supported by predictive evidence.