Time-varying relationship between SO_4^{2-}, NO_3^- and NH_4^+ concentrations in cumulative precipitation samples and gaseous pollutants

Marek Brabec 1), Iva Hunova 2), Marek Maly 1)

1) Department of Statistical Modeling and Artificial Intelligence, Institute of Computer Science, The Czech Academy of Sciences, Praha

2) Air Quality Section, Czech Hydrometeorological Institute, Praha

mbrabec@cs.cas.cz

Problem

When studying air pollution, it is desirable to model dynamics of both gaseous pollutant concentrations (NO_x, SO_2, NH_3) and concentrations of important ions (NO_3^-, SO_4^{2-}, NH_4^+) in precipitation samples as well as their relationships.

While conceptually simple, the modeling presents technical difficulties stemming from the fact that while gaseous concentrations are essentially point measurements, the precipitation samples are integrated across collection interval. Moreover, for long-term data, the length of the interval often changed profoundly. For example in CHMU data 1992–2024, it decreased from approximately 30 days to 1 day of aggregation lengths. In addition to the systematic long-term changes, smaller but non-negligible variation in the length is related to calendar and logistic irregularities (weekends, bank holidays etc.). Formally, this amounts to integration-like observation operator of length that changes both systematically (on large time scale) and randomly (on smaller scale).

Introduction

In the past, we formulated a statistical modeling approach to deal with non-negligible and possibly time-varying aggregation length in a formalized way:

- Hunova, Brabec, Maly, Skachova (2022),
- Hunova, Brabec, Maly (2024).

The model is based on a flexible random-walk dynamics and Bayesian approach using the INLA (Integrated Nested Laplace Approximation) that was computationally highly attractive at the time of publication. We successfully demonstrated that the approach can be used for reconstruction of the daily series of latent concentrations and used their estimates to study the ratios of mean NO_3^-, SO_4^{2-}, NH_4^+ concentrations both in long-term and in seasonal views. While this modeling is a useful tool for marginal disaggregation, it falls short in not being able to focus on mutual statistical relationships.

In this work, we expand the previous work to cover bivariate relationships of NO_x-NO_3^-, SO_2-SO_4^{2-}, NH_3-NH_4^+ pairs. Moreover, we allow the relationship to change in time. Time-varying model formulation gives us a statistical methodology to study both long-term and seasonal variation in gaseous-precipitation concentration pair.

Statistical model

When analyzing bivariate relations in time series, it is well known that simple-minded direct approach can easily lead to spurious and misleadingly high estimates (Granger, Newbold 1978) in case that the series are nonstationary (time series operators have unit roots). In the previous work (Hunova, Brabec, Maly, Skachova 2022), we have demonstrated that this is really the case. So, we base our bivariate disaggregation statistical model on latent cointegrated vector and VECM (Vector Error Correction, Lutkepohl 2006) with time-varying long-term coefficient.

Observation equations:

$$\begin{pmatrix} log(Z_{gas,t}) \\ log(Z_{precip,t}) \end{pmatrix} = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} \sum_{s \in I_t} \mu_{1,s} \\ \sum_{s \in I_t} \mu_{2,s} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, diag(\sigma_1^2, \sigma_2^2) \right)$$

State-equations:

$$\begin{pmatrix} \mu_{1,s} \\ \mu_{2,s} \end{pmatrix} = \begin{pmatrix} \mu_{1,s-1} \\ \mu_{2,s-1} + \alpha.(\mu_{2,s-1} - \beta_{i(s)}.\mu_{1,s-1}) \end{pmatrix} + \begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, diag(\tau_1^2, \tau_2^2) \right)$$

where $i(s)$ is the year-month index of time s.

Equilibrium coefficient dynamics:

$$\beta_i = \beta_{i-1} + \omega_i$$

$$\omega_i \sim N(0, \psi^2)$$

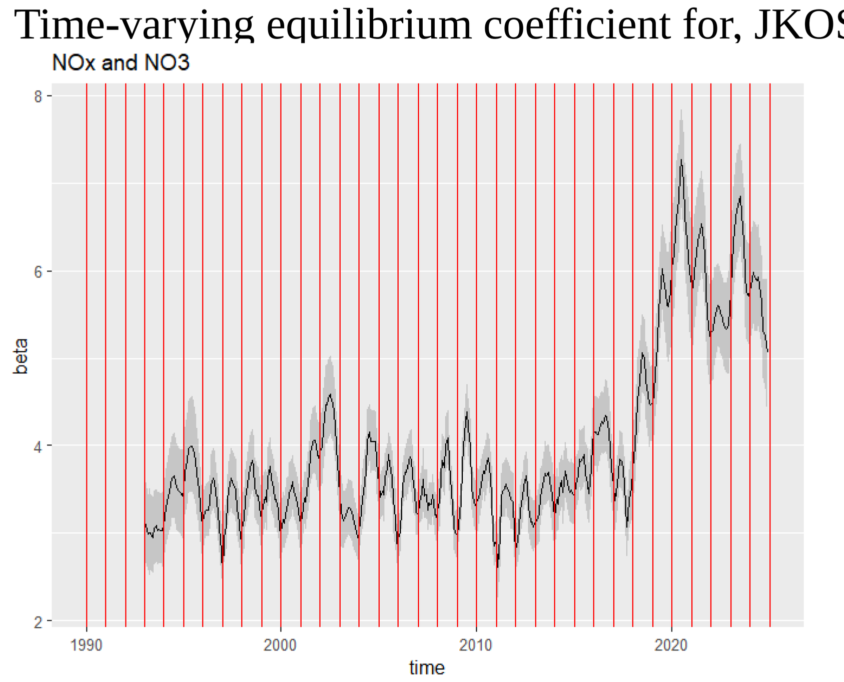
Initial states and alpha prior: Gaussian
Standard deviation priors: half-normal

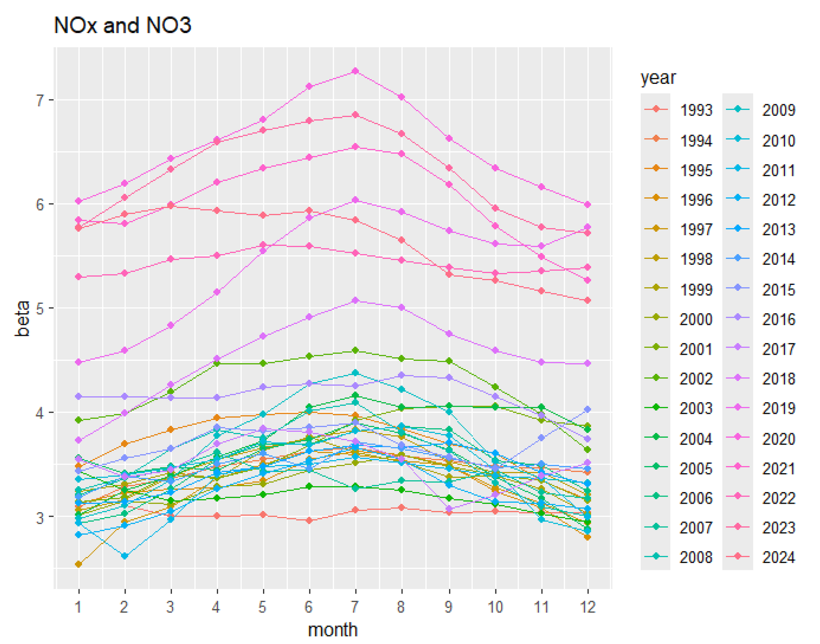
Estimation

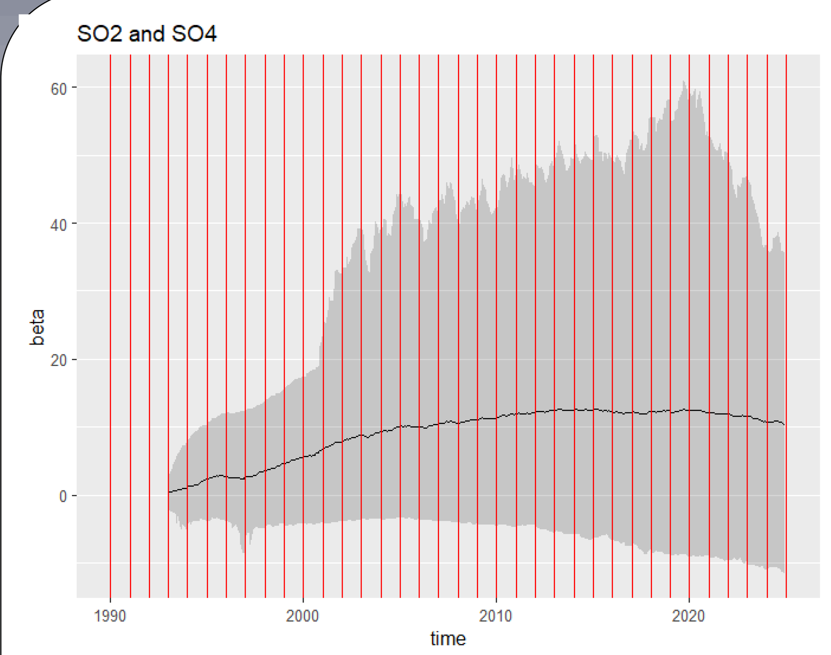
We conduct computationally very effective Hamiltonian MCMC, implemented in the probabilistic programing language Stan (Carpenter et al. 2017) to obtain posterior distributions (running 4 parallel chains for 2000 iterations each, then discarding first half as a burn-in). We present posterior means and 2.5%, 97.5% percentiles as summaries of quantities of major interest.

Results

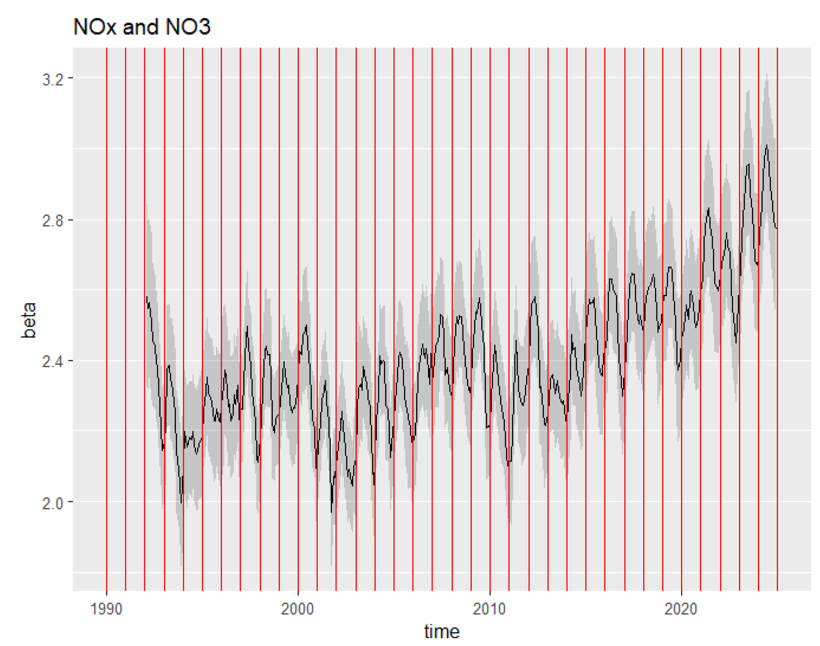
Time-varying equilibrium coefficient for, JKOS:

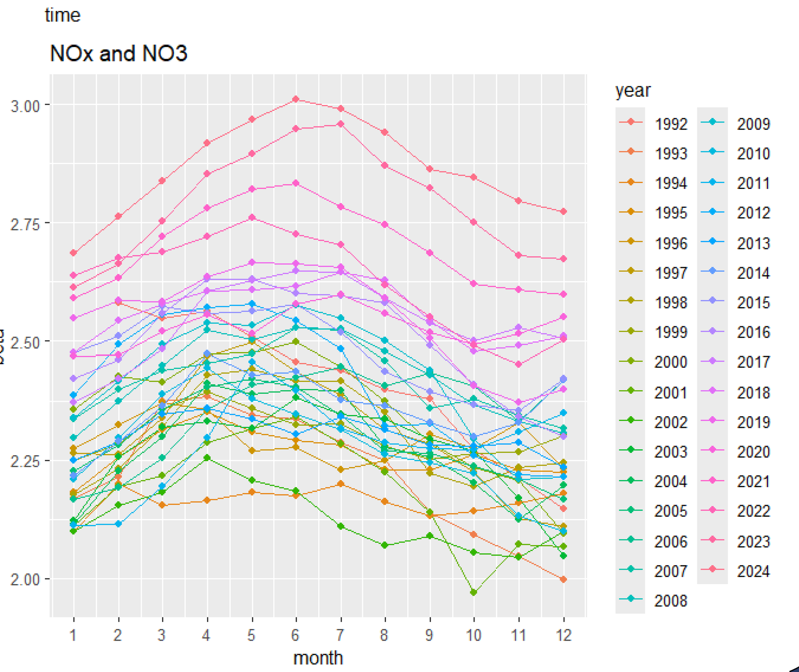




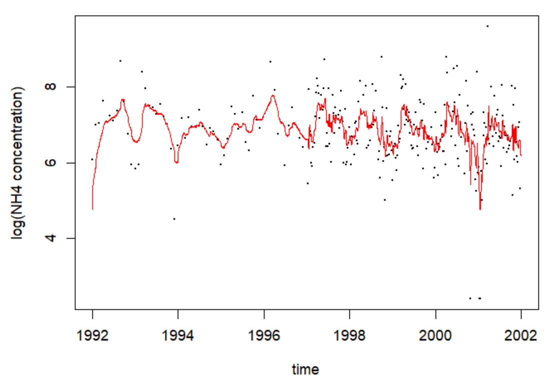


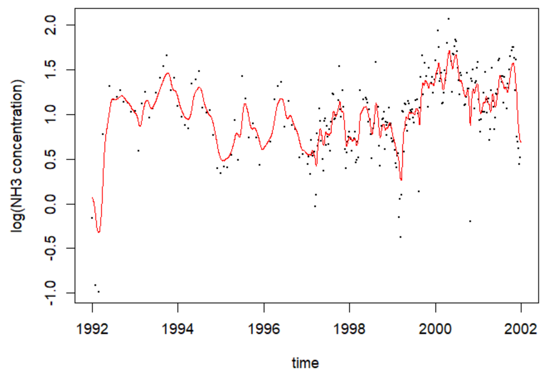
ALIB:





Data and latent components, NH_3-NH_4^+ pair, ALIB:





Summary

In this work we:

- formulate a novel comprehensive bivariate time series model
- use state-space approach
- cover technical problems with time-varying artifacts (systematically and randomly varying aggregation length)
- allow for integration and cointegration in latent bivariate component to prevent spurious regressions
- allow for time-varying relationship between gaseous and precipitation concentrations
- and study both long-term and seasonal changes in the strength of the relationship between gaseous and precip concentrations
- utilize flexible Bayesian implementation
- use computationally effective HMCMC computations in Stan
- study NO_x-NO_3^-, SO_2-SO_4^{2-}, pairs on long-term CHMU time series, 1992-2024 (and shorter series for NH_3-NH_4^+)
- compare background rural (JKOS) and background urban (ALIB) stations
- show log-term increasing trend seasonal changes in NO_x-NO_3^- efficiency

Acknowledgement

This research has been funded by the Czech Technological Agency project SS02030031, ARAMIS.

Bibliography

Carpenter,B.-Gelman,A.-Hoffman,M.-Lee,D.-Goodrich,B.-Betancourt,M.-Brubaker,M.; Guo,J.-Li,P.-Riddell,A. (2017): Stan: a probabilistic programming language. Journal of Statistical Software. 76 (1): 1–32.
Gelman,A.-Carlin,J.B.-Stern,H.S.-Dunson,D.B.-Vehtari,A.-Rubin,D.B. (2013): Bayesian data analysis. 3rd ed. CRC Press.
Granger,C.W.J.-Newbold,P. (1978): Spurious regression in Economerics. Journal of Econometrics, 2 (2), 111-120.
Hunova,I.-Brabec,M.-Maly,M.-Skachova,H. (2022): Reconstruction of daily courses courses of SO_4^{2-}, NO_3^-, NH_4^+ concentrations in precipitation from cumulative samples. Atmosphere 13: 1049
Hunova,I.-Brabec,M.-Maly,M. (2024): Major ions in Central European precipitation – insight into changes in NO3-/SO42-, NH4+/NO3- and NH4+/SO42- ratios over the last four decades. Chemosphere 349 (2024) 140986
Lutkepohl,H. (2006): New introduction to multiple time series. Springer.
R Core Team (2025): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rue,H.-Martino,S.-Chopin,N. (2009): Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). Journal of the Royal Statistical Society, Series B 71(2): 319–392.