Practicum: Spatio-Temporal Dynamic Linear Models

Carlo Zaccardi and Alexandra M. Schmidt

ABS25 Applied Bayesian Statistics School

03-06, June 2025

General DLM

- **Dynamic Linear Models (DLM)** are a class of models characterized by their ability to model time-varying relationships.
- DLMs consist of two main components: the **observation equation** and the **state equation**.

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \boldsymbol{z}_t' \gamma + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{F}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_t, \quad \boldsymbol{\beta}_t \sim N(0, \boldsymbol{W}) \tag{2}$$

where:

- y_t is the observed time series at time $t = 1, \ldots, T$.
- $\boldsymbol{x}_t$ is a p-dimensional vector of covariates at time t.
- $\boldsymbol{F}_t$ is a $p \times p$ evolution matrix.
- $\boldsymbol{\beta}_t$ is the state vector at time t, initialized with $\boldsymbol{\beta}_1 \sim N(0, \boldsymbol{D})$.
- $\boldsymbol{z}_t$ is a q-dimensional vector of covariates whose effect is constant (including an intercept term).
- γ is a vector of coefficients for $\boldsymbol{z}_t$.
- ϵ_t is the observation error.
- $\boldsymbol{W}$ is the covariance matrix of the state evolution error.

Spatio-Temporal DLM

- Spatio-temporal DLMs extend the basic DLM framework to incorporate spatial dependencies.
- The state equation can be adapted to account for spatial correlations in the state vector.
- Let p denote the number of spatial locations where data is collected, then $\boldsymbol{\beta}_t = (\beta_{1,t}, \ldots, \beta_{p,t})'$, where $\beta_{i,t}$ is the effect of the covariate $x_{i,t}$ on the outcome at location $i = 1, \ldots, p$ and time $t = 1, \ldots, T$.
- The state evolution covariance matrix $\boldsymbol{W}$ can be structured to reflect **spatial relationships**, e.g. by assuming an *exponential covariance function* if the data are point-referenced:

$$\boldsymbol{W}_{ij} = \sigma^2 \exp\left(-\frac{d_{ij}}{\phi}\right), \quad d_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \tag{3}$$

  where σ^2 is the variance, ϕ is the range parameter, and d_{ij} is the Euclidean distance between locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$.

ANOVA Decomposition of the State Vector

- The state vector $\boldsymbol{\beta}_t$ can be decomposed into components that capture different sources of variation [see, e.g., Bakar et al., 2015, Cai et al., 2013, Paez et al., 2008]:

$$\beta_{i,t} = \overline{\delta} + \delta_i + \delta_t + \delta^*_{i,t} \tag{4}$$

  where:
  - $\overline{\delta}$ is the overall mean effect.
  - δ_i is the spatial effect at location i: average (across time) spatial deviation from the overall effect
  - δ_t is the temporal effect at time t: average (across space) temporal changes from the overall effect
  - $\delta^*_{i,t}$ is the interaction effect between location i and time t: it captures the site- and time-specific deviations from the overall effect that are left after accounting for the temporal and spatial effects
- Advantages:
  - It can help **identify patterns** in the covariate's effect, such as spatial trends or temporal changes.
  - **Easier prior elicitation and better interpretability** of the covariate's effect.
  - It allows for **non-separable** spatio-temporal structure for $\beta_{i,t}$.

Bayesian Inference

- The Bayesian model is completed by specifying distributions for the single components of the state vector:

$$\bar{\delta} \sim N(0, \sigma^2_{\bar{\delta}}) \tag{5}$$

$$\delta_t \sim N(\delta_{t-1}, \sigma^2_{\delta_t}) \tag{6}$$

$$\tilde{\boldsymbol{\delta}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{7}$$

$$\boldsymbol{\delta}^*_t \sim N(\boldsymbol{\delta}^*_{t-1}, \boldsymbol{\Sigma}^*) \tag{8}$$

  where $\tilde{\boldsymbol{\delta}} = (\delta_1, \ldots, \delta_p)'$ is the vector of spatial effects and $\boldsymbol{\delta}^*_t = (\delta^*_{1,t}, \ldots, \delta^*_{p,t})'$ is the vector of interaction effects at time t.
- The covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$ are parameterized through exponential covariance functions.
- Non-informative prior distributions are assumed for all parameters in the model.

Efficient Inference and Identifiability

- The algorithm proposed by Chan and Jeliazkov [2009] can be used to build an efficient sampler.
- To make the model identifiable, we impose **constraints** on the parameters at each MCMC iteration:
  - Set $\sum_{t=1}^{T} \delta_t = 0$.
  - Set $\sum_{i=1}^{p} \delta_i = 0$.
  - Set $\sum_{i=1}^{p} \delta^*_{i,t} = 0$ for each $t = 1, \ldots, T$.
  - Set $\sum_{t=1}^{T} \delta^*_{i,t} = 0$ for each $i = 1, \ldots, p$.

Question: Pollution Spatio-Temporal Data

- The **AgrImOnIA dataset** is a comprehensive dataset relating air quality and livestock (expressed as the density of bovines and swine bred) along with weather and other variables. See Fassò et al. [2023].

- This dataset is a collection of estimated daily values for a range of measurements of different dimensions as: air quality (AQ), weather (WE), emissions (EM), livestock animals (LI) and land use (LA). Data are related to Lombardy and the surrounding area for 2016-2021, inclusive. The surrounding area is obtained by applying a 0.3° buffer on Lombardy borders.

- Visit https://zenodo.org/records/7956006 and download two files: `Agrimonia_Dataset_v_3_0_0.Rdata` and `Metadata_monitoring_network_registry_v_2_0_1.csv`

Question: Pollution Spatio-Temporal Data

- You are going to use the daily measurements of nitrogen dioxide (NO2) concentrations from the AQ dimension.
- NO2 is one of a group of highly reactive gases known as *oxides of nitrogen* or *nitrogen oxides (NOx)*. NO2 forms from emissions from cars, trucks and buses, power plants, and off-road equipment.
- It can cause significant health issues by irritating the lungs and can contribute to respiratory problems.
- Visit US EPA to know more about the pollutant's effect on human health and the environment.

Question: Pollution Spatio-Temporal Data

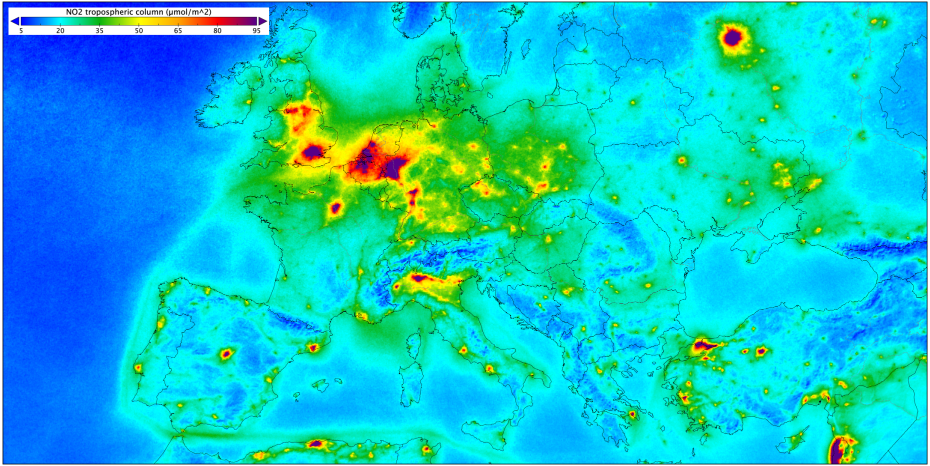

Figure: NO2 over Europe, based on measurements gathered by the Copernicus Sentinel-5P mission between April and September 2018.

Question: Pollution Spatio-Temporal Data

- The objective is to fit a spatio-temporal DLM that could help to understand the **relationship of NO2 with meteorological variables in Lombardy**:
  - boundary layer height (blh), 2 m temperature (temp2m), total precipitation (prec), surface pressure (spr), 10 m wind speed (ws10m).
- Interpret the estimated coefficients, in therms of their overall, spatial, temporal and interaction effects.
- How does the results change if you change the hyperparameters?

Programming Tip (Optional)

- The command `source('main_function.R')` sources the C++ code. However, it takes few minutes to execute because a miriad of warnings are generated by the C++ compiler that are then printed to the R console.
- If you want to source the C++ code quicker, proceed as follows (**Windows**):
  1. Open or create the folder `C:\Users\YourUsername\Documents\.R`. Note the dot before `R` in the folder name.
  2. Open or create the file `Makevars.win` in that folder.
  3. Copy the following lines into the file:
     - CXXFLAGS+=-Wno-ignored-attributes
     - CXX11FLAGS+=-Wno-ignored-attributes
     - CXX14FLAGS+=-Wno-ignored-attributes
     - CXX17FLAGS+=-Wno-ignored-attributes
  4. Ensure to end the file with an empty line. Save the file and close it. Open R Studio.
- (**Linux**) Create or open the file `Makevars` (without the extension) in the folder `/home/YourUsername/.R/` and copy the same lines as above.

References I

K Shuvo Bakar, Philip Kokic, and Huidong Jin. A spatiodynamic model for assessing frost risk in south-eastern australia. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 64(5): 755–778, 2015.

Bo Cai, Andrew B Lawson, Md Monir Hossain, Jungsoon Choi, Russell S Kirby, and Jihong Liu. Bayesian semiparametric model with spatially–temporally varying coefficients selection. *Statistics in medicine*, 32(21):3670–3685, 2013.

Joshua CC Chan and Ivan Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2): 101–120, 2009.

Alessandro Fassò, Jacopo Rodeschini, Alessandro Fusta Moro, Qendrim Shaboviq, Paolo Maranzano, Michela Cameletti, Francesco Finazzi, Natalia Golini, Rosaria Ignaccolo, and Philipp Otto. Agrimonia: a dataset on livestock, meteorology and air quality in the lombardy region, italy. *Scientific Data*, 10 (1):143, 2023.

Marina S Paez, Dani Gamerman, Flávia MPF Landim, and Esther Salazar. Spatially varying dynamic coefficient models. *Journal of Statistical Planning and Inference*, 138(4):1038–1058, 2008.