How Do Climatic Associations Between Canadian Cities Change Over Time?

Jie Jian

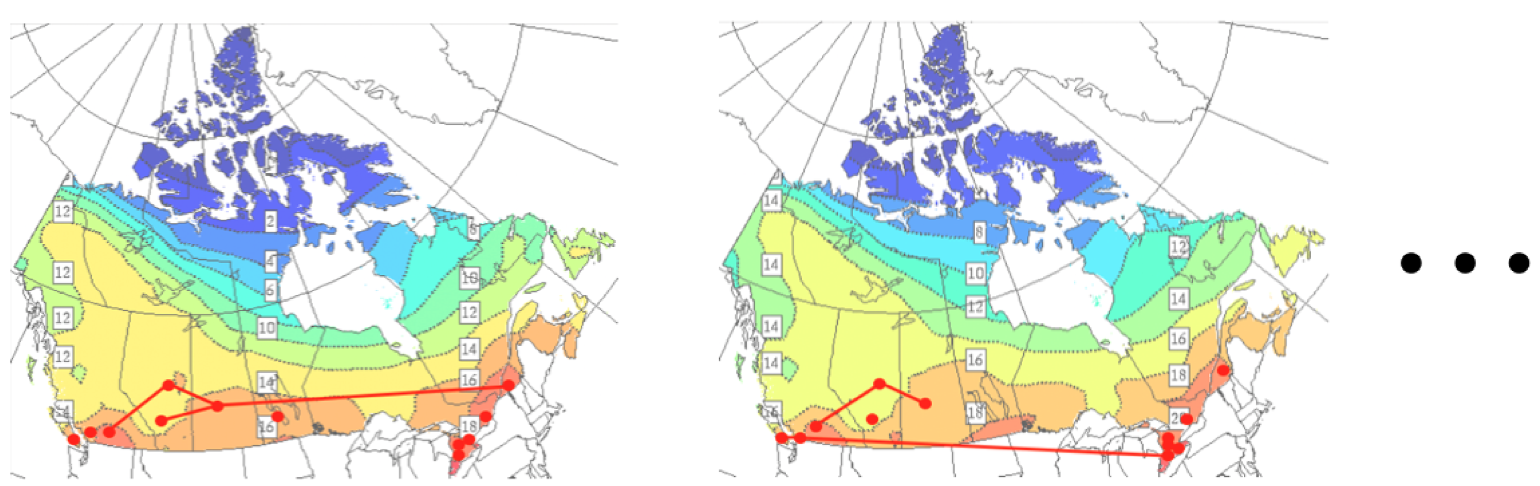
Department of Statistics and Actuarial Science, University of Waterloo

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
Department of Statistics and Actuarial Science

Detecting the Complex Dynamic Climatic Network

Climate is a complex dynamical system. Modelling the climate connectivity can uncover the spatiotemporal patterns of climatic relevance between spatially dispersed grid points.

- **Data:** Hourly temperature measurements in July for thirteen Canadian cities, taken repeatedly during the month of July between 2019 and 2022, i.e., $X_{ij}(t)$, for $j = 1, 2, ..., 13$ stations, $t = 1, 2, ..., 24$ time points, and $i = 1, 2, ..., 31 \times 4 = 124$ measurements. [1]
- **Model:** Each cities are **nodes**, and their partial correlations are **edges**.
- **Assumption:**
  1. At each temporal time point, $\boldsymbol{X}(t)$ follows multivariate Gaussian distribution. Namely, $(X_1(t), \cdots, X_p(t))^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{p\times p}(t))$.
  2. The network structure characterized by $\boldsymbol{\Sigma}_{p\times p}(t)$ changed smoothly over time.
- **To estimate:** All $\boldsymbol{\Sigma}_{p\times p}(t)$'s.



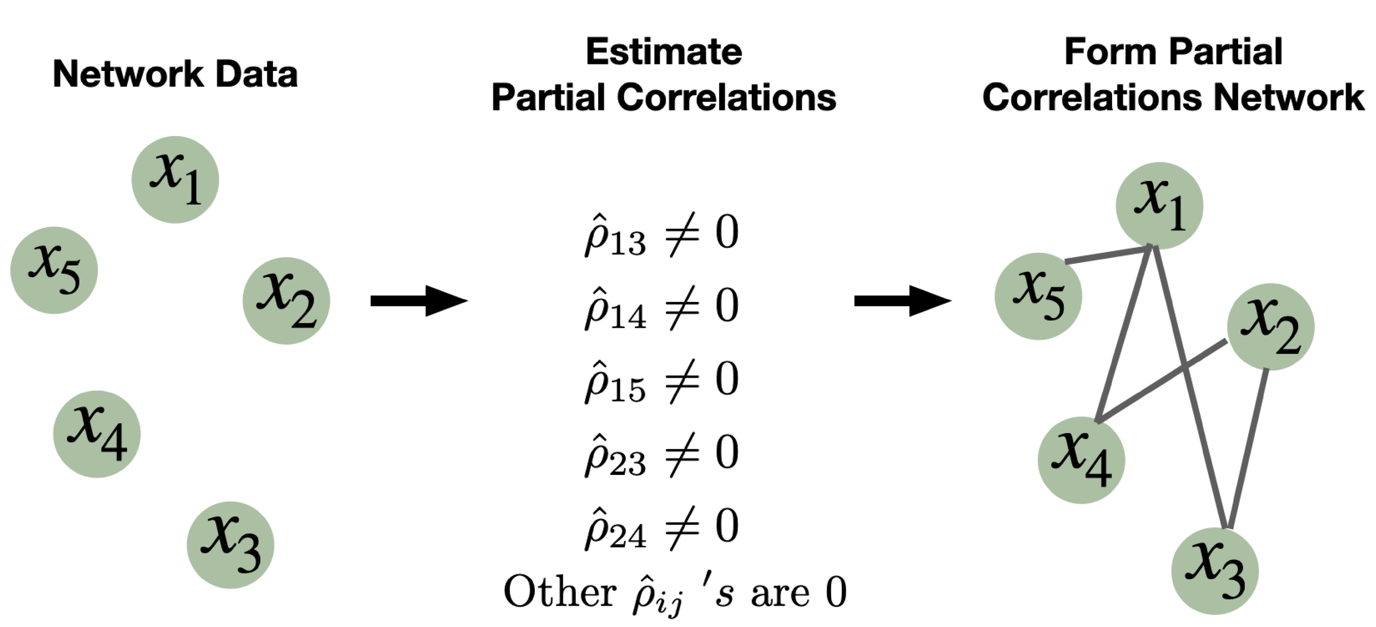
t=1 (1 AM) t=2 (2 AM)

Static Network Estimation Gaussian Graphical Model (GGM)

In static case, we observe the repeated sample from multivariate Gaussian random variables $(x_1, \cdots, x_p)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{p\times p})$, $(\boldsymbol{\Sigma}^{-1})_{ij} := \sigma^{ij}$.

- **Partial correlation:** Partial correlation between x_i and x_j, ρ_{ij}, measures the degree of association between two random variables conditional on the rest random variables. Moreover, $\rho_{ij} = -\sigma^{ij}/\sqrt{\sigma^{ii}\sigma^{jj}}$.
- **Conditional independence:** Under Gaussian assumption, $x_i \perp x_j | x \backslash \{i, j\} \Leftrightarrow \sigma^{ij} = 0$.
- **Estimate Partial Correlations by SPACE [3]:**

$$\min_{\{\rho_{ij}, i<j\}} \frac{1}{n}\sum_{i=1}^{p} \|X_i - \sum_{j<i}\rho_{ji}\sqrt{\frac{\sigma^{jj}}{\sigma^{ii}}}X_j - \sum_{j>i}\rho_{ij}\sqrt{\frac{\sigma^{jj}}{\sigma^{ii}}}X_j\|^2 + \lambda_1 \sum_{1\le i<j\le p}|\rho_{ij}|$$



Time-varying Network Estimation: Achieve Two Goals Simultaneously

- **Goal 1**: To obtain estimated network at each temporal time point: Stack the all temporal objective functions together;
- **Goal 2**: To have smoothness between two neighboring networks: Propose a penalty on the change of neighbouring networks.

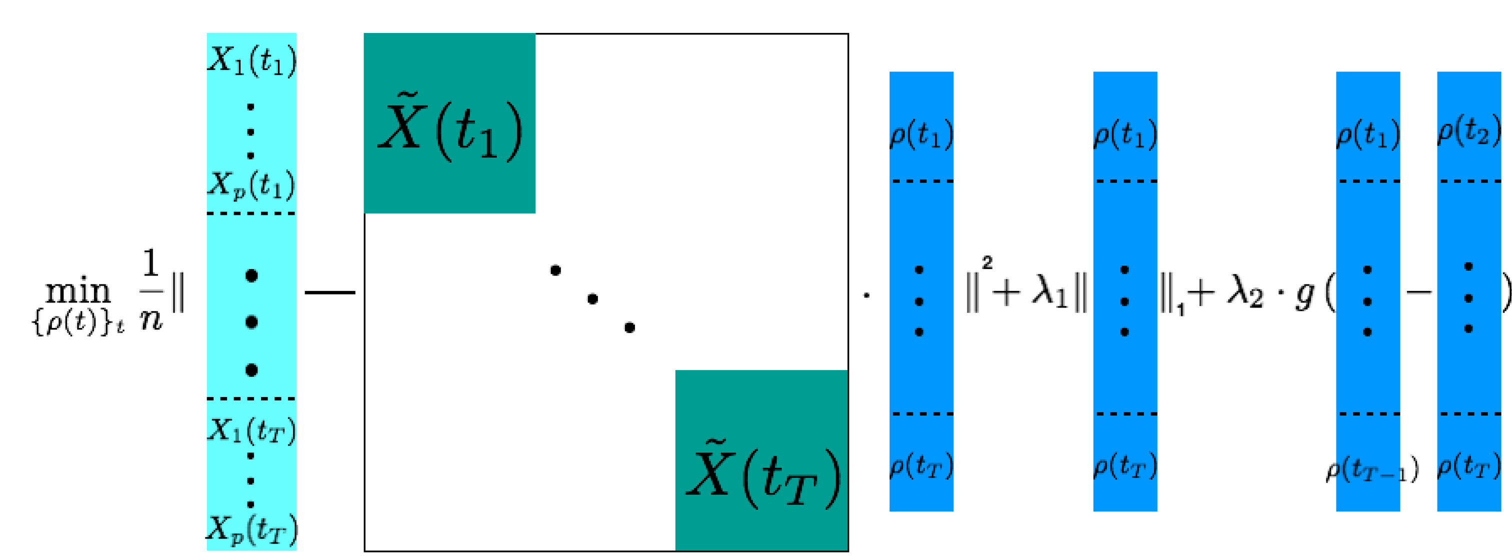
Objective Function

$$\min_{\{\rho_{ij}(t), i<j\}} \frac{1}{n}\sum_{k=1}^{T}\sum_{i=1}^{p} \|X_i(t_k) - \sum_{j<i}\rho_{ji}(t_k)\sqrt{\frac{\sigma^{jj}(t_k)}{\sigma^{ii}(t_k)}}X_j(t_k) - \sum_{j>i}\rho_{ij}(t_k)\sqrt{\frac{\sigma^{jj}(t_k)}{\sigma^{ii}(t_k)}}X_j(t_k)\|^2$$
$$+ \lambda_1 \sum_{k=1}^{T}\sum_{1\le i<j\le p}|\rho_{ij}(t_k)| + \lambda_2 \cdot \sum_{k=2}^{T}\sum_{1\le i<j\le p} g\left(\rho_{ij}(t_k) - \rho_{ij}(t_{k-1})\right),$$

where $g(\cdot) = (\cdot)^2$ for **generalized elastic net (GEN)** and $g(\cdot) = |\cdot|$ for **generalized fused lasso (GFL)**.

Challenges and How to Tackle Them

1. **Hard to compute and High computational cost:** We exploit the block structure of the matrix form and apply very fancy computational tricks when using ADMM algorithms for both the GEN and the GLF problems.



$$\Rightarrow \min_{\rho} \frac{1}{n}\| \ Y - \tilde X \rho \ \| + \lambda_1 \| \rho \|_1 + \lambda_2 \cdot g(\ D\rho \ )$$

where the matrix D takes the form of

$$\boldsymbol{D}_{(T-1)p(p-1)/2 \times Tp(p-1)/2} = \begin{bmatrix} I & -I & 0 & 0 & 0 \\ 0 & I & -I & 0 & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & 0 & I & -I \end{bmatrix} \quad (1)$$

In ADMM algorithm, we need to calculate the inverse of a large symmetric block tri-diagonal matrix $\left(\frac{2}{n}\tilde X^\top \tilde X + 2\lambda_2 D^\top D + aI\right)$. To help improve efficiency and accuracy of the inversion, we take advantage of the tridiagonal structure of the matrix to do parallel block-wise inversions first and then apply left and right multiplication of smaller block matrices.

2. **Tuning parameter selection:** We derive approximated degree-of-freedom formulae in BIC criterion to facilitate the tuning parameter selections.

For those interested in further exploration, I kindly invite you to [2] for more comprehensive insights.

What Do We Find?

We display the estimated time-varying temperature partial correlations between Canadian cities as follows. The blue is used for negative values and red is for positive values. All city pairs are listed in descending order of the distances between cities from top to the bottom.

(a) GEN (b) GFL

Main Findings

- More connections between cities that are geographically closer.
- Their estimated partial correlations tend to be positive.
- Connections occur more frequently in the afternoon.
- Saint John has connections with cities that are far away.

References

[1] Environment and Climate Change Canada. Environment and Climate Change Canada Historical Climate Data website. https://climate.weather.gc.ca/index_e.html, 2022. Accessed: 2022-10-16.

[2] Jie Jian, Peijun Sang, and Mu Zhu. Two gaussian regularization methods for time-varying networks, 2022.

[3] Jie Peng, Pei Wang, Nengfeng Zhou, and Ji Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746, 2009.

Acknowledgement

This is joint work with Dr. Peijun Sang and Dr. Mu Zhu.