Matrix Auto-regressive Model with Vector Time-series Covariates

Hu Sun [1], Zuofeng Shang [2], Yang Chen [1]

[1]Department of Statistics, University of Michigan, Ann Arbor

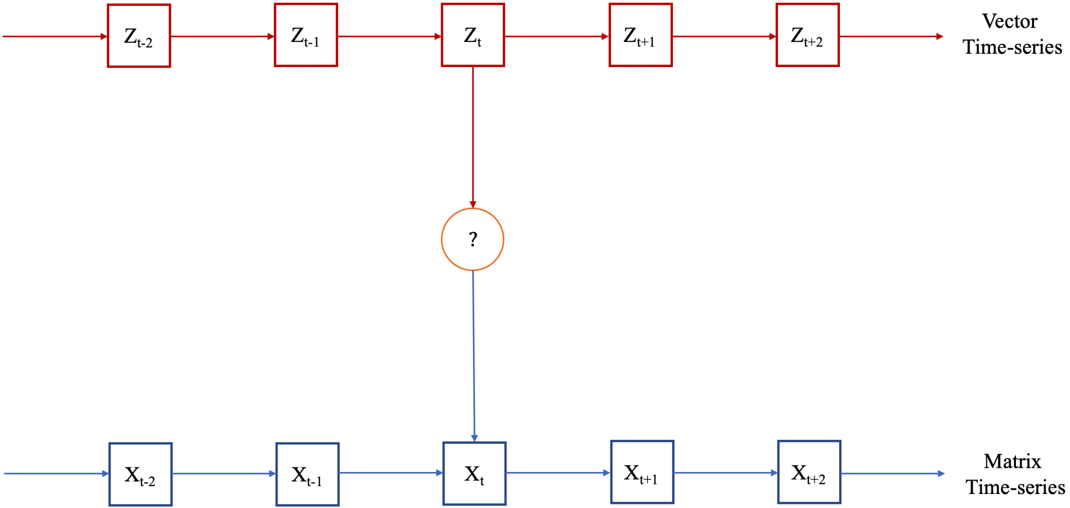
[2]Department of Mathematical Sciences, New Jersey Institute of Technology

January 29, 2023

Research Problem

- Data I: Matrix time series: $X_1, X_2, \ldots, X_t, \ldots, \ldots$, where $X_t \in \mathbb{R}^{M \times N}, \forall t$.
- Data II: Auxiliary multivariate time-series $z_1, z_2, \ldots, z_t, \ldots$, where $z_t \in \mathbb{R}^{D}, \forall t$.
- Problem: Forecast X_t with $X_{t-p}, X_{t-p+1}, \ldots, X_{t-1}$ together with $z_{t-q}, z_{t-q+1}, \ldots, z_{t-1}$

Research Problem



Model

- A general additive model can be formulated as:

$$X_t = \sum_{p=1}^{P} F_p(X_{t-p}) + \sum_{q=1}^{Q} G_q(z_{t-q}) + E_t$$

[1] R. Chen, H. Xiao, and D. Yang, "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, vol. 222, no. 1, pp. 539–560, 2021.

[2] N.-J. Hsu, H.-C. Huang, and R. S. Tsay, "Matrix autoregressive spatio-temporal models," *Journal of Computational and Graphical Statistics*, vol. 30, no. 4, pp. 1143–1155, 2021.

Model

- A general additive model can be formulated as:

$$X_t = \sum_{p=1}^{P} F_p(X_{t-p}) + \sum_{q=1}^{Q} G_q(z_{t-q}) + E_t$$

- $F_p(X_{t-p})$ ([1], [2]): $F_p(X_{t-p}) = A_p X_{t-p} B_p^T$, where $A_p \in \mathbb{R}^{M\times M}, B_p \in \mathbb{R}^{N\times N}$.

[1] R. Chen, H. Xiao, and D. Yang, "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, vol. 222, no. 1, pp. 539–560, 2021.
[2] N.-J. Hsu, H.-C. Huang, and R. S. Tsay, "Matrix autoregressive spatio-temporal models," *Journal of Computational and Graphical Statistics*, vol. 30, no. 4, pp. 1143–1155, 2021.

Model

- A general additive model can be formulated as:

$$X_t = \sum_{p=1}^{P} F_p(X_{t-p}) + \sum_{q=1}^{Q} G_q(z_{t-q}) + E_t$$

- $F_p(X_{t-p})$ ([1], [2]): $F_p(X_{t-p}) = A_p X_{t-p} B_p^T$, where $A_p \in \mathbb{R}^{M \times M}, B_p \in \mathbb{R}^{N \times N}$.
- E_t ([1]): $\text{vec}\,(E_t) \sim \mathcal{N}\,(0, \Sigma_{col} \otimes \Sigma_{row})$.

[1] R. Chen, H. Xiao, and D. Yang, "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, vol. 222, no. 1, pp. 539–560, 2021.

[2] N.-J. Hsu, H.-C. Huang, and R. S. Tsay, "Matrix autoregressive spatio-temporal models," *Journal of Computational and Graphical Statistics*, vol. 30, no. 4, pp. 1143–1155, 2021.

Model

- A general additive model can be formulated as:

$$X_t = \sum_{p=1}^{P} F_p(X_{t-p}) + \sum_{q=1}^{Q} G_q(z_{t-q}) + E_t$$

- $F_p(X_{t-p})$ ([1], [2]): $F_p(X_{t-p}) = A_p X_{t-p} B_p^T$, where $A_p \in \mathbb{R}^{M\times M}, B_p \in \mathbb{R}^{N\times N}$.
- E_t ([1]): $\mathrm{vec}\,(E_t) \sim \mathcal{N}\,(0, \Sigma_{col} \otimes \Sigma_{row})$.
- $G_q(z_{t-q})$: $\sum_{d=1}^{D} z_{t-q,d} \cdot h_{q,d}, h_{q,d}(.) \in \mathcal{H}_{\mathcal{K}}$, a matrix function from an RKHS with kernel $\mathcal{K}$.

[1] R. Chen, H. Xiao, and D. Yang, "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, vol. 222, no. 1, pp. 539–560, 2021.
[2] N.-J. Hsu, H.-C. Huang, and R. S. Tsay, "Matrix autoregressive spatio-temporal models," *Journal of Computational and Graphical Statistics*, vol. 30, no. 4, pp. 1143–1155, 2021.

Model

Formally, our model is formulated as (red: parameter, blue: data):

$$X_t = \sum_{p=1}^{P} A_p X_{t-p} {B_p}^T + \sum_{q=1}^{Q} \sum_{d=1}^{D} z_{t-q,d} \cdot h_{q,d} + E_t$$

$$\text{vec}\left(E_t\right) \sim \mathcal{N}\left(0, \Sigma_{col} \otimes \Sigma_{row}\right)$$

RKHS for Matrix Function

- Each (i, j) cell of X_t has a spatial location $s_{ij} \in \mathbb{R}^2$.

RKHS for Matrix Function

- Each (i, j) cell of X_t has a spatial location $s_{ij} \in \mathbb{R}^2$.
- Kernel $\mathcal{K}(s_{ij}, s_{kl})$ defines the "similarity" of the regression coefficient at (i, j) and (k, l) for predictor z_{t-q}.

RKHS for Matrix Function

- Each (i, j) cell of X_t has a spatial location $s_{ij} \in \mathbb{R}^2$.
- Kernel $\mathcal{K}(s_{ij}, s_{kl})$ defines the "similarity" of the regression coefficient at (i, j) and (k, l) for predictor z_{t-q}.
- Example of $\mathcal{K}(.,.)$:

$$K_\eta(s_1, s_2) = \left(\frac{1}{4\pi} + \frac{\eta}{12\pi}\right) - \frac{\eta}{8\pi}\sqrt{\frac{1 - \langle s_1, s_2 \rangle}{2}}, \quad s_1, s_2 \in \mathbb{S}$$

where $\mathbb{S}$ is a unit-sphere.

RKHS for Matrix Function

- Each (i, j) cell of X_t has a spatial location $s_{ij} \in \mathbb{R}^2$.
- Kernel $\mathcal{K}(s_{ij}, s_{kl})$ defines the "similarity" of the regression coefficient at (i, j) and (k, l) for predictor z_{t-q}.
- Example of $\mathcal{K}(.,.)$:

$$K_\eta(s_1, s_2) = \left(\frac{1}{4\pi} + \frac{\eta}{12\pi}\right) - \frac{\eta}{8\pi}\sqrt{\frac{1 - \langle s_1, s_2\rangle}{2}}, \quad s_1, s_2 \in \mathbb{S}$$

  where $\mathbb{S}$ is a unit-sphere.
- In practice, we use a limited basis function to express the kernel:

$$K_\eta(s_1, s_2) \approx \frac{1}{4\pi} + \sum_{l=1}^{K^*} \frac{\eta}{(4l^2 - 1)(2l + 3)} \sum_{m=-l}^{l} Y_l^m(s_1) Y_l^m(s_2)$$

  where $Y_l^m(.) : \mathbb{R}^2 \mapsto \mathbb{R}$ are matrix basis functions.

Model Estimation

Model estimation is conducted via penalized maximum likelihood estimation (PMLE):

$$\min_{\substack{\tilde{A}_1,\tilde{A}_2,\ldots,\tilde{A}_P,\Sigma_{row}\in\mathbb{R}^{M\times M}\\ \tilde{B}_1,\tilde{B}_2,\ldots,\tilde{B}_P,\Sigma_{col}\in\mathbb{R}^{N\times N}\\ \tilde{h}_1,\tilde{h}_2,\ldots,\tilde{h}_D\in\mathcal{H}_{\mathcal{K}}}} (T-P)\log\left(|\widetilde{\Sigma_{col}}|^M|\widetilde{\Sigma_{row}}|^N\right) + \sum_{t=P+1}^{T} r_t'\left(\widetilde{\Sigma_{col}}\otimes\widetilde{\Sigma_{row}}\right)^{-1} r_t$$
$$+\lambda\sum_{d=1}^{D}\|\tilde{h}_d\|^2_{\mathcal{H}_{\mathcal{K}}}$$

where r_t is the prediction residual:

$$r_t = x_t - \sum_{p=1}^{P}\left(\tilde{B}_p\otimes\tilde{A}_p\right)x_{t-p} - \sum_{d=1}^{D} z_{t-1,d}\cdot\tilde{h}_d$$

Model Estimation

Two computational compromises are made:

- $\tilde{h}_d \in \mathcal{H}_{\mathcal{K}}$ is approximated by a truncated set of matrix basis functions: $\Psi_1, \Psi_2, \ldots, \Psi_{K^*}$:

$$\tilde{h}_d \approx \sum_{k=1}^{K^*} \tilde{b}_{d,k} \Psi_k$$

- A cyclic minimization scheme is used for the PMLE:

$$\tilde{A}_1 \to \tilde{B}_1 \to \tilde{A}_2 \to \tilde{B}_2 \to \cdots \to \tilde{A}_P \to \tilde{B}_P \to \left\{\tilde{b}_{d,k}\right\}_{d=1,k=1}^{D,K^*} \to \widetilde{\Sigma_{col}} \to \widetilde{\Sigma_{row}} \to \ldots$$

Model Estimation

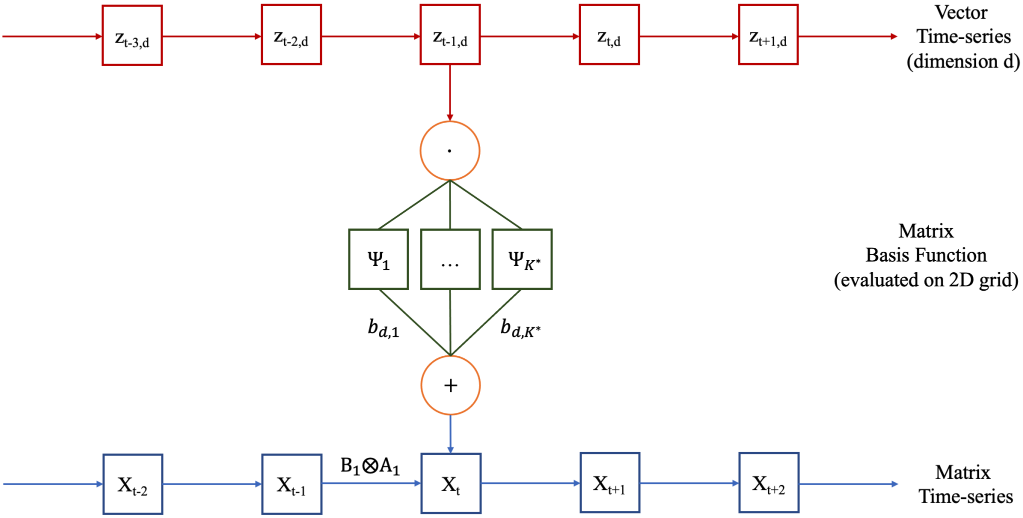


Figure: Graphical Illustration of the Model (both lags equal to 1)

Real Data Application: Spatial-Temporal Process Forecast

- We use a remote-sensing data set called the **T**otal **E**lection **C**ontent (TEC) as an application context. In this dataset, $M = 181, N = 361, T \approx 300$. There are two vector time-series covariates accompanying the data.

Real Data Application: Spatial-Temporal Process Forecast

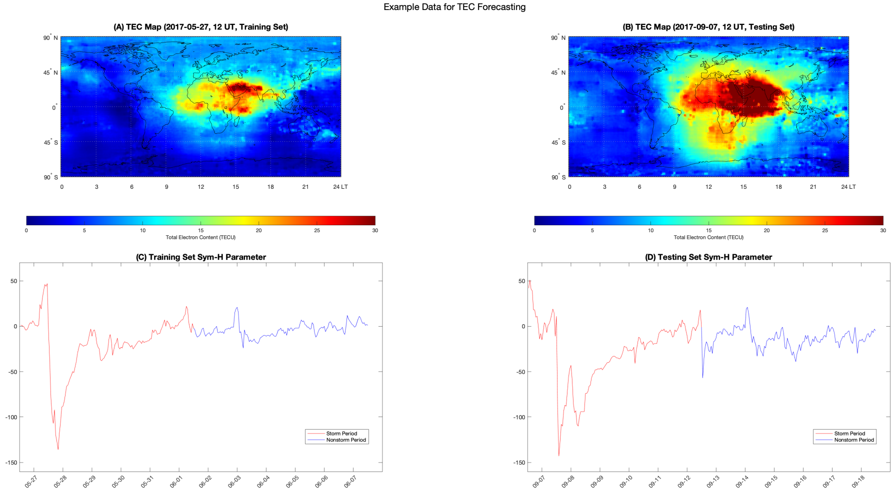


Figure: Example of the Total Electron Content (TEC) data and its auxiliary parameter Sym-H data.

Real Data Application: Spatial-Temporal Process Forecast

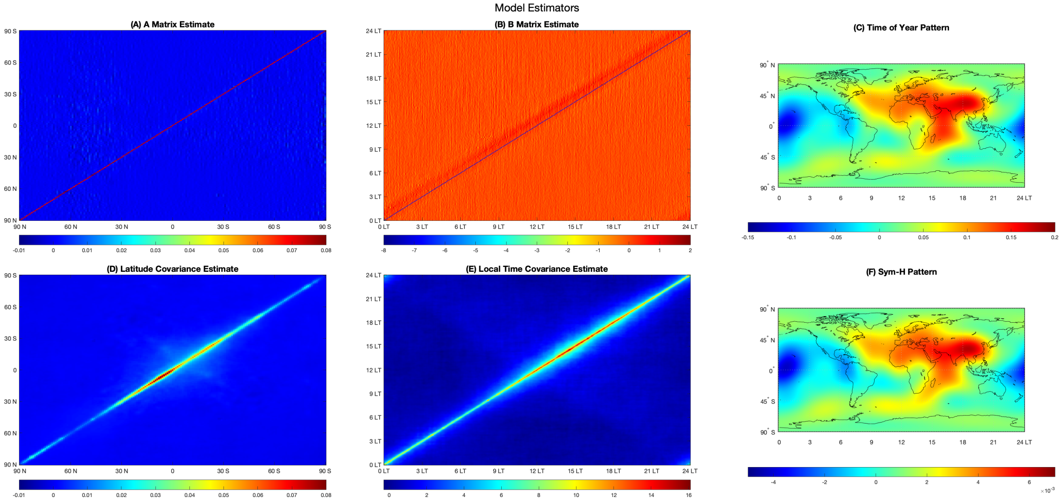


Figure: Parameter estimates of the matrix auto-regressive model with vector time-series covariates.

Real Data Application: Spatial-Temporal Process Forecast

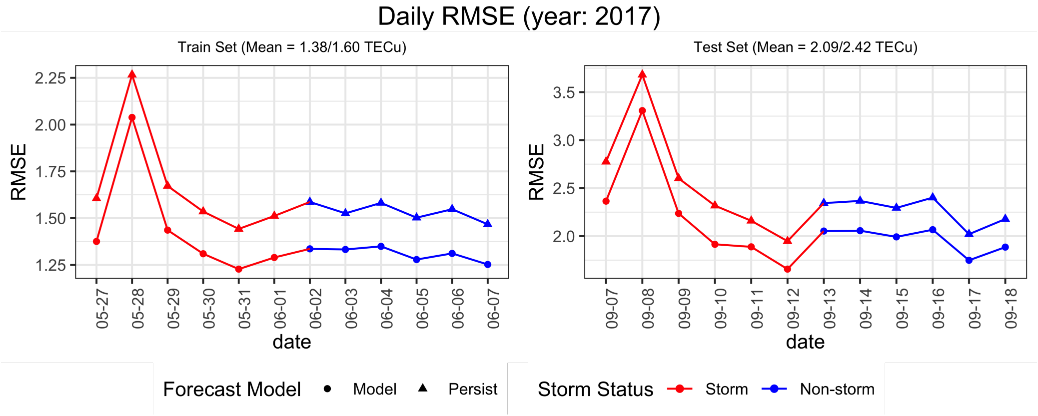


Figure: Daily Rooted Mean Squared Error (RMSE) for forecasting the TEC. Left panel shows the training set RMSE and right panel shows the testing set RMSE.