Matrix Autoregressive Model with Vector Time Series Covariates for Spatio-Temporal Data

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

Background: Multi-Modality Time Series Joint Modeling

In this paper, we investigate a matrix time series autoregression problem where we observe:

[Spatio-Temporal Matrix Time Series]: $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$ and each $\mathbf{X}_t$ is a 2D spatial data collected on an $M \times N$ grid $\mathcal{S}$. So $\mathbf{X}_t(i,j)$ is *local data* at location (i,j) and time t.

[Auxiliary Vector Time Series]: $\mathbf{z}_1, \ldots, \mathbf{z}_T \in \mathbb{R}^D$, and each $\mathbf{z}_t$ is *global data* shared across $\mathcal{S}$ at t.

The autoregression problem is trying to model $\mathrm{E}[\mathbf{X}_t | (\mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-P}), (\mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-Q})]$. A motivating example regarding a space weather real data application is shown in Figure 1:

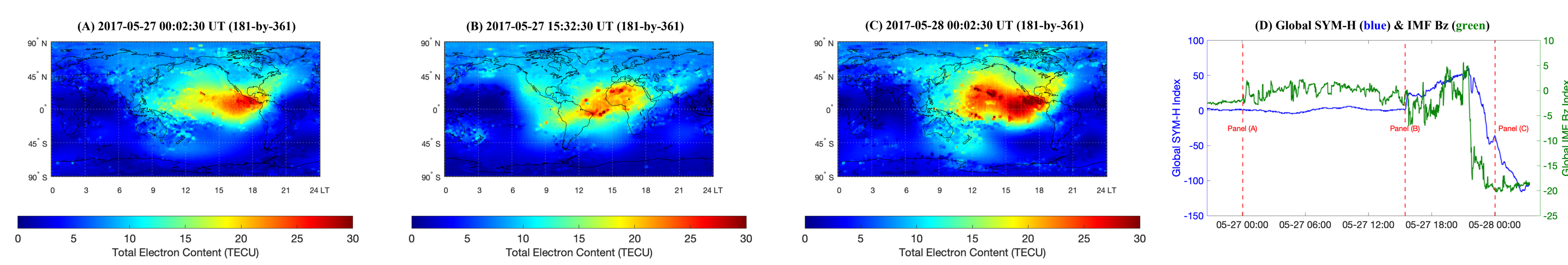


Figure 1. (A)-(C) 181×361 Earth's Total Electron Content (TEC) Matrix Time Series; (D) 2-dimensional Solar Wind Parameters Vector Time Series. The forecasting problem is to forecast (C) with (A), (B) and (D).

Existing Work 1 (*Matrix Autoregression or MAR*):

$$\mathbf{X}_t = \sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top + \mathbf{E}_t, \quad \mathbf{vec}(\mathbf{E}_t) \overset{i.i.d.}{\sim} N(\mathbf{0}_{MN}, \Sigma), \Sigma = \overbrace{\Sigma_c \otimes \Sigma_r}^{\text{Multi-Way Covariance}}. \tag{1}$$

Existing Work 2 (*Spatio-Temporal MAR* with Fixed-Rank Co-kriging):

$$\mathbf{X}_t = \sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top + \mathbf{E}_t, \quad \mathbf{vec}(\mathbf{E}_t) \overset{i.i.d.}{\sim} N(\mathbf{0}_{MN}, \Sigma), \Sigma = \sigma^2 \mathbf{I} + \overbrace{\mathbf{FMF}^\top}^{\text{rank}-k,\ k << MN}, \tag{2}$$

where $\mathbf{F} \in \mathbb{R}^{MN \times k}$ contains k spatial basis functions and $\mathbf{M} \in \mathbb{R}^{k \times k}$ is a co-kriging parameter. There is no existing work that can incorporate *vector predictors* under *MAR*.

Our Model: Matrix Autoregression with Auxiliary Covariates (MARAC)

$$\mathbf{X}_t = \sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top + \sum_{q=1}^{Q} \mathcal{G}_q \times_3 \mathbf{z}_{t-q}^\top + \mathbf{E}_t, \quad \mathbf{vec}(\mathbf{E}_t) \overset{i.i.d.}{\sim} N(\mathbf{0}_{MN}, \Sigma_c \otimes \Sigma_r), \tag{3}$$

where $\mathcal{G}_q \in \mathbb{R}^{M \times N \times D}$ and $\times_3$ is the mode-3 tensor-matrix product. Element-wisely,

$$\mathbf{X}_t(i,j) = \underbrace{\sum_{p=1}^{P} \left\langle \mathbf{A}_p(i,:)^\top \mathbf{B}_p(j,:), \mathbf{X}_{t-p} \right\rangle}_{\text{Matrix TS Autoregression}} + \underbrace{\sum_{q=1}^{Q} \mathcal{G}_q(i,j,:)^\top \mathbf{z}_{t-q}}_{\text{Vector TS Local Linear Model}} + \mathbf{E}_t(i,j), \quad (i,j) \in \mathcal{S} \tag{4}$$

$\mathbf{X}_{t-p}$ is *local data* (unique to each spatial location), but has *global* parameters $\mathbf{A}_p, \mathbf{B}_p$;
$\mathbf{z}_{t-q}$ is *global data* (shared across spatial domain), but has *local* parameters $\mathcal{G}_q$.

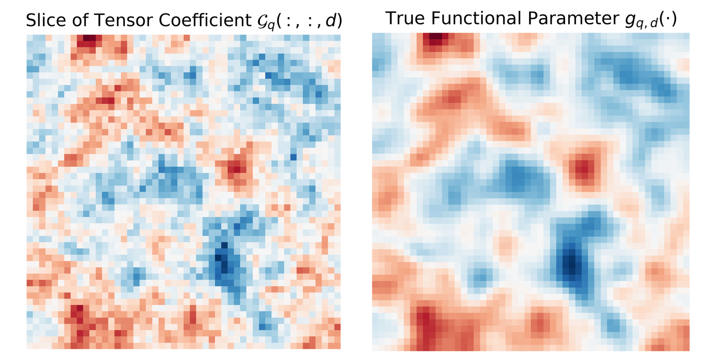
If one re-groups the last two terms in (3) as a new error matrix time series $\widetilde{\mathbf{E}}_t$:

$$\mathrm{Var}\left[\mathbf{vec}\left(\sum_{q=1}^{Q} \mathcal{G}_q \times_3 \mathbf{z}_{t-q}^\top + \mathbf{E}_t\right)\right] = \mathrm{Var}\left[\mathbf{vec}\left(\widetilde{\mathbf{E}}_t\right)\right] = \underbrace{\Sigma_c \otimes \Sigma_r}_{\text{temporally-independent spatial covariance}} + \overbrace{\mathbf{FMF}^\top}^{\text{temporally-dependent low-rank covariance}}, \tag{5}$$

where $\mathbf{F} = [\mathbf{G}_1^\top \ldots \mathbf{G}_Q^\top]$, $\mathbf{G}_q \in \mathbb{R}^{MN \times D}$ is obtained via unfolding $\mathcal{G}_q$ on mode-3 and $\mathbf{M} = [\mathrm{Cov}(\mathbf{z}_{t-k}, \mathbf{z}_{t-l})]_{1 \le k, l \le Q}$ is the covariance matrix of the auxiliary vector time series.

Computational Algorithm: Alternating Penalized MLE

We make an assumption that the coefficient tensors $\mathcal{G}_1, \ldots, \mathcal{G}_Q$ are spatially-smooth:



- $(i,j) \in \mathcal{S} \longleftrightarrow s_{ij} \in [0,1]^2$
- $\mathcal{G}_q(i,j,d)$: tensor coefficient at (i,j) for the d^{th} covariate at lag q
- $g_{q,d}(s_{ij})$: functional parameter evaluated at s_{ij} for the d^{th} covariate at lag q
- $\mathcal{G}_q(i,j,d) = g_{q,d}(s_{ij}), \forall (i,j) \in \mathcal{S}$
- $g_{q,d} \in \mathcal{H}_K, \forall q \in [Q], d \in [D]$, where $\mathcal{H}_K$ is an RKHS
- $K(\cdot,\cdot) : [0,1]^2 \times [0,1]^2 \mapsto \mathbf{R}$ is a spatial kernel function

We estimate the model parameters via penalized maximum likelihood estimation (PMLE):

$$\widehat{\Theta} = \underset{\substack{\{\mathbf{A}_p, \mathbf{B}_p\}_{p=1}^P, \Sigma_r, \Sigma_c \\ \{g_{q,d}\}_{q=1,d=1}^{Q,D}}}{\arg\min} \underbrace{-\frac{1}{T} \sum_{t=1}^{T} \ell\left(\mathbf{X}_t | \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-P}, \mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-Q}, \Theta\right)}_{\text{Matrix TS Residual Negative Log-Likelihood}} + \underbrace{\lambda \sum_{q=1}^{Q} \sum_{d=1}^{D} \|g_{q,d}\|_{\mathcal{H}_K}^2}_{\text{Tensor Smoothness Penalty}} \tag{6}$$

Given $\lambda > 0$, we can convert the infinite-dimensional optimization problem above into a finite-dimensional problem (the Representer Theorem):

$$\widehat{\Theta} = \underset{\substack{\{\mathbf{A}_p, \mathbf{B}_p\}_{p=1}^P, \Sigma_r, \Sigma_c \\ \{\gamma_{q,d}\}_{q=1,d=1}^{Q,D}}}{\arg\min} -\frac{1}{T} \sum_{t=1}^{T} \ell\left(\mathbf{X}_t | \mathbf{X}_{t-1}, \ldots, \mathbf{X}_{t-P}, \mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-Q}, \Theta\right) + \underbrace{\lambda \sum_{q=1}^{Q} \sum_{d=1}^{D} \gamma_{q,d}^\top \mathbf{K} \gamma_{q,d}}_{\text{Kernel Ridge Penalty}},$$

where $\mathbf{K} \in \mathbb{R}^{MN \times MN}$ is the kernel Gram matrix based on the spatial kernel function $K(\cdot,\cdot)$.

We update the parameters one at a time until convergence. Updating $\mathbf{A}_p$ becomes:

$$\mathbf{A}_p \leftarrow \left[\sum_t \widetilde{\mathbf{X}}_{t,-p} \Sigma_c^{-1} \mathbf{B}_p \mathbf{X}_{t-p}^\top\right] \left[\sum_t \mathbf{X}_{t-p} \mathbf{B}_p^\top \Sigma_c^{-1} \mathbf{B}_p \mathbf{X}_{t-p}^\top\right]^{-1}, \tag{7}$$

where $\widetilde{\mathbf{X}}_{t,-p}$ is the running residual matrix without the lag-p autoregressive term. For $\gamma_q = [\gamma_{q,1}, \cdots, \gamma_{q,D}] \in \mathbb{R}^{MN \times D}$, the optimization is equivalent to a kernel ridge regression:

$$\mathbf{vec}(\gamma_q) \leftarrow \left[\left(\sum_{t=1}^{T} \mathbf{z}_{t-q} \mathbf{z}_{t-q}^\top\right) \otimes \Sigma^{-1} \mathbf{K} + \lambda T^2 \mathbf{I}\right]^{-1} \left[\sum_{t=1}^{T} \left(\mathbf{z}_{t-q} \otimes \Sigma^{-1}\right) \mathbf{vec}\left(\widetilde{\mathbf{X}}_{t,-q}\right)\right], \tag{8}$$

where $\widetilde{\mathbf{X}}_{t,-q}$ is the running residual matrix without the lag-q auxiliary covariate term.

Theory: MARAC Estimators Asymptotics

Main Result 1: Finite-Dimensional Asymptotics Given fixed M, N and $\lambda = o(T^{-1/2})$, and assume that $\{\mathbf{X}_t\}_{t=1}^T$ is generated by MARAC in (3), and $\{\mathbf{z}_t\}_{t=1}^T$ is a covariance-stationary time series, then the alternating PMLE estimator of the MARAC model is asymptotically normal:

$$\sqrt{T}\left[\mathbf{vec}(\widehat{\Theta} - \Theta^*)\right] \Longrightarrow N(\mathbf{0}, \Xi),$$

where Θ^* contains all the true values of $\mathbf{A}_p, \mathbf{B}_p$ and all γ_q, and $\widehat{\Theta}$ is its estimator.

Main Result 2: High-Dimensional Asymptotics If $MN \to \infty$ as $T \to \infty$, and assume that the spatial kernel function $K(\cdot,\cdot)$ bears a Mercer decomposition $K(\cdot,\cdot) = \sum_s \lambda_s \phi_s(\cdot)\phi_s(\cdot)$, where $\lambda_s \sim s^{-r}, r > 1$, then under additional mild regularity conditions, the autoregressive coefficient $\Phi = [\mathbf{B}_1 \otimes \mathbf{A}_1 \ldots \mathbf{B}_P \otimes \mathbf{A}_P]$ has element-wise estimation error bound at $O_P(1/\sqrt{MN})$, or equivalently:

$$\|\mathbf{vec}\left(\widehat{\Phi} - \Phi^*\right)\| \lesssim \sqrt{MN}, \tag{9}$$

with high probability. This is different from the finite-dimensionality result (i.e. $O_P(1/\sqrt{T})$).

Simulation Study

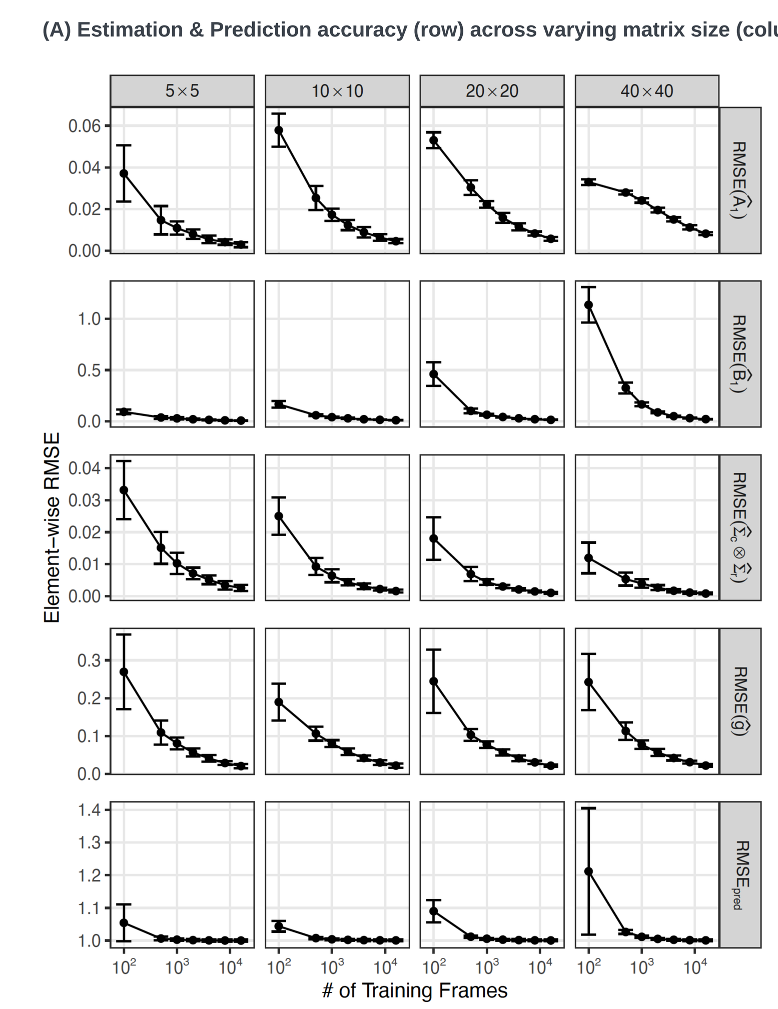
In our simulation study, we conduct four sets of experiments:

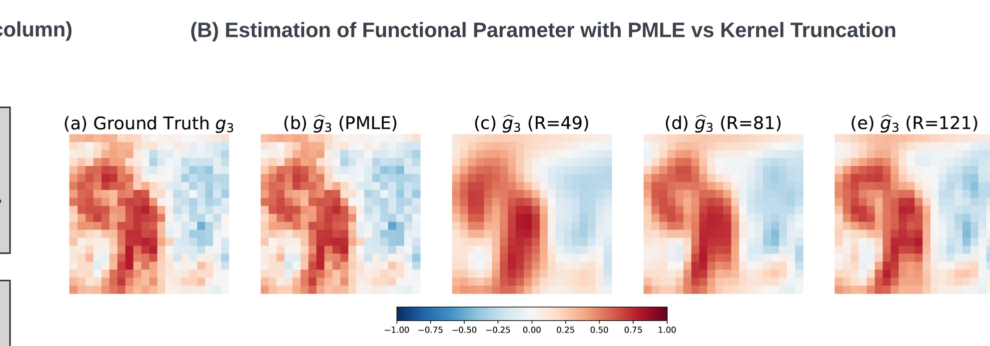
(A) Estimation Validation: check the accuracy of the parameter estimators and predictions.

(B) Fast Computation with Kernel Truncation: instead of estimating $\mathcal{G}_q$ with kernel ridge regression as in (8), we use a series of $R \in \{49, 81, 121\}$ basis functions to approximate $\mathcal{G}_q$, which speeds up the computation a lot in high-dimensional settings at the cost of accuracy.

(C) Lag Selection: check the consistency of selecting the correct lag with AIC & BIC.

(D) Method Comparison: compare MARAC model with competing methods on a prediction task.





(C) Lag Selection Consistency with AIC & BIC

	$T = 1000$	$T = 2000$	$T = 4000$	$T = 8000$
AIC	(.54, .99, .53)	(.55, .97, .53)	(.59, .96, .55)	(.65, .94, .59)
BIC	(1.00, .09, .09)	(.99, .56, .56)	(.97, .97, .94)	(.96, .99, .95)

Table 1: Probability that AIC and BIC select the correct P (first number), Q (second number) and (P, Q) (third number) from 100 repeated runs.

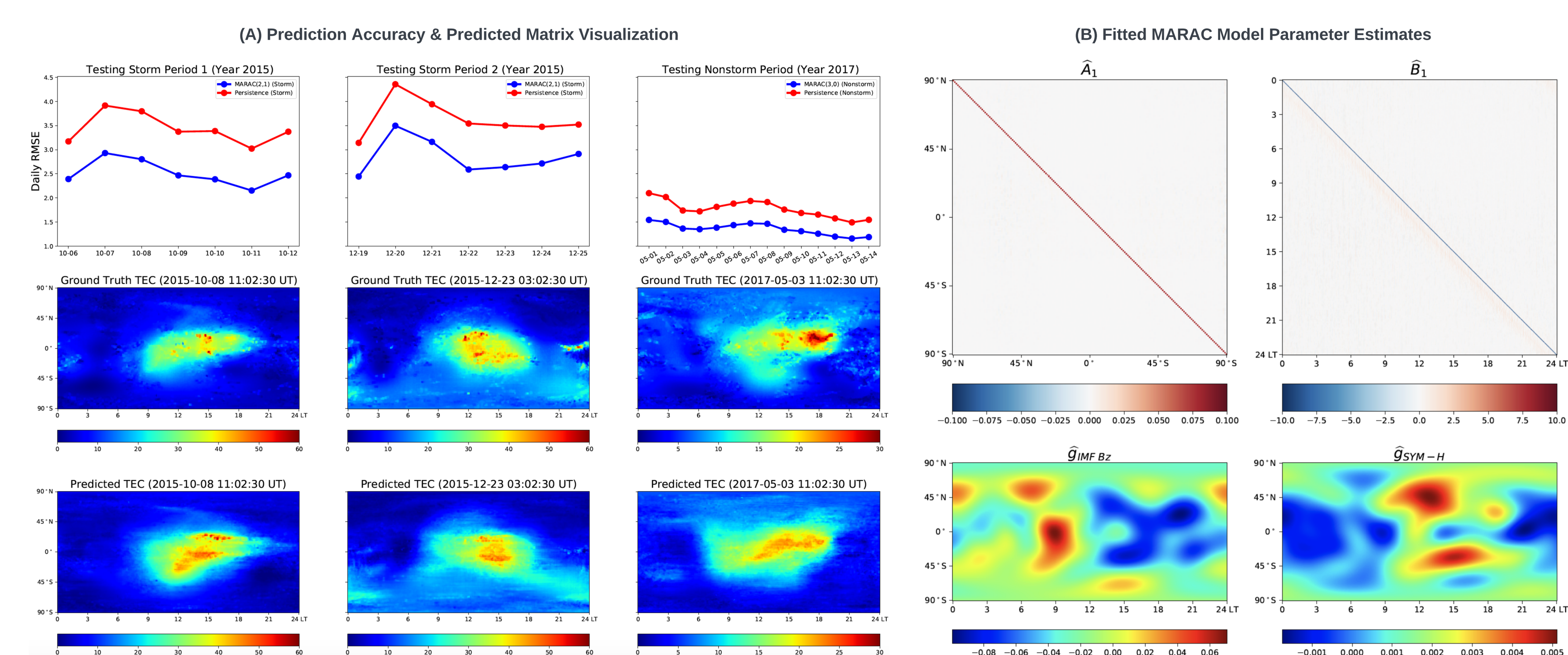
(D) Prediction Accuracy Comparison across Competing Methods

Matrix Size	MARAC(1, 1)	MAR(1)	PixelARX(1, 1)	VARX(1, 1)
5×5	**1.000±0.008**	1.031±0.024	1.143±0.029	**1.007±0.009**
10×10	**1.002±0.006**	1.028±0.019	1.101±0.033	1.027±0.006
20×20	**1.002±0.004**	1.028±0.017	1.112±0.038	1.120±0.006
40×40	**1.005±0.005**	1.043±0.020	1.118±0.042	2.253±0.025

Table 2: Testing set ($T_{\text{test}} = 2000$) RMSE across competing methods. Optimal prediction RMSE should be at 1.0 given our setup. Experiments repeated for 20 times.

Real Data Application: Global TEC Forecast

In real data application, we consider the problem of forecasting global Total Electron Contents (TEC) with solar wind parameters as the auxiliary time series, as detailed in Figure 1:



[1] **Sun, H.**, Hua, Z., Ren, J., Zou, S., Sun, Y., & Chen, Y. (2022). Matrix Completion Methods for the Total Electron Content Video Reconstruction. *The Annals of Applied Statistics*, 16(3), 1333-1358.
[2] **Sun, H.**, Chen, Y., Zou, S., Ren, J., Chang, Y., Wang, Z. & Coster, A. (2023) Complete Global Total Electron Content Map Dataset based on a Video Imputation Algorithm VISTA. *Scientific Data*, 10(1), 236.
[3] **Sun, H.**, Shang, Z. & Chen, Y. (2023) Matrix Autoregressive Model with Vector Time Series Covariates for Spatio-Temporal Data. *Submitted*.