Variational Propagation for Exact Spatiotemporal Dynamic Modeling

Luca Presicce

Ph.D. student in Statistics

University of Milan-Bicocca, Department of Economics, Management & Statistics

SIS 2025 - Statistics for Innovation, June 16-18, 2025, Genoa (Italy)

Introduction & Motivation

How it started

This is a preliminary work in collaboration with

Sudipto Banerjee

University of California, Los Angeles

- L. Presicce, S. Banerjee (2025+) **“Adaptive Markovian Spatiotemporal Propagation in Multivariate Bayesian Modeling”**, In Preparation.
- L. Presicce, S. Banerjee (2025) **“Bayesian Transfer Learning for Artificially Intelligent Geospatial Systems: A Predictive Stacking Approach”**, Under Review.
- S. Banerjee (2020) **“Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework”**, Spatial Statistics, vol. 37.

Motivation: Data-Rich Spatiotemporal Environment

Spatiotemporal phenomena are pervasive in many research areas (e.g., Environmental and climate sciences, Biomedical applications, Epidemiology and health analytics, Remote sensing and geostatistics)

Facing these problems main challenges arise:

- ☞ High-dimensional observations over time
- ☞ Complex spatial and temporal dependencies
- ☞ Provide inferences in real time

☛ **Key point:** Necessity for spatiotemporal models offering on-demand inferences and predictions for large-scale online frameworks.

(Why) Existing Methods (Struggle)

Dynamic linear models (DLMs) offer a convenient framework [6, 3], along with the forward filtering backward sampling [2] algorithm (FFBS).

Conjugacy is lost when incorporating spatiotemporal covariance structures, as introduce (weakly identifiable) non-conjugate parameters.

Classical simulation-based or iterative algorithms are computationally intensive: infeasible for large-scale real-time tasks.

Most contributions focus on empirical Bayes [9], stochastic differential equations [5], or other iterative strategies as INLA [8], which may require strong prior assumptions [5].

Proposal & Model Overview

Chasing - exact, full - conjugacy

We propose a scalable online learning framework using variational propagation to restore exact conjugacy.

- ☞ Encode time-evolving latent states and multivariate spatial dependencies using Matrix-variate DLM.
- ☞ Leverage Bayesian Predictive Stacking (BPS) to combine multiple models with different spatial parameters.
- ☞ Derive Variational approximation to recover full conjugacy (and scalability) by projecting mixtures of posteriors into conjugate families.

Let $Y = \{ Y_t \,:\, t \in \mathcal{T} \subset \mathbb{N} \}$ be a spatiotemporal tensor of outcomes: Y_t $(n \times q)$ matrix with n fixed locations $\mathcal{S} = \{s_1, \dots, s_n\}$, for q correlated outcomes.

A matrix-variate dynamic linear model can be represented as:

$$\begin{aligned} Y_t &= F_t \Theta_t + \Upsilon_t, \qquad & \Upsilon_t &\sim \mathrm{MN}(0, V_t, \Sigma) \\ \Theta_t &= G_t \Theta_{t-1} + \Xi_t, \qquad & \Xi_t &\sim \mathrm{MN}(0, W_t, \Sigma). \end{aligned} \tag{1}$$

Consider now the following reparameterization

- ☞ $\Theta_t = [B_t^\top : \Omega_t^\top]^\top$: regression coefficients and latent spatial process
- ☞ $V_t = V_t(\alpha) = \frac{1-\alpha}{\alpha} \mathbb{I}_n$ introducing discontinuity with proportion of spatial variability α
- ☞ $W_t = W_t(\phi)$ includes spatial kernel $R_t(S, S; \phi)$ $(B_t \perp\!\!\!\perp \Omega_t \ \forall t$ → W_t block-diagonal).

We can cast (1) as multivariate autoregressive latent spatial regression

$$\begin{aligned} Y_t \mid B_t, \Omega_t, \Sigma &\sim \mathrm{MN}(X_t B_t + \Omega_t,\ (\alpha - 1)^{-1}\, \mathbb{I}_n,\ \Sigma) \\ B_t \mid \Sigma &\sim \mathrm{MN}(B_{t-1},\ W_t^{(B)},\ \Sigma) \\ \Omega_t \mid \Sigma &\sim \mathrm{MN}(\Omega_{t-1},\ R_t(\mathcal{S}, \mathcal{S}; \phi),\ \Sigma), \end{aligned} \tag{2}$$

with prior information on state matrix, and common column covariance matrix defined as

- ☞ $[B_0^\top : \Omega_0^\top]^\top = \Theta_0 \mid \Sigma \sim \mathrm{MN}(m_0, C_0, \Sigma)$
- ☞ $\Sigma \sim \mathrm{IW}(\nu_0, \Psi_0)$

where m_0, C_0, ν_0, and Ψ_0 are considered known quantities.

Methodological Details

Conjugacy of Matrix-Normal-Inverse-Wishart family within FFBS

Here FFBS provides convenient conjugate framework to propagate posteriors through time:

- Given a fixed couple $\{\alpha, \phi\}$, Model (2) is fully conjugate

Posterior and posterior predictive are available in closed form as MNIW and matrix-variate Student's t distributions [1, 7].

Avoiding simulation-based or iterative approaches to evaluate non-conjugate parameters for dynamic spatiotemporal models.

- **Key point:** combinations of $\{\alpha_j, \phi_j\}$ (characterizing different models $\mathscr{M}_j$, for $j = 1, \dots, J$) yield distinct but tractable posteriors.

Collect inferences - (Dynamic) Bayesian Predictive Stacking of Predictive Densities

BPS of predictive densities assimilates models using a weighted distribution in the convex hull, $C_t = \left\{\sum_{j=1}^{J} w_{t,j}\, p(\cdot \mid Y_{1:t-1}, \mathscr{M}_j) : \sum_j w_{t,j} = 1,\; w_{t,j} \geq 0\right\}$, of individual posterior distributions by maximizing the logarithm score [4, 10] to fetch

$$\hat{w}_t = (\hat{w}_{t,1}, \ldots, \hat{w}_{t,J})^\top = \arg\max_{w_t \in S_1^J} \frac{1}{n} \sum_{i=1}^{n} \log \sum_{j=1}^{J} w_{t,j}\, p\left(Y_{t,i} \mid Y_{1:t-1}, \mathscr{M}_j\right) \;, \tag{3}$$

where any 1-step-ahead predictive $p(\,\cdot \mid Y_{1:t-1}, \mathscr{M}_j)$ available in closed-form, and each model $\mathscr{M}_j$ corresponds to fixed couple $\{\alpha_j, \phi_j\}$.

Solving (3) minimizes the Kullback-Leibler divergence from the true 1-step-ahead predictive distribution: since unknown, we use leave-future-out (LFO) to estimate the expected value of the score [10].

Challenges & Solutions

All that glitters is not gold - BPS breaks online conjugacy

Once obtained $\hat{w}_t$, posterior inference follow by stacked posterior distributions:

$$\hat{p}(\cdot \mid Y_{1:t}) = \sum_{j=1}^{J} \hat{w}_{t,j}\, p(\,\cdot \mid Y_{1:t}, \mathscr{M}_j), \tag{4}$$

Stacked posteriors $\hat{p}(\Theta_t, \Sigma \mid Y_{1:t}) = \sum_{j=1}^{J} \hat{w}_{t,j}\, p(\Theta_t, \Sigma \mid Y_{1:t}, \mathscr{M}_j)$ are mixtures of MNIW distributions, no longer belonging conjugate families

Leading to non-conjugate posterior-to-prior update: we cannot propagate to future time point using FFBS machinery again.

☛ **Solution:** use variational approach to find the MNIW distribution that minimize KL divergence from stacked posterior.

We obtain the variational approximating posterior distribution

$$\hat{p}_{KL}(\Theta_t, \Sigma | Y_{1:t}) = \mathrm{MNIW}(\Theta_t, \Sigma \mid \tilde{m}_t, \tilde{C}_t, \tilde{\Psi}_t, \tilde{\nu}_t) \tag{5}$$

with $\tilde{m}_t = \sum w_j m_t^{(j)}$, $\tilde{C}_t = \sum w_j \left(C_t^{(j)} + (m_t^{(j)} - \tilde{m}_t)^\top \Sigma^{-1} (m_t^{(j)} - \tilde{m}_t) \right)$, $\tilde{\Psi}_t = \tilde{\nu}_t \left[\sum_{j=1}^J \hat{w}_j \nu_t^{(j)} \Psi_t^{-1(j)} \right]^{-1}$

$\nu_t^{(j)} = \nu_{t-1}^{(j)} + \frac{n}{2}$ is constant across models $\mathscr{M}_j$, defining $\tilde{\nu}_t$ as $\tilde{\nu}_t = \sum_j \hat{w}_j \nu_t^{(j)}$ allow direct computation of $\tilde{\Psi}_t$ otherwise not possible.

Variational posterior in (5) belongs to MNIW family, restoring exact temporal posterior-to-prior conjugate update.

👉 Using $\hat{p}_{KL}(\Theta_t, \Sigma \mid Y_{1:t})$ instead of (4) permits conjugate online propagation to future time point with FFBS machinery.

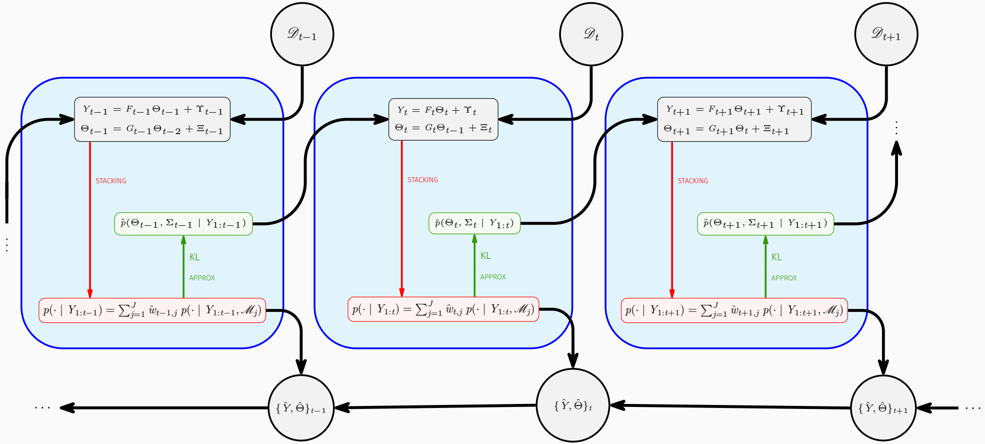

Figure 1: Temporal datasets dynamic propagation representation

Empirical Results

Establish effectiveness using synthetic generated data from model in Equation (2)

We consider $n = 300$ fixed locations, over $T = 20$ time instants (6000 multivariate observations), for $q = 3$ correlated outcomes, and $p = 4$ predictors.

True parameters set as $\Sigma = \begin{bmatrix} 1 & -0.3 & 0.6 \\ -0.3 & 1.2 & 0.4 \\ 0.6 & 0.4 & 1 \end{bmatrix}$, $\alpha = 0.8$, $\phi = 4$ (exponential spatial kernel), state matrix initialized at $\Theta_0 = 0_{(p+n)\times q}$.

Implementing BPS uses $J = 9$ models: $\alpha \in \{0.65, 0.8, 0.95\}$, $\phi \in \{2, 4, 6\}$

Simulation Results - Coefficient Dynamics

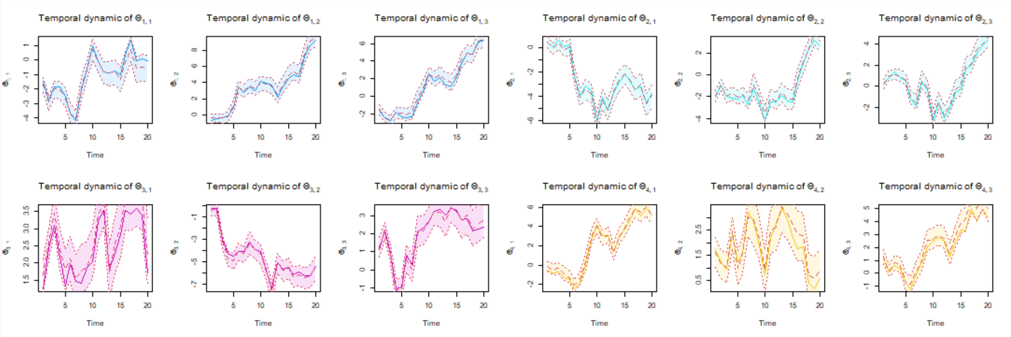

Figure 2: Regression coefficients dynamics: true (solid line), map (dashed line), and 95% credible interval (shade).

- Strong tracking of true dynamics for regression coefficient and spatial process.
- Credible intervals (95%) show excellent calibration with 95.28% empirical coverage.

Simulation Results - Temporal Forecasting

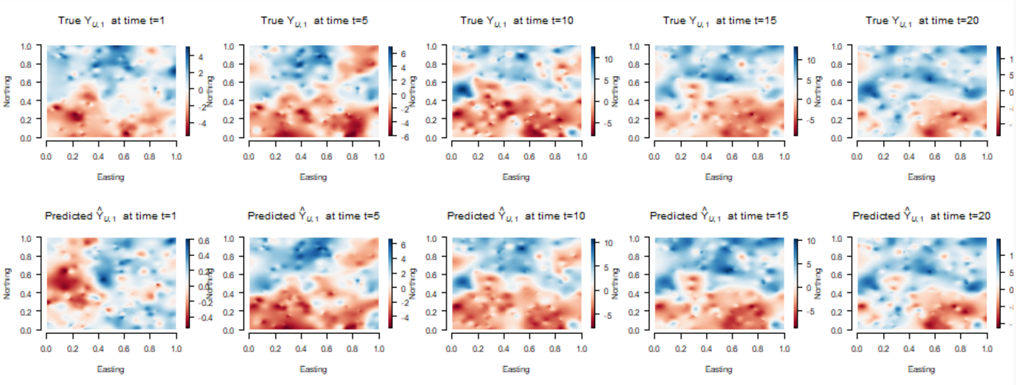

Figure 3: Temporal forecast for outcome 1 at selected time points: true (top row), and predicted surfaces (bottom row).

- Rapid temporal dynamics learning from time $t = 2$ onward.

Simulation Results - Spatial Interpolation

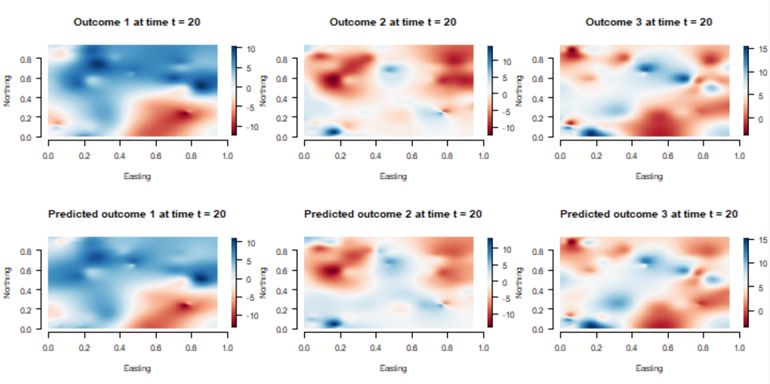

Figure 4: Spatial interpolation at unobserved points for t=20: true (top row), and predicted surfaces (bottom row).

- Spatial interpolations indistinguishable from raw truth.

Conclusions

Exact conjugate inference restored for dynamic spatiotemporal models, avoiding simulation-based algorithms or strong prior information.

- BPS permits within-time point conjugacy → parallel learning
- KL approximation permits between-time point conjugacy → sequential learning

variational approximation is only used to propagate information across time, while stacked posteriors are used for accurate dynamic inferences and predictions.

Simulation experiments show strong empirical performance and computationally efficient online learning.

spFFBS: Fully Conjugate Matrix-variate DLMs for Spatiotemporal modeling

Working in progress → spFFBS R package:

- Introduce an easy framework to fully conjugate matrix-variate DLMS framework for spatiotemporal geostatistical modeling.

- Use Rcpp/C++ -based code, allowing faster and scalable parallel-sequential computations for dynamic spatiotemporal model (2).

- Available on Github @lucapresicce/spFFBS (hopefully sooon on CRAN).

Check it out on my GitHub!

References

[1] Banerjee, S. (2020). Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework. Spatial Statistics 37, 100417.

[2] Carter, C. K. and R. Kohn (1994, September). On Gibbs sampling for state space models. Biometrika 81(3), 541–553. _eprint: https://academic.oup.com/biomet/article-pdf/81/3/541/714321/81-3-541.pdf.

[3] Gamerman, D., H. F. Lopes, and E. Salazar (2008). Spatial dynamic factor analysis. Bayesian Analysis 3(4), 759 – 792. Publisher: International Society for Bayesian Analysis.

[4] Gneiting, T. and A. E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association 102(477), 359–378. Publisher: Taylor & Francis.

[5] Leeds, W. B. and C. K. Wikle (2012). Science-based parameterizations for dynamical spatiotemporal models. WIREs Computational Statistics 4(6), 554–560.

[6] Nobre, A. A., A. M. Schmidt, and H. F. Lopes (2005, May). Spatio-temporal models for mapping the incidence of malaria in Pará. Environmetrics 16(3), 291–304. Publisher: John Wiley and Sons Ltd.

[7] Presicce, L. and S. Banerjee (2024). Bayesian Transfer Learning for Artificially Intelligent Geospatial Systems: A Predictive Stacking Approach. arXiv preprint, arXiv:2410.09504. Publisher: Cornell University _eprint: stat.ME/2410.09504.

[8] Rue, H., S. Martino, and N. Chopin (2009, April). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. Journal of the Royal Statistical Society Series B: Statistical Methodology 71(2), 319–392.

[9] Xu, K. and C. K. Wikle (2007). Estimation of parameterized spatio-temporal dynamic models. Journal of Statistical Planning and Inference 137(2), 567–588.

[10] Yao, Y., A. Vehtari, D. Simpson, and A. Gelman (2018). Using Stacking to Average Bayesian Predictive Distributions (with Discussion). Bayesian Analysis 13(3), 917–1007. Publisher: International Society for Bayesian Analysis.

Thanks for your attention!