Conjugate Bayesian Modeling and Inference In High-dimensional Spatial Statistics: Conquering New Challenges

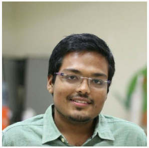
Sudipto Banerjee

SIAM, February 24th, 2022

Department of Biostatistics UCLA

Collaborators

Abhirup Datta (JHU)



Debangan Dey (JHU)

Barbara Engelhardt (Princeton)

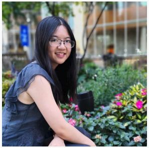
Andrew Finley (MSU)

Rajarshi Guhaniyogi (TAMU)

Andrew Jones (Princeton)

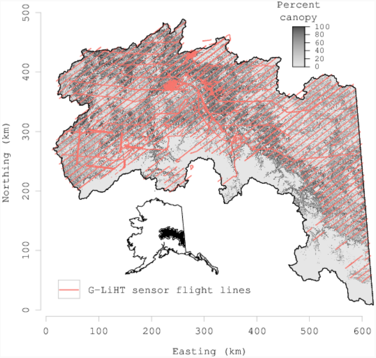
Didong Li (Princeton/UCLA)

Michele Peruzzi (Duke)

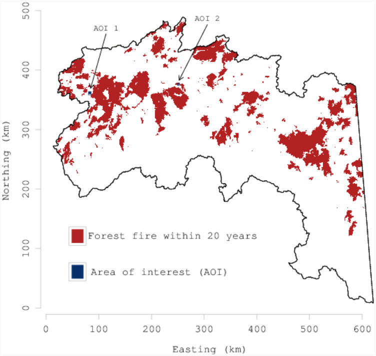


Lu Zhang (Columbia)

Example: Alaska Tanana Valley Forest Height Dataset (FD-CMAB, *JCGS*, 2019)



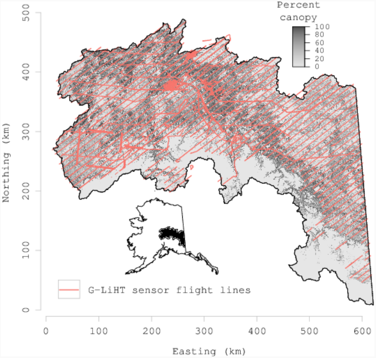
Forest height and tree cover



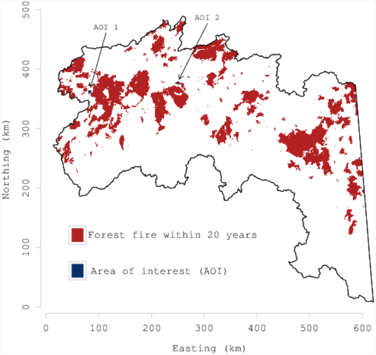
Forest fire history

- Forest height (red lines) data from LiDAR at 10×10^6 locations
- Knowledge of forest height is important for biomass assessment, carbon management etc

Example: Alaska Tanana Valley Forest Height Dataset (FD-CMAB, *JCGS*, 2019)



Forest height and tree cover



Forest fire history

- Goal: High-resolution domain-wide prediction maps of forest height
- Covariates: Domain-wide tree cover (grey) and forest fire history (red patches) in the last 20 years

Analyzing the data

Models used:

- Non-spatial regression: $y_{FH} = \beta_0 + \beta_{tree} x_{tree} + \beta_{fire} x_{fire} + \epsilon$

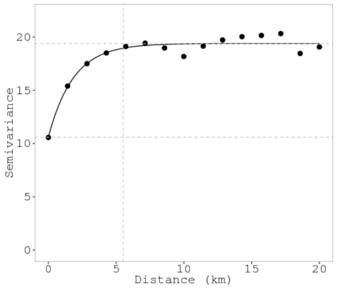


Figure: Variogram (defined as $\mathrm{var}\{Z(\ell + h) - Z(\ell)\}$) of the residuals from non-spatial regression indicates strong spatial pattern

Bayesian regression for BIG DATA

- Conjugate Bayesian hierarchical linear model:

$$y_i \mid \beta, \sigma^2 \overset{ind}{\sim} N(x_i^\top \beta, \sigma^2)\ ,\ i = 1, 2, \ldots, n\ ;$$
$$\beta \mid \sigma^2 \sim N(\mu_\beta, \sigma^2 V_\beta)\ ; \quad \sigma^2 \sim IG(a, b)\ .$$

- Exact Bayesian inference:

$$\sigma^2 \mid y \sim IG(a^*, b^*) \quad \beta \mid \sigma^2, y \sim N(Mm, \sigma^2 M)\ , \quad \text{where}$$
$$m = V_\beta^{-1}\mu_\beta + X^\top y\ , \quad M^{-1} = V_\beta^{-1} + X^\top X\ ,$$
$$a^* = a + n/2\ , \quad b^* = \mu_\beta^\top V_\beta^{-1}\mu_\beta + y^\top y - m^\top M^{-1} m\ .$$

- What if the data cannot be stored/loaded into available workspace?

- HADOOP: Map-Reduce framework (Divide & Conquer) with cloud computing.

Bayesian regression on HADOOP

- Partition data as $D_k = \{y_k, X_k\}$, $k = 1, 2, \ldots, K$, where each y_k is $n_k \times 1$, X_k is $n_k \times p$ and $N = \sum_{k=1}^{K} n_k$.

- Sequential ("streaming") updates:

$$p(\beta, \sigma^2 \mid D_1, \ldots, D_{k+1}) \propto p(\beta, \sigma^2 \mid D_1, \ldots, D_k) \times p(D_{k+1} \mid \beta, \sigma^2)$$

- Parallel architecture: compute simultaneously...

$$m_k = V_\beta^{-1} + X_k^\top y_k \text{ and } M_k^{-1} = V_\beta^{-1} + X_k^\top X_k \ ;$$

$$m = \sum_{k=1}^{K} (m_k - (1 - 1/K)\mu_\beta) \text{ and } M^{-1} = \sum_{k=1}^{K} (M_k^{-1} - (1 - 1/K)V_\beta^{-1}) \ .$$

- Depends (crucially) on independence across subsets; not suitable for spatial random fields.

Geostatistical models for parallel architectures

- $y_{FH}(\ell) = \beta_0 + \beta_{tree} x_{tree}(\ell) + \beta_{fire} x_{fire}(\ell) + w(\ell) + \epsilon(\ell)$
- $w(\ell) \sim GP(0, C(\cdot,\cdot\,|\,\sigma^2,\phi))$
- $y_{FH} \sim N(X\beta, K_\theta)$ where K_θ is the spatial covariance matrix:
  $$K_\theta = C_{(\sigma,\phi)} + \tau^2 I\,, \;\; \text{where} \;\; \theta = \{\sigma,\phi,\tau\}$$
  where $C_{(\sigma^2,\phi)}$ is the GP covariance matrix derived from $C(\cdot,\cdot\,|\,\sigma^2,\phi)$.
- Massive data: divide and conquer?
- Bayesian model averaging? Predictive stacking? Exchangability?
- Meta-Kriging (GB, *Technometrics* 2018): find convex combination of subset-posteriors closest to the full posterior.
- Analyze "compressed data": Compressive sensing; Data sketching.

Bayesian Hierarchical Models

$$[\text{data} \,|\, \text{process}, \text{parameters}] \times [\text{process} \,|\, \text{parameters}] \times [\text{parameters}]$$

- Construct a joint model...

$$p(\theta, \tau, \beta) \times p(w \,|\, \theta) \times p(\tilde{w} \,|\, w, \theta) \times p(y \,|\, \beta, w, \tau) \times p(\tilde{Y} \,|\, \tilde{w}, \theta, \tau)$$

- Posterior inference for parameters and the process:

$$p(\theta, \tau, \beta, w, \tilde{w}, \tilde{Y} \,|\, y) \propto p(\theta, \tau, \beta, w \,|\, y) \times p(\tilde{w} \,|\, w, \theta) \times p(\tilde{Y} \,|\, \tilde{w}, \theta, \tau)$$

- Multivariate example with $Y = \{Y_j(s_i)\}$ for $j = 1, 2, \ldots, m$ variables:

$$MN(Y \,|\, XB, K_{\theta,\tau}, \Sigma) \times MN(B \,|\, \mu_B, V_B, \Sigma) \times IW(\Sigma \,|\, a, S) \times p(\theta, \tau)\ .$$

Constructing GPs from Graphs

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2016, VOL. 111, NO. 514, 800–812, Theory and Methods
http://dx.doi.org/10.1080/01621459.2015.1044091

Taylor & Francis
Taylor & Francis Group

Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets

Abhirup Datta, Sudipto Banerjee, Andrew O. Finley, and Alan E. Gelfand

ABSTRACT
Spatial process models for analyzing geostatistical data entail computations that become prohibitive as the number of spatial locations become large. This article develops a class of highly scalable nearest-neighbor Gaussian process (NNGP) models to provide fully model-based inference for large geostatistical datasets. We establish that the NNGP is a well-defined spatial process providing legitimate finite-dimensional Gaussian densities with sparse precision matrices. We embed the NNGP as a sparsity-inducing prior within a rich hierarchical modeling framework and outline how computationally efficient Markov chain Monte Carlo (MCMC) algorithms can be executed without storing or decomposing large matrices. The floating point operations (flops) per iteration of this algorithm is linear in the number of spatial locations, thereby rendering substantial scalability. We illustrate the computational and inferential benefits of the NNGP over competing methods using simulation studies and also analyze forest biomass from a massive U.S. Forest Inventory dataset at a scale that precludes alternative dimension-reducing methods. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2014
Revised April 2015

KEYWORDS
Bayesian modeling; Gaussian process; Hierarchical models; Markov chain Monte Carlo; Nearest neighbors; Predictive process; Reduced-rank models; Sparse precision matrices; Spatial cross-covariance functions

1. Introduction

With the growing capabilities of Geographical Information Systems (GIS) and user-friendly software, statisticians today routinely encounter geographically referenced datasets containing a large number of irregularly located observations on multiple variables. This has, in turn, fueled considerable interest in statistical modeling for location-referenced spatial data; see, for example, the books by Stein (1999), Moller and Waagepetersen (2003), Schabenberger and Gotway (2004), and Cressie and Wikle (2011), and Banerjee, Carlin, and Gelfand (2014) for a variety of methods and applications. Spatial process models introduce spatial dependence between observations using an underlying random field, $\{w(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, over a region of interest $\mathcal{D}$, which is endowed with a probability law that specifies the joint distribution for any finite set of random variables. For example, a zero-centered Gaussian process ensures that $\mathbf{w} = (w(\mathbf{s}_1), w(\mathbf{s}_2) \ldots, w(\mathbf{s}_n))' \sim N(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}))$, where $\mathbf{C}(\boldsymbol{\theta})$ is a family of covariance matrices, indexed by an unknown set of parameters $\boldsymbol{\theta}$. Such processes offer a rich modeling framework and are being widely deployed to help researchers comprehend complex spatial phenomena in the sciences. However, model fitting usually involves the inverse and determinant of $\mathbf{C}(\boldsymbol{\theta})$, which typically require $\sim n^3$ floating point operations (flops) and storage of the order of n^2. These become prohibitive when n is large and $\mathbf{C}(\boldsymbol{\theta})$ has no exploitable structure.

Broadly speaking, modeling large spatial datasets proceeds from either exploiting "low-rank" models or using sparsity. The former attempts to construct spatial processes on a lower-dimensional subspace (see, e.g., Higdon 2001; Kammann and Wand 2003; Rasmussen and Williams 2005; Stein 2007, 2008; Banerjee et al. 2008; Crainiceanu et al. 2008; Cressie and Johannesson 2008; Finley, Banerjee, and McRoberts 2009) by regressing the original (*parent*) process on its realizations over a smaller set of $r << n$ locations ("knots" or "centers"). The algorithmic cost for model fitting typically decreases from $O(n^3)$ to $O(nr^2 + r^3) \approx O(nr^2)$ flops since $n >> r$. However, when n is large, empirical investigations suggest that r must be fairly large to adequately approximate the parent process and the nr^2 flops become exorbitant (see Section 5.1). Furthermore, low-rank models perform poorly when neighboring observations are strongly correlated and the spatial signal dominates the noise (Stein 2014). Although bias-adjusted low-rank models tend to perform better (Finley, Banerjee, and McRoberts 2009; Banerjee et al. 2010; Sang and Huang 2012), they increase the computational burden.

Sparse methods include covariance tapering (see, e.g., Furrer, Genton, and Nychka 2006; Kaufman, Scheverish, and Nychka 2008; Du, Zhang, and Mandrekar 2009; Shaby and Ruppert 2012), which introduces sparsity in $\mathbf{C}(\boldsymbol{\theta})$ using compactly supported covariance functions. This is effective for parameter estimation and interpolation of the response ("kriging"), but it has not been fully developed or explored for more general inference on residual or latent processes. Introducing sparsity in $\mathbf{C}(\boldsymbol{\theta})^{-1}$ is prevalent in approximating Gaussian process likelihoods using Markov random fields (e.g., Rue and Held 2005), products of lower-dimensional conditional distributions (Vecchia 1988, 1992; Stein, Chi, and Welty 2004), or composite likelihoods (e.g., Bevilacqua and Gaetan 2014; Eidsvik et al. 2014). However, unlike low-rank processes, these do not, necessarily,

CONTACT Sudipto Banerjee sudipto@ucla.edu Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA 90095-1772
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

spNNGP

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 00, NO. 0, 1–14: Theory and Methods
https://doi.org/10.1080/01621459.2020.1833889

Taylor & Francis
Taylor & Francis Group

Check for updates

Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains

Michele Peruzzi[a,b], Sudipto Banerjee[c], and Andrew O. Finley[a]

[a]Department of Forestry, Michigan State University, East Lansing, MI; [b]Department of Statistical Science, Duke University, Durham, NC; [c]Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA

ABSTRACT
We introduce a class of scalable Bayesian hierarchical models for the analysis of massive geostatistical datasets. The underlying idea combines ideas on high-dimensional geostatistics by partitioning the spatial domain and modeling the regions in the partition using a sparsity-inducing directed acyclic graph (DAG). We extend the model over the DAG to a well-defined spatial process, which we call the meshed Gaussian process (MGP). A major contribution is the development of an MGPs on tessellated domains, accompanied by a Gibbs sampler for the efficient recovery of spatial random effects. In particular, the cubic MGP (Q-MGP) can harness high-performance computing resources by executing all large-scale operations in parallel within the Gibbs sampler, improving mixing and computing time compared to sequential updating schemes. Unlike some existing models for large spatial data, a Q-MGP facilitates massive caching of expensive matrix operations, making it particularly apt in dealing with spatiotemporal remote-sensing data. We compare Q-MGPs with large synthetic and real world data against state-of-the-art methods. We also illustrate using Normalized Difference Vegetation Index data from the Serengeti park region to recover latent multivariate spatiotemporal random effects at millions of locations. The source code is available at *github.com/mkln/meshgp*. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2020
Accepted October 2020

KEYWORDS
Bayesian; Domain partitioning; Graphical models; Large *n*; Sparsity; Spatial

1. Introduction

Collecting large quantities of spatial and spatiotemporal data is now commonplace in many fields. In ecology and forestry, massive datasets are collected using satellite imaging and other remote sensing instruments such as LiDAR that periodically record high-resolution images. Unfortunately, clouds frequently obstruct the view resulting in large regions with missing information. Figure 1 shows this phenomenon in Normalized Difference Vegetation Index (NDVI) data from the Serengeti region. Filling such gaps in the data is an important goal as is quantifying uncertainty in predictions. This goal is achieved through stochastic modeling of the underlying phenomenon, which involves the specification of a spatial or spatiotemporal process characterizing dependence from a finite realization. Gaussian processes (GPs) are a customary choice to characterize spatial dependence, but their implementation is notoriously burdened by their $O(n^3)$ computational complexity. Consequently, intense research has been devoted in recent years to developing scalable models for large spatial datasets—see detailed reviews by Sun, Li, and Genton (2011) and Banerjee (2017).

Computational complexity can be reduced by considering low-rank models; among these, knot-based methods motivated by "kriging" ideas enjoy some optimality properties but oversmooth the estimates of spatial random effects unless the number of knots is large, and require corrections to avoid overestimation of the nugget (Banerjee et al. 2008; Cressie and Johannesson 2008; Banerjee et al. 2010; Guhaniyogi et al. 2011; Finley, Banerjee, and Gelfand 2012). Other methods reduce the computational burden by introducing sparsity in the covariance matrix; strategies include tapering (Furrer, Genton, and Nychka 2006; Kaufman, Schervish, and Nychka 2008) or partitioning of the spatial domain into regions with a typical assumption of independence across regions (Sang and Huang 2012; Stein 2014). These can be improved by considering a recursive partitioning scheme, resulting in a multi-resolution approximation (MRA; Katzfuss 2017). Other assumptions on conditional independence assumptions also have a good track record in terms of scalability to large spatial datasets: Gaussian random Markov random fields (GMRF; Rue and Held 2005), composite likelihood methods (Eidsvik et al. 2014), and neighbor-based likelihood approximations (Vecchia 1988) belong to this family.

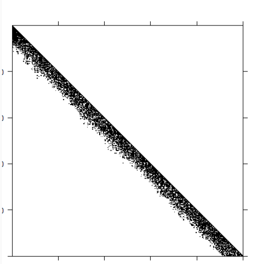
The recent literature has witnessed substantial activity surrounding the so called Vecchia approximation (Vecchia 1988). This approximation can be regarded as a special case of the GMRF approximations with a simplified neighborhood structure motivated from a directed acyclic graphical (DAG) representation of a GP likelihood. Extensions leading to well-defined spatial processes to accommodate inference at arbitrary locations by extending the DAG representation to the entire domain include nearest neighbor Gaussian processes (NNGPs; Datta,

CONTACT Sudipto Banerjee *sudipto@ucla.edu* Department of Biostatistics, UCLA Fielding School of Public Health, 650 Charles E. Young Drive South, Los Angeles, CA 90095.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

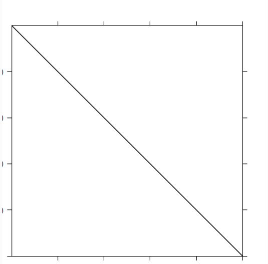
meshed

Sparse precision matrices (e.g., Vecchia's approximation; NNGP)

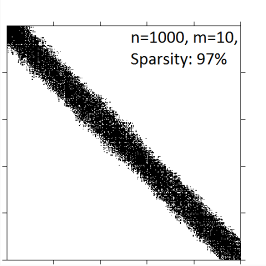
$$N(w \mid 0, K_\theta) \approx N(w \mid 0, \tilde{K}_\theta) \; ; \tilde{K}_\theta^{-1} = (I - A)^\top D^{-1} (I - A)$$



$I - A$



D^{-1}



$\tilde{K}_\theta^{-1}$

- $\det(\tilde{K}_\theta^{-1}) = \prod_{i=1}^{n} D_{ii}^{-1}$, $\tilde{K}_\theta^{-1}$ is sparse with $O(nm^2)$ entries

- Computing A and D

```
for(i in 1:(n-1) {
  Pa = N[i+1] # neighbors of i+1
  a[i+1,Pa] = solve(K[Pa,Pa], K[i+1, Pa])
  d[i+1,i+1] = K[i+1,i+1] - dot(K[i+1, Pa],a[i+1,Pa])
}
```

- We need to solve $n-1$ linear systems of size at most $m \times m$ in parallel.
- Quadratic form:

```
qf(u,v,A,D) = u[1] * v[1] / D[1,1]
for(i in 2:n) {
    qf(u,v,A,D) = qf(u,v,A,D) + (u[i] - dot(A[i,N(i)], u[N(i)]))
        *(v[i] - dot(A[i,N(i)], v[N(i)]))/D[i,i]
    }
```

- Determinant: $\det(\tilde{K}_\theta) = \prod_{i=1}^{n}$ `d[i,i]`

Alaska Tanana Valley data (Finley et al., *JCGS*, 2019)

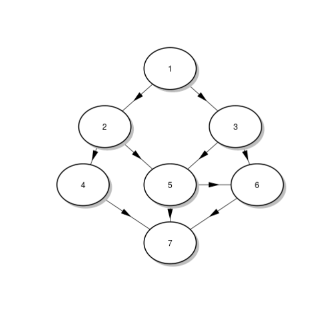
	Conjugate NNGP	Collapsed NNGP	Response NNGP
β_0	2.51	2.41 (2.35, 2.47)	2.37 (2.31,2.42)
β_{TC}	0.02	0.02 (0.02, 0.02)	0.02 (0.02, 0.02)
β_{Fire}	0.35	0.39 (0.34, 0.43)	0.43 (0.39, 0.48)
σ^2	23.21	18.67 (18.50, 18.81)	17.29 (17.13, 17.41)
τ^2	1.21	1.56 (1.55, 1.56)	1.55 (1.54, 1.55)
ϕ	3.83	3.73 (3.70, 3.77)	4.15 (4.13, 4.19)
CRPS	0.84	0.86	0.86
RMSPE	1.71	1.73	1.72
time (hrs.)	0.002	319	38

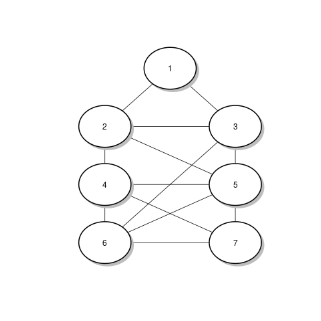
Table: Parameter estimates and model comparison metrics for the Tanana valley dataset

- Conjugate model produces estimates and model comparison numbers very similar to the MCMC based NNGP models
- For 5×10^6 locations, conjugate model takes 7 seconds

Highly Multivariate Graphical Gaussian Processes DDB, 2021

- Complex dependencies are often modeled using CI graphs (Cox & Wermuth, 1996)





- But what about complex dependencies among processes (Each node is $\{w_i(s) : s \in \mathbb{R}^d\}$)? And a very large number of nodes, too?

Highly Multivariate Graphical Gaussian Processes DDB, 2021

- What does this mean : $w_i(\cdot) \perp w_j(\cdot) \,|\, \{w_{-(ij)}(\cdot)\}$? Dalhaus (2000):

  $$\operatorname{cov}(z_i(s), z_j(s)) = 0 \quad \text{for all } s, s' \in \mathcal{D},$$

  where $z_i(s) = w_i(s) - \mathbb{E}[w_i(s) \,|\, \sigma(\{w_k(\cdot) : k \in \mathcal{V} \setminus \{i,j\}\})]$.

- Graphical GP (GGP): $\{w_i(\cdot) : i = 1, 2, \ldots, q\} \sim GGP_{\mathcal{G}}$ if $w_i(\cdot) \perp w_j(\cdot) \,|\, \{w_{-(i,j)}(\cdot)\}$ according to CI graph $\mathcal{G}$.

- Given a CI graph $\mathcal{G}$ and any cross-covariance function, there exists a unique (and optimal) $GGP_{\mathcal{G}}$ whose cross-covariance agrees with the given cross-covariance for all adjacent pairs in the graph.

Highly Multivariate Graphical Gaussian Processes DDB, 2021

- Constructing a GGP from a given $C(\mathcal{S})$ over a fixed finite set $\mathcal{S}$:
    1. Form an extended graph over $\mathcal{V} \times \mathcal{S}$ using *strong product* adjacency rules (to allow "stitching" across random fields);
    2. Use Dempster's (1972) covariance selection to specify $w(\mathcal{S}) \sim N(0, M(\mathcal{S}))$,
        - 2.1 $M_{ii}(\mathcal{S}) = C_{ii}(\mathcal{S})$ for each node i;
        - 2.2 Zeroes in $M(\mathcal{S})^{-1}$ correspond to CI relations in $\mathcal{G}$;
        - 2.3 $M_{ij}(\mathcal{S}) = C_{ij}(\mathcal{S})$ for all adjacent pairs in $\mathcal{G}$.
    3. Extend from finite set $\mathcal{S}$ to entire domain using predictive process with $\mathcal{S}$ as knots (Banerjee et al., 2008).
- DDB, 2021 also implement Bayesian inference for an unknown $\mathcal{G}$ using RJMCMC for (embeddable) decomposable graphs (Green & Thomas, 2013).

Parallelizable Stitching of Gaussian Processes

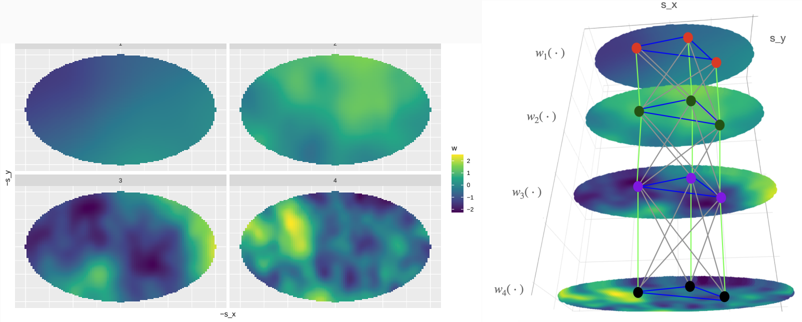
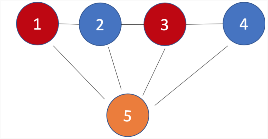


Figure: Stitching Gaussian Processes. Left: Realizations of 4 univariate GPs. Right: Realization of a multivariate (4-dimensional) GGP created by stitching together the 4 univariate GPs from the left figure using the strong product graph over the 4 variables and 3 locations.

Parallelizable Chromatic Gibbs Samplers



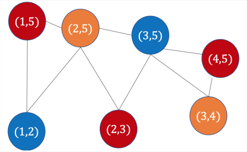
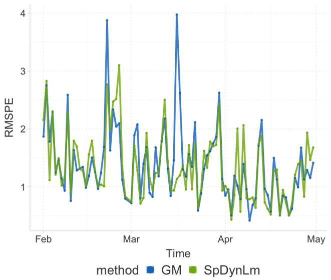
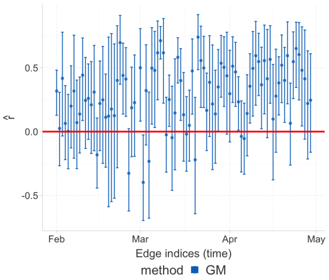


Figure: Chromatic sampling for GGP with a gem graph between 5 variables: Left: Gem graph and coloring used for chromatic sampling of the variable-specific parameters. Right: Coloring of the corresponding edge graph $\mathcal{G}_E(\mathcal{G}_V)$ used for chromatic sampling of the cross-covariance parameters. In chromatic sampling, we can use this coloring to sample nodes belonging to same color in parallel bringing down the complexity by significant amount.

Example DDB, 2021: 99 stations over 89 days



Prediction performance for full analysis



Estimates of time-specific cross-correlations

Burgeoning literature on DAG-based spatial models...

- Vecchia, A.V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 50, 297–312. DOI: https://doi.org/10.1111/j.2517-6161.1988.tb01729.x
- Stein, M.L., Chi, Z. and Welty, L.J. (2004), Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66, 275–296. DOI: https://doi.org/10.1046/j.1369-7412.2003.05512.x
- Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical Nearest-Neighbor Gaussian Process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111, 800–812. DOI: https://doi.org/10.1080/01621459.2015.1044091.
- Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A. S., and Schaap, M. (2016b). Non-separable dynamic Nearest-Neighbour Gaussian Process models for large spatio-temporal data with an application to particulate matter analysis. *Annals of Applied Statistics*, 10, 1286–1316. DOI: https://doi.org/10.1214/16-AOAS931
- Zhang, L., Datta, A. and Banerjee, S. (2018). Practical Bayesian modelling and inference for massive spatial datasets on modest computing environments. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **12**, 197–209. DOI: https://doi.org/10.1002/sam.11413
- Taylor-Rodriguez, D., Finley, A.O., Datta, A., Babcock, C., Andersen, H.E., Cook, B.C., Morton, D.C. and **Banerjee, S.** (2019). Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping. *Statistica Sinica*, 29, 1155–1180. DOI: https://doi.org/10.5705/ss.202018.0005.
- Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of gaussian processes. *Statistical Science*, 36, 124–141. DOI: https://doi.org/10.1214/19-STS755
- Peruzzi, M., Banerjee, S. and Finley, A.O. (in press). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *Journal of the American Statistical Association*, DOI: https://doi.org/10.1080/01621459.2020.1833889

Thank You!