Graphical Gaussian process models for highly multivariate spatial data

D. Dey, A. Datta, & S. Banerjee

진수정, 조형곤

공간정보의 통계분석 (2022A)

2022. 11. 30

Contents

Introduction

Multivariate spatial data

- Multivariate spatial data abound in the natural and environmental science when studying features of the joint distribution of multiple spatially dependent variables
- $w(s) = (w_1(s), w_2(s), ..., w_q(s))^T$ is modeled as zero-centered multivariate Gaussian process (GP)

Introduction

Multivariate spatial data

- The cross covariance is a matrix-valued function $C = (C_{ij}) : \mathcal{D} \times \mathcal{D} \to \mathbb{R}^{q \times q}$ with $C_{ij}(s, s') = Cov(w_i(s), w_j(s'))$
- $nq \times nq$ matrix $C(\mathcal{S}, \mathcal{S}) = \{C(s_i, s_j)\}$ for $S = \{s_1, ..., s_n\}$
- Mat$\acute{e}$rn cross-covariance function involve $O(q^2)$ parameters and $O(q^3)$ floating point operations
- Multivariate Mat$\acute{e}$rn models have typically restricted to applications with $q <= 5$

Introduction

Covariance Selection

- Covariance selection ensures the existence of multivariate distributions that retain univariate marginals while satisfying conditional independence relations specified by an inter-variable graph
- Existing covariance selection methods do not hold for spatial covariance functions over $\mathcal{D} \subset \mathbb{R}^d$

Introduction

Key contributions

- Identified the construction of a marginal retaining graphical Gaussian process (GGP) as a process-level covariance selection problem
- Proved existence, uniqueness, and optimality of the GGP
- Introduced a practical method for approximating the optimal GGP by 'stitching' GPs using an inter-variable graph
- Demonstrated computational scalability

Method

Graphical Gaussian process (GGP)

Definition (Graphical Gaussian Process)

A $q \times 1$ *Gaussian process* $w(\cdot)$ is *a graphical Gaussian process with respect to a graph* $\mathcal{G}_{\mathcal{V}} = (\mathcal{V}, E_{\mathcal{V}})$ *if the univariate Gaussian processes* $w_i(\cdot)$ *and* $w_j(\cdot)$ *are conditionally independent for every* $(i, j) \notin E_{\mathcal{V}}$. *We denote such a process by* $GGP(\mathcal{G}_{\mathcal{V}})$

Method

Graphical Gaussian process (GGP)

Theorem 1 *Let $\mathcal{G}_{\mathcal{V}} = (V, E_{\mathcal{V}})$ be any given graph and $C = (C_{ij})$ a $q \times q$ stationary cross-covariance function. Let $F(w) = \{f_{ij}(w)\}$ be the spectral density matrix corresponding to C at frequency w, and assume that $f_{ij}(\cdot)$ is square-integrable for all i.*

- *There exists a unique $q \times 1$ $GGP(\mathcal{G}_{\mathcal{V}})$ $W(\cdot)$ with cross-covariance function $M = (M_{ij})$ such that $M_{ij} = C_{ij}$ for $i = j$ and for all $(i, j) \in E_{\mathcal{V}}$.*
- *If $\tilde{F}(w)$ denotes the spectral density matrix of $w(\cdot)$ and F is the set of spectral density matrices of all possible $GGP(\mathcal{G}_{\mathcal{V}})$ $W(\cdot)$, then*

$$\tilde{F}(w) = \underset{K(\cdot) \in \mathcal{F}}{\operatorname{argmin}} \int_w d_{KL}\{F(w) \parallel k(W)\} dw \tag{1}$$

Method

Stitching of Gaussian processes

Given any $\mathcal{G}_{\mathcal{V}}$ and a cross-covariance function C, we seek a multivariate Gaussian process $w(\cdot)$ that fulfils the following conditions:

- Condition 1. The process exactly preserves the marginal distributions specified by C
- Condition 2. It is a $GGP(\mathcal{G}_{\mathcal{V}})$
- Condition 3. It exactly or approximately retains the cross-covariance specified by C for pairs of variables included in $\mathcal{G}_{\mathcal{V}}$

Method

Stitching of Gaussian processes

Stitching process:

- Begin construction on $\mathcal{L}$, a finite arbitrary set of locations in $\mathcal{D}$
- Stitch the variables at the locations in $\mathcal{L}$ *s.t.* there is a 'thread' between two location-variable pairs if and only if there is an edge between the two variables in the graph
- Stitch each of the remaining surfaces independently so that they have the same distribution as the univariate surfaces

Method

Stitching of Gaussian processes

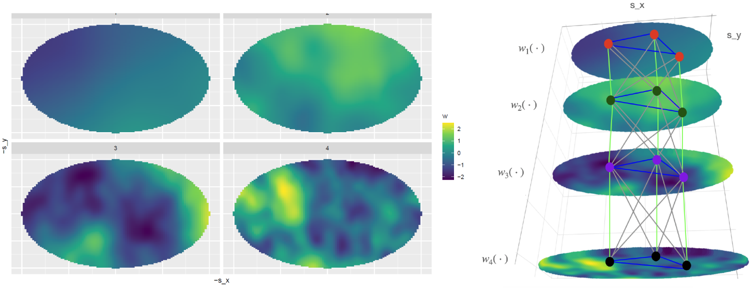


Figure: Stitching Gaussian processes

Method

Stitching of Gaussian processes

To fulfil the rerquirements, $w(\mathcal{L}) \sim N(0, M(\mathcal{L}, \mathcal{L}))$ is modeled *s.t.*

- Condition 4. $M_{ii}(\mathcal{L}, \mathcal{L}) = C_{ii}(\mathcal{L}, \mathcal{L})$ for all $i = 1, 2, ..., q$
- Condition 5. $(M(\mathcal{L}, \mathcal{L})^{-1})_{ij} = 0$ for all $(i, j) \notin E_{\mathcal{V}}$
- Condition 6. $M_{ij}(\mathcal{L}, \mathcal{L}) = C_{ij}(\mathcal{L}, \mathcal{L})$ for all $(i, j) \in E_{\mathcal{V}}$

Method

Stitching of Gaussian processes

Extension to infinite-dimensional GP

- Predictive process + independent residual process

$$w_i(s) = w_i^* + z_i(s) = C_{ii}(s, \mathcal{L})C_{ii}(\mathcal{L}, \mathcal{L})^{-1}w_i(\mathcal{L}) + z_i(s) \qquad (2)$$

Calculation process

Covariance construction

- The existence and the construction of a marginal-preserving GGP were covered, given any valid cross-covariance C and any inter-variable graph $\mathcal{G}_{\mathcal{V}}$.
- Of particular interest are developing a multivariate graphical Mat$\acute{e}$rn GPs such that each univariate process is a Mat$\acute{e}$rn GP (Gneiting et al., 2010; Apanasovich et al., 2012).
  - We retain the ability to interpret the parameters for each univariate spatial process (Most other multivariate covariance functions fail to retain this property).

Calculation process

Covariance construction

The isotropic multivariate Mat$\acute{e}$rn cross-covariance function on a *d*-dimensional domain is

$$C_{ij}(s, s^{'}) = \sigma_{ij} H_{ij}(\|s - s^{'}\|) \tag{3}$$

where $H_{ij}(\cdot) = H(\cdot|\nu_{ij}, \phi_{ij})$, H being the Mat$\acute{e}$rn correlation function (Apanasovich et al, 2012).

Calculation process

Covariance construction

- For validity, it is sufficient to constrain the intrasite covariance matrix $\Sigma = (\sigma_{ij})$ to be of the form (Apanasovich et al., 2012):

$$\sigma_{ij} = b_{ij} \frac{\Gamma\{(\nu_{ii} + \nu_{jj} + d)/2\}\Gamma(\nu_{ij})}{{\phi_{ij}}^{2\Delta_A + \nu_{ii} + \nu_{jj}}\Gamma(\nu_{ij} + d/2)} \tag{4}$$

where $\Delta_A \geq 0$ and $B = (b_{ij})$ is positive definite

Calculation process

Covariance construction

When q is large,

- Stitching needs to constrain $B = (b_{ij})$ to be p.d. on an $O(q^2)$-dimensional parameter space, and verifying positive definiteness on B incurs an additional cost of $O(q^3)$ flops.
- Evaluating $\omega(\mathcal{L}) \sim N(0, M(\mathcal{L}, \mathcal{L}))$ involves matrix operations for the $nq \times nq$ matrix $M(\mathcal{L}, \mathcal{L})$. While the precision matrix, $M(\mathcal{L}, \mathcal{L})^{-1}$, is sparse, its determinant is usually not available in closed form and the calculation can become prohibitive even for small n.
- To facilitate scalability in highly multivariate settings, consider decomposable inter-variable graphs.

Calculation process

Graph construction

Definition (Decomposable graph)

A graph is **decomposable** if it is *complete*, or if there exists a proper decomposition into *decomposable subgraphs*.

- A decomposable graph is equivalent to a chordal graph, which possesses no chordless cycles of length four or greater.

Calculation process

Graph construction

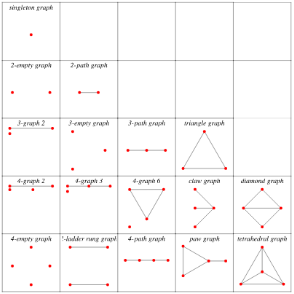


Figure: Examples of decomposable graph

- Several naturally occuring dependence structures like autoregressive dependence correspond to decomposable graphs.

Calculation process

Graph construction

- Assuming decomposability is conspicuous in graphical models (Dobra et al., 2003; Wang & West, 2009) since
  - Fitting Bayesian graphical models is cumbersome for non-decomposable graphs (Roverato, 2002; Atay-Kayis & Massam, 2005).
  - Non-decomposable graph can be embedded in a larger decomposable graph.
- If $\mathcal{G}_V$ is decomposable, then it has a perfect clique sequence (Lauritzen, 1996).

Calculation process

Graph construction

If the decomposable graph $\mathcal{G}_\nu$ has a perfect clique sequence $\{K_1, K_2, ..., K_p\}$ with separators $\{S_2, ..., S_m\}$, then

- the GGP likelihood on $\mathcal{L}$ can be decomposed as

$$f_M\{\omega(\mathcal{L})\} = \frac{\prod_{m=1}^{p} f_C\{\omega_{K_m}(\mathcal{L})\}}{\prod_{m=2}^{p} f_C\{\omega_{S_m}(\mathcal{L})\}} \tag{5}$$

  where f_A denotes the density of a Gaussian process over $\mathcal{L}$ with covariance function $A \in \{M, C\}$

  - Helps to manage the dimension and constraints of the parameter space and the computational complexity of stitching.

Calculation process

Graph construction

- and the precision matrix of $\omega(\mathcal{L})$ satisfies (Lauritzen, 1996)

$$M(\mathcal{L},\mathcal{L})^{-1} = \sum_{m=1}^{p} \{C^{-1}_{[K_m \boxtimes \mathcal{G}_\mathcal{L}]}\}^{\mathcal{V}\times\mathcal{L}} - \sum_{m=2}^{p} \{C^{-1}_{[S_m \boxtimes \mathcal{G}_\mathcal{L}]}\}^{\mathcal{V}\times\mathcal{L}} \quad (6)$$

where for any symmetric matrix $A = (a_{ij})$ with rows and columns indexed by $\mathcal{U} \subset \mathcal{V} \times \mathcal{L}$, $A^{\mathcal{V}\times\mathcal{L}}$ is a $|\mathcal{V}\times\mathcal{L}| \times |\mathcal{V}\times\mathcal{L}|$ matrix s.t.

$$(A^{\mathcal{V}\times\mathcal{L}})_{ij} = \begin{cases} a_{ij} & (i,j) \in \mathcal{U} \\ 0 & o.w. \end{cases}$$

  - The stitching likelihood evaluation avoids the large matrix $M(\mathcal{L},\mathcal{L})$ and all matrix operations are limited to the sub-matrices of $M(\mathcal{L},\mathcal{L})$.

Calculation process

Graph construction

Table 1 summarizes these gains from stitching with decomposable graphs.

Table 1: *Properties of any q-dimensional multivariate Matérn GP of Gneiting et al. (2010) or Apanasovich et al. (2012) and a multivariate graphical Matérn GP stitched using a decomposable graph $\mathcal{G}_V$ with largest clique size q^*, length of perfect ordering p, and maximal number of cliques p^* sharing a common vertex.*

Model attributes	Multivariate Matérn	Graphical model Matérn
Number of parameters	$O(q^2)$	$O(\|E_{\mathcal{V}}\| + q)$
Parameter constraints	$O(q^3)$	$O(p^*(q^{*3}))$ (worst case)
Storage	$O(n^2q^2)$	$O(pn^2q^{*2})$
Time complexity	$O(n^3q^3)$	pn^3q^{*3}
Conditionally independent processes	No	Yes
Univariate components are Matern GPs	Yes	Yes

Calculation process

Graph construction

- The construction of GGP assumes a known graphical model.
- When a graph is unknown:Adapt an MCMC sampler for decomposable graphs, extending the framework below.

$$y_i(s) = x_i(s)^T \beta_i + \omega_i(s) + \epsilon(s),\ i = 1, 2, ..., q,\ s \in \mathcal{D}$$

$$p(\beta, \tau, \theta) \times N(\omega(\mathcal{S}|0, \mathcal{C}_\theta(\mathcal{S}, \mathcal{S})) \times \prod_{j=1}^{n} N(y(s_j)|X(s_j)\beta + \omega(s_j), D_\tau))$$

- Then we can infer about the graphical model itself along with the GGP parameters (Green & Thomas, 2013).

Calculation process

Graph construction

Implementation

- Known graph: We implement a chromatic Gibbs sampler.
- Unknown graph: We agument the sampler above with a reversible jump MCMC sampler (rjMCMC).

Calculation process

Graph construction

Chromatic Gibbs sampler

- Uses the graph coloring to facilitate parallel simulation.
- Coloring: A coloring is a collection of labels assigned to nodes on a graph so that no two nodes that share an edge have the same label.

Fig. 2. Chromatic sampling for a graphical Gaussian process with a gem graph between five variables: (a) gem graph and colouring used for chromatic sampling of the variable-specific parameters; (b) colouring of the corresponding edge graph $\mathcal{G}_E(\mathcal{G}_V)$ used for chromatic sampling of the cross-covariance parameters b_{ij}.

Figure: Chromatic sampling for GGP with a gem graph between 5 variables

Simulation

Known graph

Want to compare three models:

- (a) PM: Parsimonious Multivariate Mat$\acute{e}$rn of Gneiting et al. (2010)
  $\nu_{ij} = (\nu_{ii} + \nu_{jj})/2$ and $\phi_{ij} = \phi$
- (b) MM: Multivariate Mat$\acute{e}$rn of Apanasovich et al. (2012)
  $\nu_{ii} = \nu_{ii} = \nu_{jj} = \frac{1}{2}$, $\Delta_A = 0$, and $\phi_{ij}^2 = (\phi_{ii}^2 + \phi_{jj}^2/2)$
- (c) GM: Graphical Mat$\acute{e}$rn (GGP on the latent process, stitched using multivariate Mat$\acute{e}$rn model (b))

Simulation

Known graph

Scenario of Simulation

Table 2: Different simulation scenarios considered for the comparison between methods.

Set	q	Graph $\mathcal{G}_\mathcal{V}$	B	Nugget	Locations	Data model	Fitted models
1A	5	Gem (Figure 2(a))	Random	No	Same location for all variables	GM	GM, MM, PM
1B	5	Gem (Figure 2(a))	Random	No	Same location for all variables	MM	GM, MM, PM
2A	15	Path	$b_{i-1,i} = \rho_i$	Yes	Partial overlap in locations for variables	GM	GM, PM
2B	15	Path	$b_{i-1,i} = \rho_i$	Yes	Partial overlap in locations for variables	MM	GM, PM
3A	100	Path	$b_{i-1,i} = \rho_i$	Yes	Partial overlap in locations for variables	GM	GM
3B	100	Path	$b_{i-1,i} = \rho_i$	Yes	Partial overlap in locations for variables	MM	GM

For all scenarios,

- Generate data on $n = 250$ locations uniformly chosen over a grid
- 1 covariate $x_j(s_i)$ for each variable j, independently from a $N(0,4)$
- $\beta_j \sim Unif(-2,2)$ for $j = 1,2,...,q$
- ϕ_{ii}, σ_{ii}: equispaced numbers in $(1,5)$

For implementing a known graph, we can deploy a chromatic Gibbs sampler and R package *BRISC* was used for estimation.

Simulation

Known graph

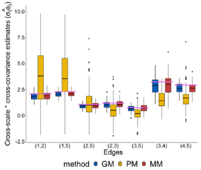
Comparing Estimation Performance

- Note that b_{ij}, $(i, j) \in E_{\mathcal{V}}$ specify the cross-covariances in stitching.
- Compare the estimates of scaled b_{ij}, $\sigma_{ij}\phi_{ij} = \Gamma(1/2)b_{ij}$.

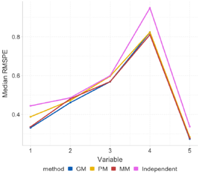
Simulation

Known graph

Result



(a) Set 1B



(d) Set 1B

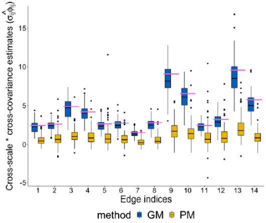
Figure: (a): Estimates of the cross-covariance parameters $\sigma_{ij}\phi_{ij}$ for the sets 1B, (d): Median RMSPE for GM, MM, PM and Independent GP model (reference for the impact of not modelling dependence) for set 1B

- (a) MM, and GM produce reasonable estimates of the true cross-covariance parameters included whereas the estimates from PM are biased and more variable.
- (d) GM performs competitively with MM (the correctly specified model).
- (d) PM yields higher RMSPE for variables 1 and 3, while the independent model is, unsurprisingly, the least accurate.

Simulation

Known graph

Result



(b) Set 2B

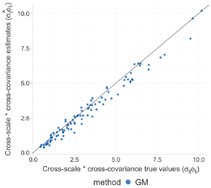
Figure: (b): Estimates of the cross-covariance parameters $\sigma_{ij}\phi_{ij}$ for the sets 2B

- The estimates of PM are biased, while GM is more accurate.

Simulation

Known graph

Result



(c) Set 3B

Figure: (c): Estimates of the cross-covariance parameters $\sigma_{ij}\phi_{ij}$ for the sets 3B

- Since neither PM nor MM can be implemented for the highly multivariate setting, only compare the estimates from GGP to the truth.
- GM accurately estimates all the b_{ij}'s.

Simulation
Unknown graph

Scenario of Simulation

- Consider Set 1A and 2A from Table 2, where the true multivariate process is a graphical Mat$\acute{e}$rn.
- For implementing an unknown graph, we can deploy a reverse jump MCMC.
- Assess the accuracy of inferring about the graphical model and the estimates of the cross-covariance parameters.

Simulation

Unknown graph

Results

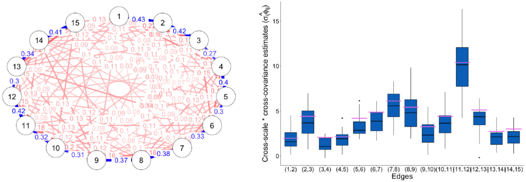


Figure: (a): Posterior edge selection probabilities for Set 2A, (b): Cross-covariance parameter estimates for Set 2A while estimating the unknown graph

- (a): Most of the false edges have narrow width indicating their low selection probability.

  blue edge: true edge / red edge: false edge

  width of the edge: proportional to the posterior probability of selecting that edge
- (b): The cross-covariance parameters corresponding to true edges are also estimated correctly.

Application

Problem statements

Spatio-temporal modelling of PM$_{2.5}$

Goal

- We model daily levels of PM$_{2.5}$ measured at monitoring stations across 11 states of the north-eastern US and Washington DC.

Setting

- Duration: $T = 89$ days (February, 01, 2020 $\sim$ April, 30th, 2020)
- Locations: $n = 99$ are selected (with at least two months of measured data for both 2020 and 2019).
- Covariates: baseline covariate (The daily 2019 PM$_{2.5}$ data) $+$ Meteorogical variables (temparature, barometric pressure, wind-speed and relative humidity)

Model Description

- **GM** (GGP Model)
  - Used Mat$\acute{e}$rn cross-covariance function with AR(1) graph
- **PM** (Parsimonious Mat$\acute{e}$rn)
  - Not implementable for full dataset
- **SpDynLM**
  - Set in the GP-based mixed-effect modelling setup
  - Models the spatial process $w_t(\cdot) = w(\cdot, t)$ at time t as

$$w_t(s) = w_{t-1}(s) + \delta_t(s);\ \delta_t(\cdot) \sim GP(0, C_{tt})$$

Application

Problem solving process

Comparison 1 (GM vs. PM)

- Break 89 days into 6 fornights subgroups.
- Analyse each chunk separately using GM and PM.

Comparison 2 (GM vs. SpDynLM)

- Analyse the full dataset using GM and SpDynLM.

Application

Result (GM vs. PM)

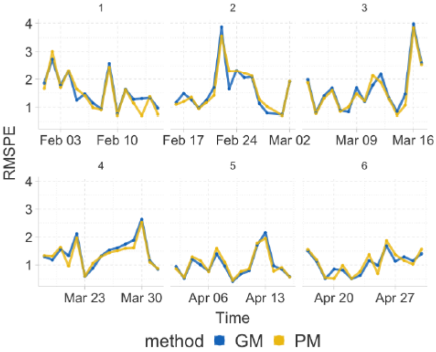


Figure: daily RMSPE for six fortnightly analyses

- GM and PM produce very similar predictive performance when analysing each fortnight of data separately.

Application

Result (GM vs. SpDynLM)

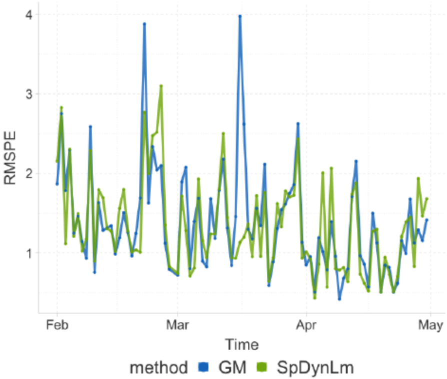


Figure: daily RMSPE for full analyses

- Prediction performance is similar for both models with respect to both point predictions.

Application

Result (GM vs. SpDynLM)

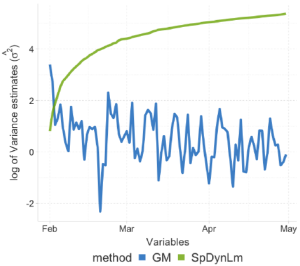


Figure: Estimates of the time-specific process variances

- Compared estimates of the marginal process variances from GGP and SpDynLM.
  - SpDynLM shows monotonically increasing variance over time, which is unrealistic.
  - GM shows substantial variation across time with generally decreasing trend going from Feb. to Apr.

Application

Result (GM vs. SpDynLM)

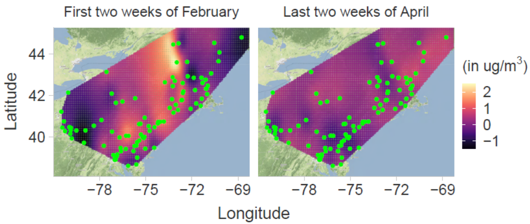


Figure: Estimates of the residual spatial processes from the graphical model after adjusting for the covariates and the baseline, for the first two weeks of February and last two weeks of April

- Slight decrease in the magnitude of the residual process
- The residuals for Apr. also showed much less variability than in Feb., which agrees with the previous result of GM.

Conclusion

- Existence, uniqueness, and optimality of a marginal retaining GGP by covariance selection.
- Approximate this optimal GGP by stitching GPs together using inter-variable graph.
- Decomposable graph assumption reduces computational complexity.
- Distinctly different from prior work referring to the massive number of spatial locations.

감사합니다.