NNGP Models with Spatial Examples of Simulated Data and Forest Canopy Heights

Frances Lin

Dec 2022

1. Introduction

Inferences and predictions on large spatial data or data with locations $\approx 10^6$ have either been too computationally challenging or infeasible. Methods for large spatial data that are under active development. However, most of the existing methods has focused primarily on theoretical and methodological developments and have not focused enough on the algorithmic details nor made use of high-performance computing (HPC) libraries to expedite expensive computations (Finley et al., 2019).

On the other hand, while the original nearest neighbor Gaussian process (NNGP) model (Datta et al., 2016), which is also referred to as the latent NNGP model, appears promising, the latent NNGP model is prone to high autocorrelations and slow convergence because the Gibbs sampler involves updating a high-dimensional vector of latent random effect *sequentially*.

Three alternate formulations of the NNGP model that are more efficient and practical than the latent NNGP model (2016) are proposed: (1) the collapsed NNGP model, (2) the response NNGP model, and (3) the conjugate NNGP model, and all except the collapsed model are implemented and can be accessed through the `R` package `spNNGP` (Finley et al., 2019; Finley et al., 2021).

Section 2 reviews Gaussian process (GP) and formulates nearest neighbor Gaussian process (NNGP) through graphical models. *Section 3* introduces the latent NNGP model, which is followed by three alternate formulations of the latent NNGP model: (1) a collapsed NNGP model, (2) a NNGP model for the response (with no latent process), and (3) a conjugate NNGP model that allows for MCMC-free exact inference. *Section 4* considers the simulated data set and real data set in the `spNNGP` package and uses the package and/or hpc (high performance computing) cluster at OSU to compare run time and replicate results. *Section 5* includes the discussion.

2. Nearest Neighbor Gaussian Processes

2.1. Review of mixed-effects and GP model for spatial data

Let $(s_i, y(s_i), x(s_i))$ be a triplet, where s_i denotes the location of measurement, $y(s_i)$ denotes the response of interest and $x(s_i)$ denotes the known or observed covariates, for $i = 1, ..., n$, a spatial linear mixed-effects

model is given as

$$y(s_i) = x(s_i)^T\beta + w(s_i) + \epsilon(s_i), \tag{1}$$

where β is the vector of coefficients, $w(s_i)$ is the vector of unknown or unobserved covariates (random effects), and $\epsilon(s_i) \sim^{iid} N(0, \tau^2)$ is the random noise (Finley et al., 2021).

Gaussian processes (GPs) are widely used in machine learning to model smooth functions for regression, classification, and other tasks (Rasmussen, 2003, as cited in Finley et al., 2021). In spatial statistics, GPs are typically used to model the latent surface $\{w(s)\}$. A GP model for the spatial surface, which is given as

$$w(s) \sim GP(0, C(\cdot, \cdot|\theta)),$$

where $C(\cdot, \cdot|\theta)$ is a covariance function, implies that the vector of random effects $w = (w(s_1), ..., w(s_n))^T$ follows a multivariate Gaussian distribution with mean zero and covariance matrix $C(\theta) = C = (c_{ij})$, where $c_{ij} = C(s_i, s_j|\theta)$ and θ is the covariance parameters of the GP. The process is completely specified by a valid covariance function $C(\cdot, \cdot|\theta)$ (Datta et al., 2016). A popular choice of θ for $C(\cdot, \cdot|\theta)$ is selected from the Matérn covariance function or Matérn kernel (Matérn, 1960). For example, let s_i and s_j be two points in $\mathcal{D}$, then the Matérn covariance function is given as

$$C(s_i, s_j; \sigma^2, \phi, \nu) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(||s_i - s_j||\phi)^\nu \mathcal{K}_\nu(||s_i - s_j||\phi), \quad \phi > 0, \nu > 0,$$

where $||s_i - s_j||$ is the Euclidean distance between locations s_i and s_j, $\theta = \{\sigma^2, \phi, \nu\}$ are respectively the marginal variance, scale (inverse of range) or decay and smoothness parameter, Γ is the gamma function, $||\cdot||$ denotes the Euclidean distance in $\mathbb{R}^d$, and $\mathcal{K}$ is the Bessel function of second kind (Stein, 1999, as cited in Datta et al., 2016; Finley et al., 2019).

The hierarchical model can be constructed by combining the mixed-effects model for the response y and the GP model for the random effects w and is given as

$$p(\beta, \theta, \tau^2) \times N(w|0, C(\theta)) \times N(y|X\beta + w, \tau^2 I), \tag{2}$$

where $p(\beta, \theta, \tau^2)$ is specified by assigning priors to β, θ and τ^2 (Finley et al., 2021). Alternatively, the marginal model can be constructed by integrating w out from (2) and is given as

$$N(y|X\beta, \Sigma = C(\theta) + \tau^2 I), \tag{3}$$

(Finley et al., 2021).

In a frequentist paradigm, parameter estimation can be obtained from (3) via MLE (maximum likelihood estimation), whereas in a Bayesian framework, after assigning priors to the parameters, posterior inference can be obtained from either (2) or (3) using MCMC (Markov chain Monte Carlo; Finley et al., 2021).

Alternatively, (2) is the same as

$$(\beta, \tau^2, \theta) \sim p(\beta, \tau^2, \theta)$$

$$w|\theta \sim N(0, C(\theta))$$

$$y|\beta, w, \tau^2 \sim N(X\beta + w, \tau^2 I).$$

2.2. Nearest Neighbor Gaussian Processes for large spatial data

Unfortunately, when n is large, evaluating both (2) and (3) can be computationally challenging or infeasible. More specifically, evaluating $N(w|0, C)$ requires $\mathcal{O}(n^3)$ computations and storing the covariance matrix C ($C(\theta)$ and $C(\theta) + \tau^2 I$, respectively) requires $\mathcal{O}(n^2)$ storage (Finley et al., 2021). In addition, predicting the response at new locations K requires additional $\mathcal{O}(kn^2)$ operations. Integrating w out from (2) does not always give computational advantages either (Finley et al., 2019).

One of the solutions is to replace GP prior for the spatial random effects w with a NNGP prior (Datta et al., 2016, as cited in Finley et al., 2021). Let $\mathcal{R} = \{s_1, ..., s_n\}$ be any finite set of locations in the spatial domain $\mathcal{D}$ and $w_{\mathcal{R}} = (w(s_1), ..., w(s_n))^T$. For any location in $\mathcal{D}$, define neighbor sets as

$$N(s_1) = \{\} \text{ (empty set)},$$

$$N(s_i) = \min(m, i-1) \text{ nearest neighbors of } s_i \text{ in } s_1, ..., s_{i-1}, \text{ for } i = 2, ..., n, \tag{4}$$

$$N(s) = m \text{ nearest neighbors of } s \text{ in } \mathcal{R},$$

then the NNGP is given as

$$w_{\mathcal{R}} \sim \prod_{i=1}^{n} p(w_i | w(N(s_i))) \tag{5}$$

(Finley et al., 2021). The above construction can be shown to be a multivariate Gaussian distribution on $w = w_{\mathcal{R}}$

$$w \sim N(0, \tilde{C}(\theta)), \tag{6}$$

where $\tilde{C}^{-1}(\theta)$ the inverse of the NNGP covariance matrix is sparse (Datta et al., 2016, as cited in Finley et al., 2021).

The idea is do introduce sparsity through graphical models. The joint distribution for a random vector w can be viewed as a directed acyclic graph (DAG). More specifically, by the chain rule, the joint density $p(w) = p(w_1, w_2, ...w_n)$ can be written as a product of conditional densities. That is,

$$p(w) = p(w_1) \prod_{i=2}^{n} p(w_i | Pa[i]), \tag{4-1}$$

where $w_i \equiv w(s_i)$ and $Pa[i] = \{w_1, w_2, ...w_{i-1}\}$ is a set of parents of w_i, and sparse models for w can be constructed by shrinking the size of $Pa[i]$ (Finley et al., 2019). In spatial statistics, $Pa[i]$ is defined to be the set of $w(s_j)$'s corresponding to a small number m ($\ll n$) of nearest neighbor sets of s_i (Finley et al., 2019). The resulting density is then

$$\tilde{p}(w) = p(w_1) \prod_{i=2}^{n} p(w_i | w_{N(s_i)}), \tag{4-2}$$

where $N(s_i)$ is referred to as the neighbor set for s_i (Datta et al., 2016; Zheng et al., 2022).

To construct a sparse precision matrix, start with a dense $n \times n$ covariance matrix C and construct a sparse strictly lower-triangular matrix A with no more than m ($\ll n$) nonzero entries in each row and the diagonal matrix D, then the matrix $\tilde{C} = (I - A)^{-1} D (I - A)^{-T}$ is a covariance matrix and its inverse $\tilde{C}^{-1} = (I - A)^T D^{-1} (I - A)$ is sparse (Finley et al., 2019). This leads to the latent NNGP model in the section below.

NNGP can also be viewed as as a special case of a Gaussian Markov Random Field (GMRF; Rue and Held 2005, as cited in Finley et al., 2021) with the advantage of not requiring mesh-based SPDE (stochastic partial differential equation) approximation of Gaussian Random Field (Finley et al., 2021).

3. NNGP Models

3.0. Latent NNGP

The original NNGP model proposed by Datta et al. (2016) constructed the neighbor sets based on m nearest neighbors and replaced the GP prior for spatial random effects w in (2) with a NNGP prior

$$w \sim N(0, \tilde{C}(\theta)). \tag{6}$$

This model is referred to as the latent NNGP model, which uses a fully Bayesian hierarchical specification

$$p(\beta, \theta, \tau^2) \times N(w|0, \tilde{C}(\theta)) \times N(y|X\beta + w, \tau^2 I), \tag{7}$$

for running an MCMC (Markov chain Monte Carlo) algorithm, and the parameters $\{w, \beta, \theta, \tau^2\}$ are updated in a Gibb's sampler (Finley et al., 2021).

Normal priors for β and inverse Gamma priors for the variance components τ^2 ensure that they yield conjugate full conditionals in the Gibbs sampler (Finley et al., 2021). The remaining covariance parameters θ are updated using random-walk Metropolis steps for their respective full conditionals (Finley et al., 2021).

The full conditional distribution for w in (7) is

$$w|\cdot \sim (B(y - X\beta)/\tau^2, B),$$

where $B = \tilde{C}^{-1}(\theta) + I/\tau^2$ is the full conditional precision matrix. However, this block update of w is not practical. This is because even though B is as sparse as $\tilde{C}^{-1}(\theta)$ is, unlike $\tilde{C}^{-1}(\theta)$, the determinant of B cannot be calculated in $\mathcal{O}(n)$ FLOPs (Finley et al., 2021). Instead, the MCMC implementation of the latent NNGP model involves updating the n full conditions (or parameters) $w_i|\cdot$ *sequentially* (Finley et al., 2021). However, MCMC convergence for high-dimensional model like thi one is difficult to study to prove reliable and can also imply slow convergence (Finley et al., 2019; Finley et al., 2021).

Alternatively, (7) is the same as

$$(\beta, \tau^2, \theta) \sim p(\beta, \tau^2, \theta)$$

$$w|\theta \sim N(0, \tilde{C}(\theta))$$

$$y|\beta, w, \tau^2 \sim N(X\beta + w, \tau^2 I).$$

Marginalizing one or more variables out tend to lower autocorrelation and improve convergence behavior in MCMC (Liu et al., as cited in Finley et al., 2019). As a result, three alternate variants of the latent NNGP model that aim to reduce the parameter dimensionality are proposed, and these models consider marginalizing over the entire vector of spatial random effects w (Finley et al., 2019).

3.1. Collapsed NNGP

A collapsed NNGP model not only enjoys the frugality of a low-dimensional MCMC chain but also allows for full recovery of the latent random effects w (Finley et al., 2019).

Consider the two-stage hierarchical specification $N(w|0, \tilde{C}(\theta)) \times N(y|X\beta + w, \tau^2 I)$ and integrate out w to avoid sampling w in the Gibb's sampler, then the collapsed NNGP model is specified as

$$y \sim N(X\beta, \Sigma = \tilde{C}(\theta) + \tau^2 I), \tag{7*}$$

where $\theta = \{\sigma^2, \phi, \nu\}$ for Matérn covariance function (Finley et al., 2019).

A normal prior $N(\mu_\beta, V_\beta)$ is used for β, inverse-Gamma priors are used for the spatial and noise variances σ^2 and τ^2, and uniform priors are used for the range and smoothness parameters $1/\phi$ and ν.

MCMC steps for updating $\{\beta, \theta, \tau^2\}$ involves (1) updating $\{\theta, \tau^2\}$ through Metropolis Hastings, (2) updating β through Gibb's sampler, and (3) repeating step (1) and (2) to obtain N MCMC samples for $\{\beta, \theta, \tau^2\}$. Interested readers are referred to **Algorithm 1** and **2** in *Section 2.1* of Finley et al. (2019) for additional algorithmic details.

3.2. NNGP for the Response

Both the latent NNGP and collapsed NNGP model (the collapsed version of the latent NNGP model) make predication at a new location by first recovering the spatial random effects w and predicting value at the new location with kriging. However, if inference on the latent process is of interest, the recovery of w is necessary. Otherwise, it is often a computational burden (Finley et al., 2019).

Instead of using NNGP for the latent Gaussian process w, the response NNGP model directly applies the marginal Gaussian process for the response y (Finley et al., 2019, as cited in Finley et al., 2021).

Consider the GP marginal model for the response

$$Y \sim N(X\beta, \Sigma),$$

where Σ is the marginalized covariance function Σ and is specified as $\Sigma(s_i, s_j) = C(s_i, s_j|\theta) + \tau^2\delta(s_i, s_j)$, where δ is the Kronecker delta (Finley et al., 2021). Since the covariance function of an NNGP can be derived from any parent GP, next replace the full GP covariance Σ with its NNGP analogue $\tilde{\Sigma}$, then the response NNGP marginal model is given as

$$Y \sim N(X\beta, \tilde{\Sigma}), \tag{8}$$

where $\tilde{\Sigma}$ is the NNGP covariance matrix derived from $\Sigma = C(\theta) + \tau^2 I$. (Finley et al., 2021). The sparsity properties in *Section 2.2* can also be applied to $\tilde{\Sigma}^{-1}$ (Finley et al., 2019).

The dimension of the parameter space is reduced from $\mathcal{O}(n)$ to $\mathcal{O}(1)$, and the lower dimensional NNGP tends to have improved MCMC convergence (Finley et al., 2019, as cited in Finely et al., 2021).

MCMC steps for updating $\{\beta, \theta, \tau^2\}$ is similar to but less involved than those of the collapsed NNGP model. Interested readers are referred to **Algorithm 3** and **4** in *Section 2.2* of Finley et al. (2019) for additional algorithmic details.

3.3. MCMC-Free Exact Bayesian Inference Using Conjugate NNGP

MCMC methods are commonly used to obtain approximate inference since the normalizing constant often involves high-dimensional integrals and is therefore hard to compute (Salakhutdinov, 2010). However, running MCMC methods such as the Gibbs' sampler for several thousand iterations may be very slow (Finley et al., 2019).

The conjugate NNGP model offers MCMC-free exact posterior inference by fixing certain covariance parameters (i.e., ϕ and α) in the response NNGP model (Finley et al., 2021).

Recall that Σ of the GP marginal model in Section 3.2. is given as

$$\Sigma = \Sigma(s_i, s_j) = C(s_i, s_j|\theta) + \tau^2\delta(s_i, s_j),$$

express the covariance function $C(\cdot,\cdot|\theta)$ as $\sigma^2 R(\cdot,\cdot|\phi)$, where σ^2 is the marginal variance and R is the correlation function parameterized by the range ϕ, i.e., $\theta = \{\sigma^2, \phi\}$, and rewrite $\tau^2 = \alpha\sigma^2$, then

$$\Sigma = \Sigma(s_i, s_j) = \sigma^2 R(s_i, s_j|\phi) + \alpha\sigma^2\delta(s_i, s_j)$$

$$= \sigma^2(R(s_i, s_j|\phi) + \alpha\delta(s_i, s_j)).$$

This implies that the MCMC-free conjugate NNGP marginal model is

$$Y \sim N(X\beta, \sigma^2\tilde{M}), \tag{9}$$

where $\tilde{M} = \tilde{M}(\cdot,\cdot|\phi,\alpha)$ is a known covariance matrix once ϕ and α are fixed (Finley et al., 2021). The fixed values of ϕ and α are either chosen based on a variogram or, more formally, selected using K-fold cross-validation on hold-out data, which leaves β and σ^2 the only unknown parameters (Finley et al., 2021).

Normal-Inverse-Gamma prior for (β, σ^2) leads to conjugate Normal-Inverse-Gamma posterior distributions, and other summary quantities of β and σ^2 can easily and exactly be obtained (Finley et al., 2021). That is, for fixed ϕ and α, the conjugate Bayesian linear regression model can be constructed as

$$IG(\sigma^2|a_\sigma, b_\sigma) \times N(\beta|\mu_\beta, \sigma^2 V_\beta) \times N(y|X\beta, \sigma^2\tilde{M})$$

with joint posterior distribution

$$p(\beta, \sigma^2|y) \propto IG(\sigma^2|a^*_\sigma, b^*_\sigma) \times N(\beta|B^{-1}b, \sigma^2 B^{-1})$$

$$= p(\sigma^2|y) \times p(\beta|\sigma^2, y),$$

where

$$a^*_\sigma = a_\sigma + n/2, \quad b^*_\sigma = b_\sigma + \frac{1}{2}(\mu_\beta^T V_\beta^{-1}\mu_\beta + y^T\tilde{M}y - b^T B^{-1} b)$$

and

$$B = V_\beta^{-1} + X^T\tilde{M}^{-1}X, \quad b = V_\beta^{-1}\mu_\beta + X^T\tilde{M}^{-1}y$$

(Finley et al., 2019).

Marginal posterior distributions for β and σ^2 are respectively

$$\beta|y \sim MVS_{t_{2a^*_\sigma}}(B^{-1}b, \frac{b^*_\sigma}{a^*_\sigma}B^{-1})$$

and

$$\sigma^2|y \sim IG(a^*_\sigma, b^*_\sigma),$$

where $MVS_{t_\kappa}(B^{-1}b, (b/a)B^{-1})$ denotes the multivariate noncentral Student's t distribution with degrees of freedom κ, mean $B^{-1}b$ and variance $bB^{-1}/(a-1)$. The marginal posterior mean and variance for σ^2 are $b^*_\sigma/(a^*_\sigma - 1)$ and $b^{2*}_\sigma/(a^*_\sigma - 1)^2(a^*_\sigma - 2)$, respectively. However, instead of sampling from the posterior distribution (i.e., sample σ^2 and then sample β one-for-one for each σ^2 drawn), a fast evaluation of the marginal posterior distributions are implemented (Finley et al., 2019).

The conjugate NNGP model is MCMC-free, and it involves (1) fixing ϕ and α and splitting data into K folds, (2) after removing the k^{th} fold of the data, obtaining posterior means for β and σ^2, (3) predicting posterior means of $y[S(k)]$, (4) calculating RMSPE (root mean square predictive error) over K folds, and (5) performing cross-validation to choose ϕ and α. Estimation involves repeating step (2) with $(\phi_0, \alpha_0)^T$ and the full data to get $(\beta, \sigma^2)|y$, whereas predication involves repeating step (3) with $(\phi_0, \alpha_0)^T$ and the full data to obtain mean and variance of $y(s_o)|y$ at a new location s_0. Interested readers are referred to **Algorithm 5** in *Section 2.3* of Finley et al. (2019) for additional algorithmic details.

3.4. Fitted values, replicate data for model checking and evaluation, and predictions

For the latent and response NNGP model, once MCMC chains converge, posterior samples can be used to generate various quantities of interest such as fitted values, replicate data at observed location, and predictions at new locations (Finley et al., 2021).

For the latent model and response model, the posterior distributions of the fitted values is specified by the samples $\{x(s_i)^T\beta^{(l)} + w(s_i)^{(l)}\}$ and $\{x(s_i)^T\beta^{(l)}\}$, respectively, where $\beta^{(l)}$ is the l^{th} post burn-in sample. For the MCMC-free conjugate model, the closed-form expressions for the posterior means for β and σ^2 for the fitted value are in **Algorithm 5** in Finley et al. (2019).

The algorithm for prediction for the latent, response, and MCMC-free conjugate model is presented in **Algorithm 2**, **4**, and **5**, respectively, in Finley et al. (2019).

3.5. Software

Functions in the **R** package `spNNGP` is written in **C/C++** via **R**'s foreign language interface to call **Fortran** Basic Linear Algebra Subprograms (`BLAS`; www.netlib.org/blas), e.g., `openBLAS` (Zhang, 2016), and Linear Algebra Package (`LAPACK`; www.netlib.org/lapack) for computationally intensive matrix operations and uses **openMP** (Dagum & Menon, 1998) for parallelization (Finley et al., 2019; Finley et al., 2021).

4. Applications to Spatial data

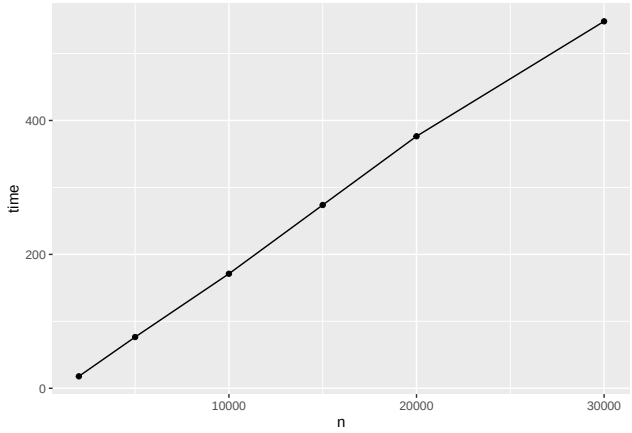
4.1. Simulated data

The simulated data sets are simulated using the exampled codes provided in the **R** documentation page of the `spConjNNGP` function in the package `spNNGP`.

We follow the model in (1) and simulate data with varying sample size n ($n = 10, 100, 1000, 10000, 20000, 30000$). For these data sets, only the conjugate NNGP model is considered.

We first note that the simulation part, rather than the model fitting part, takes much longer than expected, and we were not able to simulate pass $n = 30000$ on a personal laptop because of the dense and denser matrices that are involved. We also note that enabling **openMP** on a macOS laptop could take weeks, and even when being enabled, it did not provide much computational advantages. As a result, the results for this section were obtained running **R** locally on a personal laptop.

Even with these technical limitations, the run time for model fitting as a function of sample size is approximately linear, which is an improvement since the run time for GP usually grows in cubic time $\mathcal{O}(n^3)$ for data size n.



4.2. Real data

The `BCEF` data set contains forest canopy heights (FCHs) from NASA Goddard's LiDAR Hyperspectral and Thermal (G-LiHT; Cook et al. 2013) Airborne Imager and percent tree cover (PTC; Hansen et al. 2013) over a subset of Bonanza Creek Experimental Forest in Alaska in Summer 2014. The `BCEF` data can be accessed from the **R** package `spNNGP`.

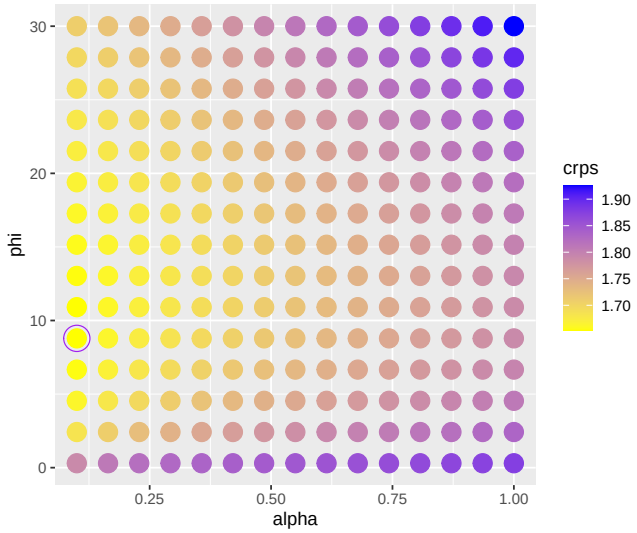
For this data set, a latent NNGP model, a response NNGP model, and a conjugate NNGP model are fitted, but the focus is again on the conjugate model. Since the data set is rather big ($n = 188,717$ locations) and **openMP** did not provide much computational advantages even when being enabled on a macOS laptop, the models are directly ran on a hpc cluster operated by OSU's College of Engineering.

We first note that compiling with **OpenMP** support and using 40 thread(s), the MCMC-based latent model took about 30 minutes to run, whereas the MCMC-based response model took a bit under 30 minutes to run. The MCMC-free conjugate model took less than 2 minutes to run.

We were able to obtain results that are similar to the paper results except for ϕ, but we note that the ϕ in the table below is the ϕ in the fixed combination (ϕ and α).

rowname	estimate	lb	ub
b_0	10.14	10.1	10.19
b_PTC	0.03621	0.0362	0.03621
sigma_2	34.73	34.7	34.75
phi	8.786	8.786	8.786
alpha	0.1	0.1	0.1
tau_2	3.473	3.47	3.475

The "optimal" combination (circled in the plot below) based on crps (continuous ranked probability score) metric is similar to the paper values. In our analysis, $\phi = 8.785714$ and $\alpha = 0.1$, but the paper does not provide specfic values. The general pattern of the crps plot is similar too. However, crps range of our plot differs. We suspect that the authors made an error since their crps value in Table 4 is 1.53.



Lastly, let us turn to model diagnostics. Again, `GPD` and `GRS` (alternative to `DIC` and `WAIC`) values are similar to paper values.

names	values
G	6490429
P	3499399
D	9989828
GRS	-541985

5. Discussion

(1) This project reviews the NNGP models formation and replicates results in Finley et al. (2021) using the `spNNGP` **R** package and hpc cluster at OSU. However, algorithmic details as well as where and how parallelization is used in each algorithm are not well explored but can be referred to Finley et al. (2019) for more.

(2) NNGP models assume Gaussian response. Exploiting the Pólya-Gamma data-augmented sampler of Polson et al. (2013), the latent NNGP model can be extended to model non-Gaussian (i.e. binary) response (Finley et al., 2021). Future study could look into extending the latent NNGP model or consider alternative models such as the GGP (generalized Gaussian process) model (Zilber & Katzfuss, 2021) via Vecchia-Laplace approximation or NNMP (nearest neighbor mixture process) model (Zheng, et al., 2022; Zheng, 2022) for other non-Gaussian response such as count, zero-inflated count, zero-inflated semicontinuous, etc. data.

(3) For the simulated data sets, running the conjugate NNGP model on a personal laptop, we show that run time as a function of sample size is approximately linear. Future study should attempt to simulate larger data sets, run all NNGP models on a hpc cluster, and record and compare their run times.

(4) For the real data set, running MCMC-based NNGP models on a hpc cluster, we were able to obtain results about or a little under 30 minutes. Running MCMC-free conjugate NNGP model took less than 2 minutes. This is a significant improvement since MCMC-based methods can take days up to weeks to run. Future study should compare the results of the MCMC-based NNGP models and investigate discrepancies of the results of the conjugate NNGP model.

Reference

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association, 111(514), 800-812.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Appendix A1 of Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association, 111(514), 800-812.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Appendix A2 of Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association, 111(514), 800-812.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., & Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. Journal of Computational and Graphical Statistics, 28(2), 401-414.

Finley, A. O., Datta, A., & Banerjee, S. (2021). spNNGP R package for nearest neighbor Gaussian process models. arXiv preprint arXiv:2001.09111.

Salakhutdinov, R. (2010) Approximate Inference using MCMC. MIT. https://www.mit.edu/~9.520/spring10/Classes/class21_mcmc_2010.pdf

Zheng, X., Kottas, A., & Sansó, B. (2022). Nearest-Neighbor Geostatistical Models for Non-Gaussian Data. arXiv preprint arXiv:2107.07736.

Zheng, X. (2022). A Modeling Framework for Non-Gaussian Spatial and Temporal Processes (Doctoral dissertation, UC Santa Cruz).

Zilber, D., & Katzfuss, M. (2021). Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. Computational Statistics & Data Analysis, 153, 107081.