Computational Methods for Large Spatial Data

Reetam Majumder

STAT 625, UMBC

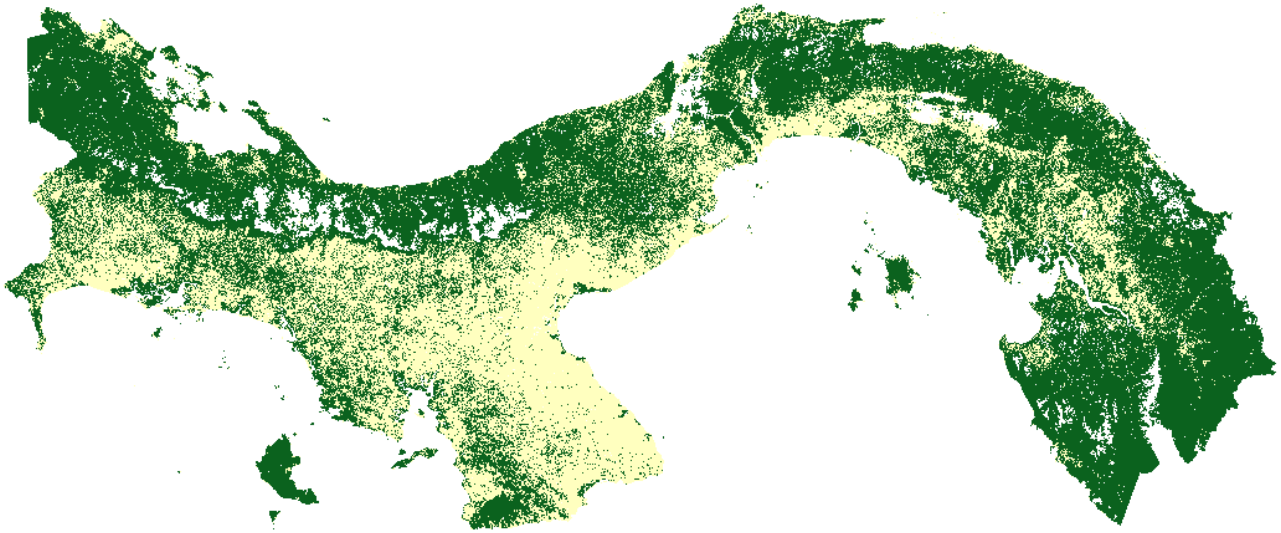
March 31, 2023

NC STATE UNIVERSITY

Overview of Kriging

Motivating example

- LANDSAT and MODIS are satellites which provide optical information of the planet
- What they actually 'measure' is spectral and thermal data - affected by cloud cover etc



- A common statistical problem is to make predictions at unobserved locations

The spatial model

- Let Y_i be a measure of NDVI, a greenness metric used to monitor changes in land use (e.g., urbanization, agriculture, fires)
- Y_i is observed at locations $\mathbf{s}_i = (s_{i1}, s_{i2}), i = 1 : n$.
- $\mathbf{X}_i$ are p covariates at location i - e.g., elevation data.
- A standard spatial model representation is $Y_i = \mu_i + Z_i + \epsilon_i$
- $\mu_i = \mathbf{X}\beta$; similar to linear regression
- There are 2 error terms:
  - $\epsilon_i \overset{iid}{\sim} N(0, \tau^2)$; called the **nugget**
  - Z_i is mean 0, spatially correlated
- Z_i captures spatial correlation not explained by $\mathbf{X}$

Spatial covariance

- $E(Y_i) = \mu_i$
- Z_i is independent of ϵ_j for all (i, j) pairs, and so:
  - $\Sigma_{ii}(\theta) := V(Y_i) = \sigma^2 + \tau^2$
  - $\Sigma_{ij}(\theta) = Cov(Y_i, Y_j) = \sigma^2 \rho(d_{ij}, \phi)$
- d_{ij} is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, ϕ is the spatial range
- Common forms for $\rho(\cdot)$ include exponential and squared exponential, and Matern.
- We'll denote the coviance matrix as $\Sigma(\theta)$; dimensions = $n \times n$
- Stationarity and isotropy are common assumptions - strong, but often necessary

Kriging

- Given all this we want to predict $\hat{Y}_0$ at $\mathbf{s}_0$
- Ideally, some uncertainty quantification (standard deviation, prediction interval etc)
- Kriging just assumes a constant mean, and known covariance
- Gaussian data is not necessary, but it makes things easier
- The 'optimal' prediction is given by

$$\hat{Y}_0 = \mu_0(\hat{\beta}) + \Sigma_0(\hat{\theta})\Sigma(\hat{\theta})^{-1}\{\mathbf{Y} - \mu(\hat{\beta})\}$$

- Inverting $\Sigma(\hat{\theta})$ takes $\mathcal{O}(n^3)$ computational cost and $\mathcal{O}(n^2)$ storage
- Panama has $\sim 1.7 \times 10^7$ observed pixels

This is a major bottleneck.

Dealing with Large Datasets

The Gaussian process

- This is pretty ubiquitous in Bayesian literature
- Data is observed at fixed spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. The joint distribution of the data is multivariate Normal
- The underlying process happens everywhere
- The multivariate Normal is then just a finite-valued subset of an infinite dimensional Gaussian process (GP) [1,2]
- Observations at any location is univariate Normal; observations at any subset of locations is multivariate Normal
- The GP is parameterized by a mean function $m(\cdot)$ and a covariance function $C(\cdot, \cdot)$

$$m(Y(\mathbf{s}_i)) = E(Y(\mathbf{s}_i))$$
$$C(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_i))$$

Some common approaches

- Brian's class notes
- The Vecchia approximation[1] has taken off again in recent years with the proliferation of large datasets in environment, ecology, epidemiology etc.

[1] Wiki article

The Vecchia approximation

- Let $y_1, \ldots, y_n$ be an ordered set of random variables
- For any ordering, you can express their joint distribution as

$$f(y_1, \ldots, y_n; \theta) = f(y_1; \theta) \prod_{i=2}^{n} f(y_i | y_{i-1}, \ldots y_1; \theta)$$

- For every $y_i, i > 1$, consider the set $\mathcal{N}_i \subset \{1, \ldots, i-1\}$
- The Vecchia approximation is

$$f(y_1, \ldots, y_n; \theta) \approx f(y_1) \prod_{i=2}^{n} f(y_i | y_{(i)}; \theta),$$

  where $y_{(i)} = \{y_j; j \in \mathcal{N}_i\}$
- $\mathcal{N}_i$ is often called the Vecchia neighbor set; $|\mathcal{N}_i| \leq m$

How to order? How to choose m?

Simplifying the precision matrix

- $\Omega(\theta) = \Sigma(\theta)^{-1}$ is defined as the precision matrix. Sparsity of the precision matrix simplifies computations
- Consider the following Vecchia approximation

$$\begin{aligned} f(y_1, \ldots, y_5) &= f(y_1) f(y_2|y_1) \ldots f(y_4|y_3, y_2, y_1) \\ &\approx f(y_1) f(y_2|y_1) \ldots f(y_5|y_4) \end{aligned}$$

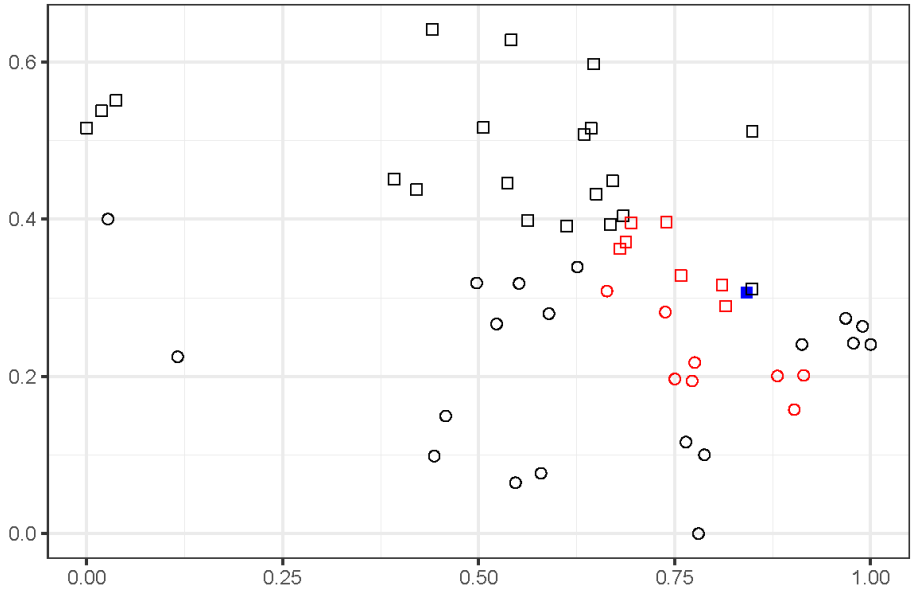
- This elicits a sparse precision matrix proportional to

$$\begin{bmatrix} 1 & k_{12} & 0 & 0 & 0 \\ & 1 & k_{23} & 0 & 0 \\ & & 1 & k_{34} & 0 \\ & & & 1 & k_{54} \\ & & & & 1 \end{bmatrix}$$

- The (structural) sparsity makes Cholesky decompositions easier
- Working with a Vecchia approximated process has $\mathcal{O}(nm^3)$ computational cost and needs $\mathcal{O}(nm^2)$ storage
- In practice, $m << n$

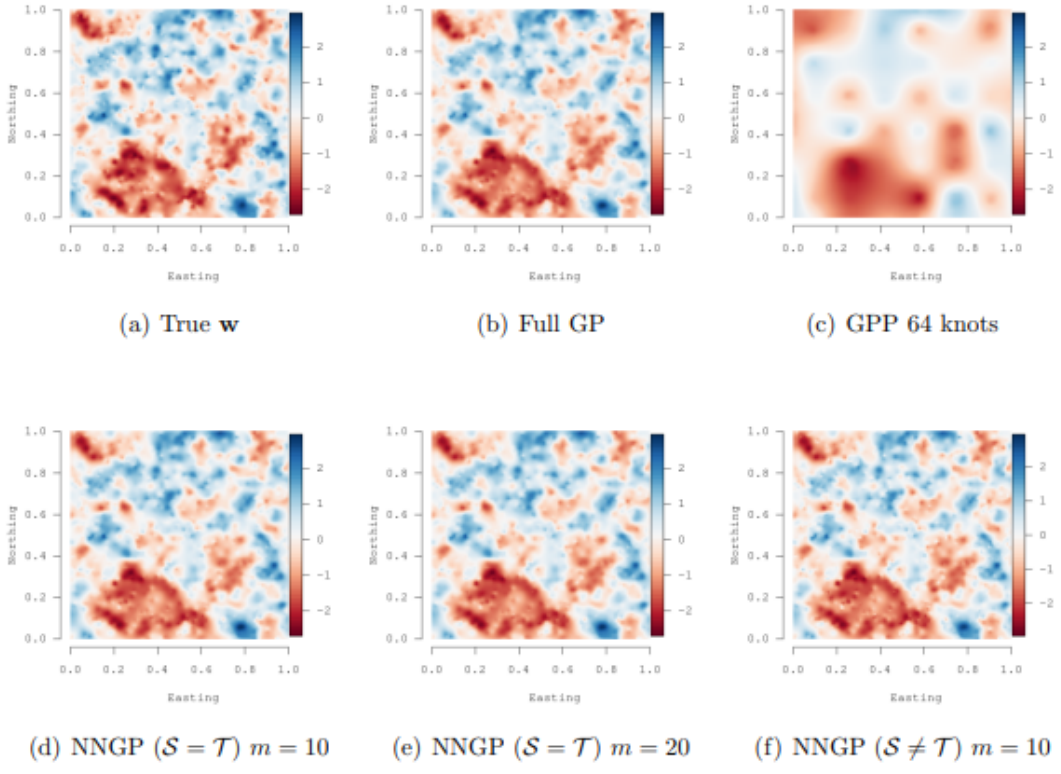
Use in spatial modeling

- The approximation is usually applied to $\{Z_i\}$ and not $\{Y_i\}$
- In its simplest form, the ordering is done based on some coordinate system



- m is often the set of nearest neighbors
- The general consensus is that for processes modeling mean behavior, the approximation is more sensitive to $|\mathcal{N}|$ than the ordering of locations

An example



(a) True $\mathbf{w}$ (b) Full GP (c) GPP 64 knots

(d) NNGP $(\mathcal{S} = \mathcal{T})$ $m = 10$ (e) NNGP $(\mathcal{S} = \mathcal{T})$ $m = 20$ (f) NNGP $(\mathcal{S} \neq \mathcal{T})$ $m = 10$

Image from Datta et al (2016)

Using machine learning

- There are connections between neural networks and GPs, which has led to some interesting methodology and applications
- Harris et al (2022) do neural network GP regression (NN-GPR) for climate modeling; the covariance function is based on an infinitely wide neural network
- Sauer et al (2022) propose deep Gaussian processes (DGP) where the covariance functions are themselves modeled using nested GPs (like NN layers)
- Chen et al (2020) directly get predictions using a neural network; alongside lat-long, they add basis functions which have spatial information, essentially doing kriging using a NN

Spatial Extremes

Spatial extremes

- Consider extreme events in streamflow, wildfires, storms.
- For example, a spatial field of annual maximum rainfall
- Data is scarce, and spatial dependence is often not in the mean
- Let $f(y_1, y_2)$ be the joint density of such a spatial process at locations $\mathbf{s}_1$ and $\mathbf{s}_2$
- Let u_1 and u_2 be the marginal CDFs, i.e., $u_i := F_i(y_i)$
- F_i are usually extreme value distribution functions
- What is of interest, then, are questions like:

As the process becomes extreme at $\mathbf{s}_1$, will it also be extreme at $\mathbf{s}_2$?

Asymptotic (in)dependence

- A measure of extremal (tail) dependence commonly used is

$$\chi_u(\mathbf{s}_1, \mathbf{s}_2) = P(u_1 > u | u_2 > u)$$

  for high quantile levels u
- For GPs, $\chi_u \to 0$ as $u \to 1$. This is called asymptotic independence
- For extreme value processes like max-stable processes, $\chi_u \to c$ for $c > 0$ as $u \to 1$. This is called asymptotic dependence

The need for spatial extremes models

- Individual data points are either maxima (max-stable processes) or peaks over a threshold (generalized Pareto process)
- These are scarce by definition; If we have 100 years of temp data, that is 100 data points of annual maximum temp
- They are more scarce at its extremes! there is exactly 1 data point above the 99^{th} percentile of annual maxima data
- What is the probability that it will be hotter this year compared to 2022? That is, what is $P[Tmax_{2023} > Tmax_{2022}]$?
- We want inference for these extreme quantiles
- Computationally very challenging; e.g., the full likelihood for the MSP can be written down only for around 13 locations (Castruccio et al, 2016)

Recent work

- Huser and Wadsworth (2022) is a great read for recent advances
- A lot of literature focuses on computational challenges
- Huser et al (2022) studied Vecchia approximation for spatial extremes
- There has been recent work on using neural networks with a few different approaches - see e.g. Sainsbury-Dale et al (2022), Richards and Huser (2022), Majumder et al (2022)
- In terms of applications, Zhang et al (2022) looked at extreme precipitation in the central US, Majumder and Reich (2022) looked at extreme streamflow for the same region
- Richards and Huser (2022) studied extreme wildfire risk across the US; Bercos-Hickey et al (2022) looked at the Pacific North West heat wave of 2021

This is an active research area in methodology and in applications

Questions?

- If you're interested in my research, I try to keep my website updated
- Feel free to reach out through email or LinkedIn