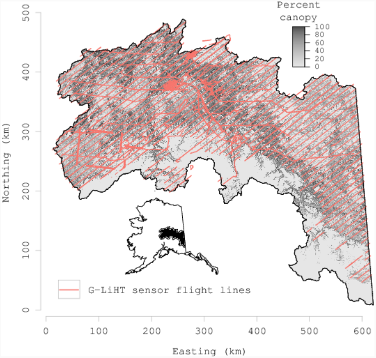
Conjugate Bayesian Models for Massive Spatial Data

Andrew Finley[1] & Jeffrey Doser[2]

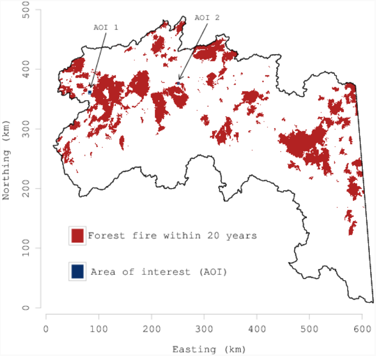
May 15, 2023

[1]Department of Forestry, Michigan State University.
[2]Department of Integrative Biology, Michigan State University.

Case Study: Alaska Tanana Valley Forest Height Dataset



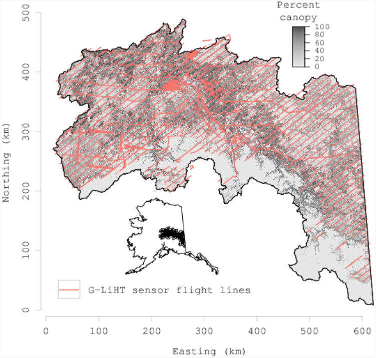
Forest height and tree cover



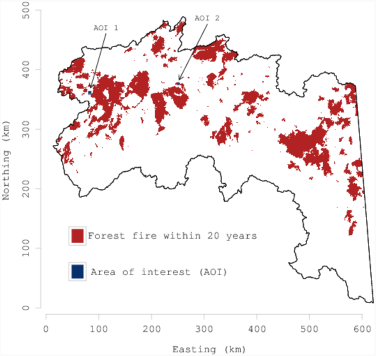
Forest fire history

- Forest height (red lines) data from LiDAR at 5×10^6 locations
- Knowledge of forest height is important for biomass assessment, carbon management etc

Case Study: Alaska Tanana Valley Forest Height Dataset



Forest height and tree cover



Forest fire history

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

Analyzing the data

Models used:

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

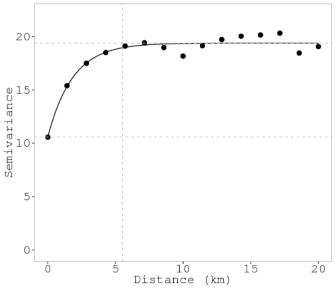


Figure: Variogram of the residuals from non-spatial regression indicates strong spatial pattern

NNGP models

- Collapsed NNGP:
    - $y_{FH}(s) = \beta_0 + \beta_{tree} x_{tree} + \beta_{fire} x_{fire} + w(s) + \epsilon(s)$
    - $w(s) \sim NNGP(0, C(\cdot, \cdot \,|\, \sigma^2, \phi))$
    - $y_{FH} \sim N(X\beta, \tilde{C} + \tau^2 I)$ where $\tilde{C}$ is the NNGP covariance matrix derived from C
- Response NNGP:
    - $y_{FH}(s) \sim NNGP(\beta_0 + \beta_{tree} x_{tree} + \beta_{fire} x_{fire}, \Sigma(\cdot, \cdot \,|\, \sigma^2, \phi, \tau^2))$
    - $y_{FH} \sim N(X\beta, \tilde{\Sigma})$ where $\tilde{\Sigma}$ is the NNGP covariance matrix derived from $\Sigma = C + \tau^2 I$

NNGP models

	Non-spatial regression	Collapsed NNGP	Response NNGP
CRPS	2.3	0.86	0.86
RMSPE	4.2	1.73	1.72
CP	93%	94%	94%
CIW	16.3	6.6	6.6

Table: Model comparison metrics for the Tanana valley dataset

- NNGP models perform significantly better than the non-spatial model
- MCMC run time for the NNGP models:
  - Collapsed model: 319 hours
  - Response model: 38 hours
- For massive spatial data, full Bayesian output for even NNGP models require substantial time

Another look at the response model

- Original full GP model: $y(s) \stackrel{ind}{\sim} N(x(s)'\beta + w(s), \tau^2)$
- $w(s) \sim GP$ with a stationary covariance function $C(\cdot, \cdot \,|\, \sigma^2, \phi)$
- $Cov(w) = \sigma^2 R(\phi)$
- Full GP model: $y \sim N(X\beta, \Sigma)$ where $\Sigma = \sigma^2 M$
- $M = R(\phi) + \alpha I$
- $\alpha = \tau^2/\sigma^2$ is the ratio of the noise to signal variance
- Response NNGP model: $y \sim N(X\beta, \tilde{\Sigma})$
- $\tilde{\Sigma} = \sigma^2 \tilde{M}$ where $\tilde{M}$ is the NNGP approximation for M

Conjugate NNGP

- $y \sim N(X\beta, \sigma^2 \tilde{M})$
- If ϕ and α are known, M, and hence $\tilde{M}$, are known matrices
- The model becomes a standard Bayesian linear model
- Assume a *Normal Inverse Gamma (NIG)* prior for $(\beta, \sigma^2)'$
- $(\beta, \sigma^2)' \sim NIG(\mu_\beta, V_\beta, a_\sigma, b_\sigma)$, i.e., $\beta \,|\, \sigma^2 \sim N(\mu_\beta, \sigma^2 V_\beta)$ and $\sigma^2 \sim IG(a_\sigma, b_\sigma)$

Conjugate NNGP

- $y \sim N(X\beta, \sigma^2 \tilde{M})$, $\tilde{M}$ is known

Joint likelihood:

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

Conjugate NNGP

- $y \sim N(X\beta, \sigma^2 \tilde{M})$, $\tilde{M}$ is known

Joint likelihood:

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

- Conjugate posterior distribution
  $(\beta, \sigma^2) \,|\, y \sim NIG(\mu_\beta^*, V_\beta^*, a_\sigma^*, b_\sigma^*)$
- Expressions for μ_β^*, V_β^*, a_σ^* and b_σ^* can be calculated in $O(n)$ time

Conjugate NNGP

- $(\beta, \sigma^2) \,|\, y \sim NIG(\mu_\beta^*, V_\beta^*, a_\sigma^*, b_\sigma^*)$
- Marginal posterior: $\beta \,|\, y \sim MVt_{2a_\sigma^*}(\mu_\beta^*, \frac{b_\sigma^*}{a_\sigma^*} V_\beta^*)$
- $MVt_k(m, V)$ is the *multivariate t* distribution with degrees of k, mean m and scale matrix V
- $E(\beta \,|\, y) = \mu_\beta^*,\ Var(\beta \,|\, y) = \frac{b_\sigma^*}{a_\sigma^* - 1} V_\beta^*$
- Marginal posterior: $\sigma^2 \,|\, y \sim IG(a_\sigma^*, b_\sigma^*)$
- $E(\sigma^2 \,|\, y) = \frac{b_\sigma^*}{a_\sigma^* - 1},\ Var(\sigma^2 \,|\, y) = \frac{b_\sigma^{*2}}{(a_\sigma^* - 1)^2 (a_\sigma^* - 2)}$
- Exact posterior distributions of β and σ^2 are available

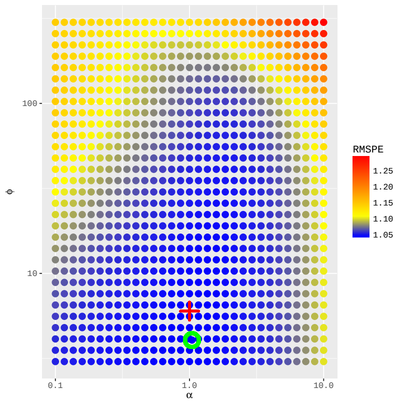
Predictive distributions

- $y(s) \,|\, y \sim t_{2a_\sigma^*}(m(s), \frac{b_\sigma^*}{a_\sigma^*} v(s))$
- $E(y(s) \,|\, y) = m(s),\; Var(y(s) \,|\, y) = \frac{b_\sigma^*}{a_\sigma^* - 1} v(s)$
- $m(s)$ and $v(s)$ can be computed using $O(m)$ flops
- Exact posterior predictive distributions of $y(s) \,|\, y$ for any s
- No MCMC required for parameter estimation or prediction

Choosing α and ϕ

- ϕ and α are chosen using K-fold cross validation over a grid of possible values
- Unlike MCMC, cross-validation can be completely parallelized
- Resolution of the grid for ϕ and α can be decided based on computing resources available
- In practice, a reasonably coarse grid often suffices

Choosing α and ϕ



RMSPE

Figure: Simulation experiment: True value ($+$) of (α, ϕ) and estimated value ($\circ$) using 5-fold cross validation

Scalability

- Computation and storage requirements are $O(n)$
- One evaluation time similar to the response NNGP model
- Unlike response NNGP, does not involve any serial MCMC iterations
- For K fold cross validation and G combinations of ϕ and α, total number of evaluations is KG
- Embarassingly parallel: Each of the KG evaluations can proceed in parallel

Scalability

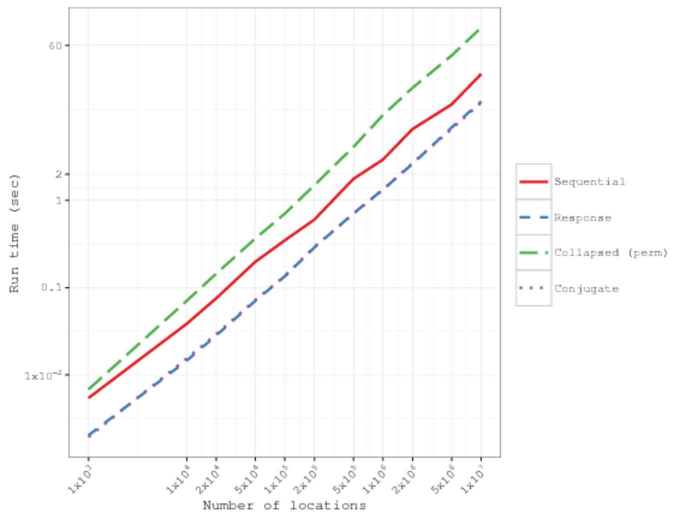


Figure: Run times of different NNGP models with increasing sample size

Alaska Tanana Valley dataset

	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

Summary

- MCMC free exact Bayesian approach by fixing some covariance parameters
- Conjugate posterior distributions of the parameters and posterior predictive distributions available in closed form
- Embarassingly parallel cross validation to identify best choices for fixed parameters
- Runs in seconds for massive spatial dataset with millions of locations
- Available in the spNNGP package in R