A Dynamic Space-Time Regression Model of Plot-Level Forest Biomass Data in Maine, USA

Elliot Shannon, Jeff Doser, Abhirup Datta, Andrew O. Finley

Department of Forestry
Department of Statistics and Probability
Michigan State University

Joint Statistical Meetings, August 6, 2023
Toronto, Ontario, Canada

Motivation

- The Forest Inventory and Analysis (FIA) program of the US Forest Service has systematically collected data from forest inventory plots across the US for more than 20 years.
- These data include estimates of forest biomass, which is an important metric for monitoring forest carbon and climate change.
- FIA plots are measured on a rotating basis every 5 years, creating a spatio-temporal dataset with both observed and unobserved data.

Motivation

Plot	1998	1999	2000	2001	2002	2003	2004	2005
1					20.1			
2	30.8					32.4		
3				19.5				
4		36.2					35.0	
5			29.4					28.3

Objectives

- Develop a dynamic space-time regression model that:
  - estimates plot-level forest biomass data for FIA plots in Maine, USA with associated uncertainty quantification for all plot locations and years,
  - incorporates wall-to-wall covariates (i.e. remotely sensed data),
  - dynamically characterizes the spatial dependence between FIA biomass measurements over time,
  - utilizes an efficient low-rank approximation for the spatial process (i.e. Predictive Process).

Model Notation

- let the data be observed at locations $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ and at time points $1, 2, \ldots, T$
- we assume that there is missing/unobserved data
- $y_t(s)$ denotes the plot-level biomass at location s and time t
- $z(s)$ is the $q \times 1$ time independent fixed effect covariate (i.e. average climate, etc.) observed at location s, γ is the corresponding $q \times 1$ coefficient vector
- $x_t(s)$ is the $p \times 1$ time dependent fixed effect covariate (i.e. NDVI, etc.) observed at location s at time t, β_t is the corresponding $p \times 1$ coefficient vector
- also let $I_t(s)$ denote the indicator variable denoting if the data is observed at location s at time t

Model Definition

The model is defined as

$$y_t(s) = z(s)'\gamma + x_t(s)'\beta_t + u_t(s) + \epsilon_t(s), \epsilon_t(s) \overset{\text{iid}}{\sim} N(0, \tau_t^2) \text{ if } I_t(s) = 1$$

$$\beta_t = \beta_{t-1} + \eta_t, \eta_t \overset{\text{iid}}{\sim} N(0, \Sigma_\eta), \beta_0 \sim N(m_0, \Sigma_0), \gamma \sim N(m, \Sigma)$$

$$u_t(s) = u_{t-1}(s) + w_t(s), w_t(s) \overset{\text{iid}}{\sim} GP(0, C_t(\cdot, \theta_t))$$

$$u_0(s) = 0, \text{ for } t = 1, 2, \ldots, T \text{ and generic location } s$$

with hyper priors: $\tau_t^2 \overset{\text{iid}}{\sim} IG(a_\tau, b_\tau), \Sigma_\eta \sim IW(r_\eta, \Omega_\eta), \theta_t \overset{\text{iid}}{\sim} p_t(\theta_t)$

Predictive Process and NNGP

- the Guassian process used to model the spatial random effects can be replaced by a lower dimensional approximation to ease computation
- the predictive process model (motivated from kriging) approximates the parent (full) process through a linear combination of 'knots', forming a low rank covariance matrix
- the nearest neighbor Gaussian process (NNGP) model generates a sparse covariance matrix through a 'screening effect' by considering m nearest neighbors

Model Implementation

We assume an exponential spatial covariance function of the form

$$C_t(s_1, s_2; \theta_t) = \sigma_t^2 exp(-\phi_t ||s_1 - s_2||)$$

where $\theta_t = \{\sigma_t^2, \phi_t\}$ with spatial variance σ_t^2 and spatial decay ϕ_t

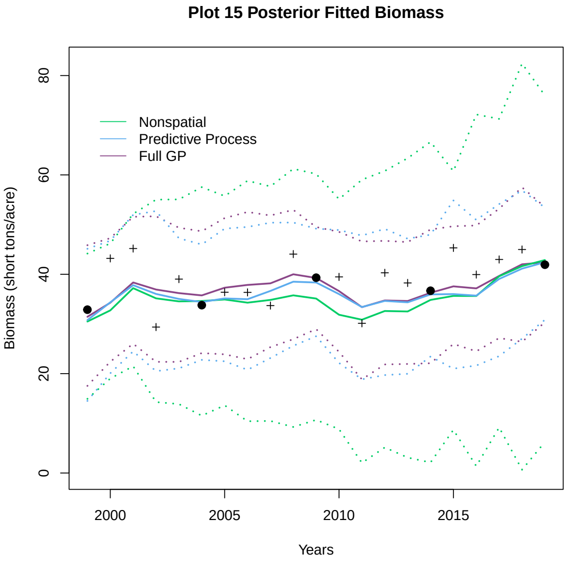
Model Implementation

- these prior specifications lead to a well-defined Bayesian hierarchical model
- Bayesian inference proceeds through a Gibbs sampling algorithm with random-walk Metropolis steps for spatial covariance parameters
- full GP and predictive process models are called from the `spBayes` R package
- samplers are written in `C++` using `LAPACK` and `BLAS` linear algebra subroutines

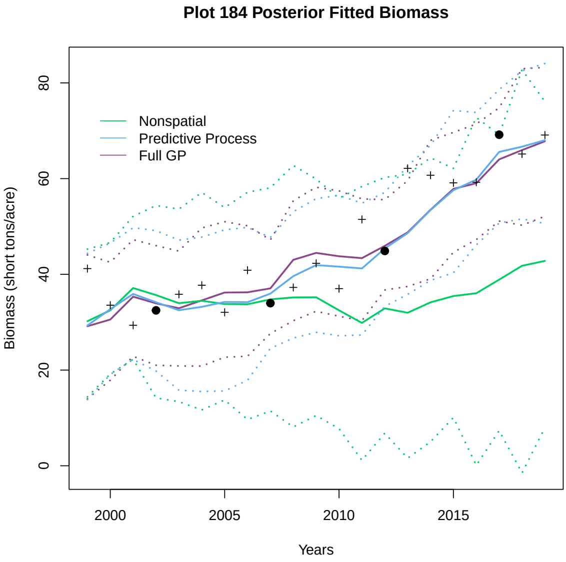
Simulation

- single temporally-varying intercept
- $T = 21$ years
- $n = 400$ locations (single county in Maine, USA)
- 6,795 of 8,400 plot-year combinations (80.9%) are unobserved
- consider nonspatial, predictive process (25 knots), and full GP models
- parameter estimates summarized from 2,500 post burn-in posterior samples

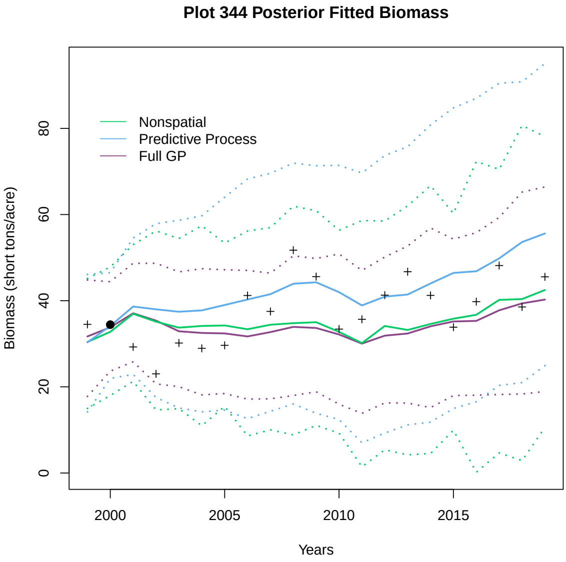
Preliminary Results



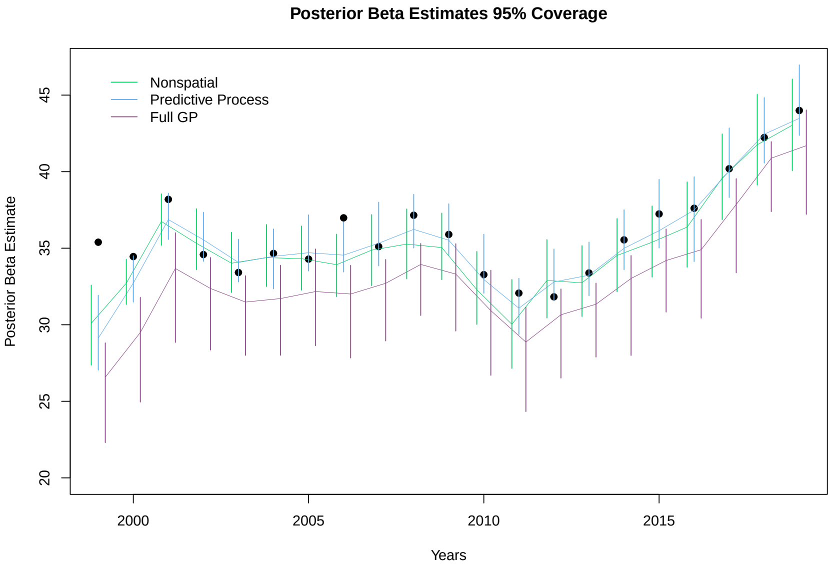
Preliminary Results



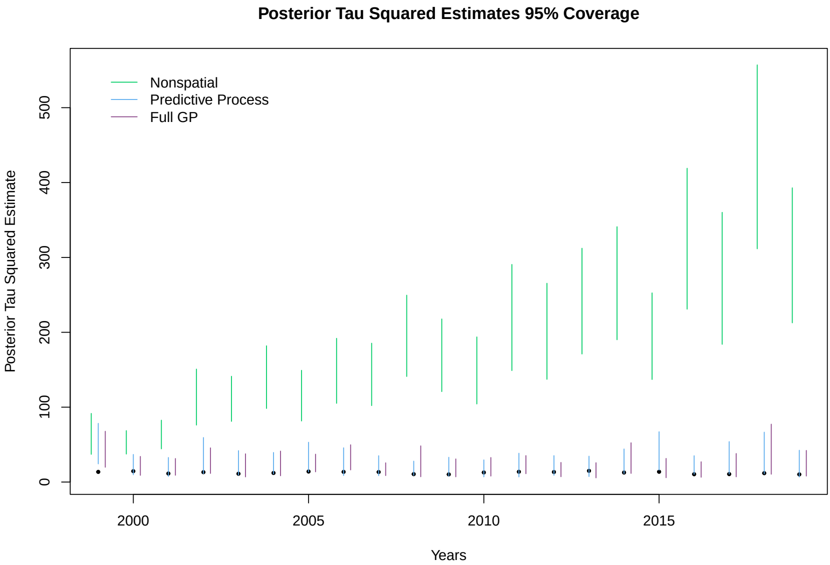
Preliminary Results



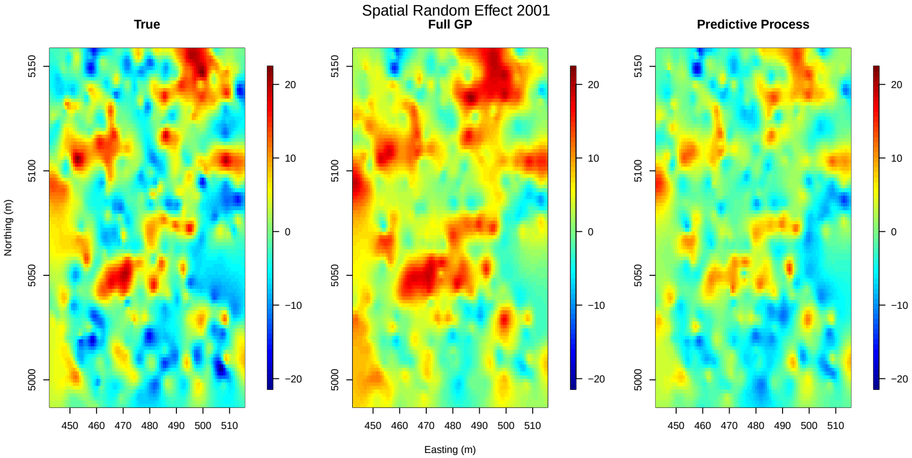
Preliminary Results



Preliminary Results



Preliminary Results



Application Data

- FIA plot-level biomass measurements
  - collected from 1999 - 2019
  - $n = 2{,}713$ plot locations
  - 10,833 total observations
- Landsat Normalized Difference Vegetation Index (NDVI)
  - annual composite images from 1999 - 2019
  - quantifies vegetation greenness and represents vegetation density
- LandTrendr
  - spectral-temporal segmentation algorithm for detecting change from satellite imagery (Landsat)
  - identifies and quantifies forest disturbance from annual image composites

Next Steps

- make predictions of forest biomass at arbitrary locations and times within the spatio-temporal domain
- implement NNGP for spatial random effect

References

Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A. S., and Schaap, M. (2016). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *The Annals of Applied Statistics*, 10(3):1286 – 1316.

Finley, A. O., Banerjee, S., and Gelfand, A. E. (2012). Bayesian dynamic modeling for large space–time datasets using gaussian predictive processes. *Journal of Geographical Systems*, 14:29 – 47.

Finley, A. O., Banerjee, S., and Gelfand, A. E. (2015). spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, 63(1):1 – 28.

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16(5):465 – 479.