Geostatistical model-based estimation of forest biomass in interior Alaska combining landsat derived tree cover, sampled airborne lidar and field observations

Chad Babcock[1], Andrew O. Finley[2], Hans-Erik Andersen[3], Bruce D. Cook[4], Douglas C. Morton[4]

[1]University of Washington, [2]Michigan State University, [3]USFS FIA Program, [4]NASA Goddard Space Flight Center

Abstract

Lidar data provide critical information on the three-dimensional structure of forests. However, collecting wall-to-wall laser altimetry data at regional and global scales is cost prohibitive. As a result, studies employing lidar data for large area estimation typically collect data via strip sampling; leaving large swaths of the forest domain unmeasured by the instrument. The goal of this research was to develop and examine the performance of a geostatistical coregionalization modeling approach for combining field inventory measurements, incomplete coverage airborne lidar, and Landsat-based remote sensing data products to predict aboveground biomass (AGB) in interior Alaska's Tanana Valley. The proposed spatial modeling strategy facilitates point-level predictive mapping of AGB density and provides a statistically sound approach for estimating total AGB for areal units. This research is keenly focused on the appropriate characterization of prediction uncertainty in the form of posterior predictive coverage intervals and standard errors. Using the framework detailed here, it is possible to quantify estimation uncertainty for any spatial extent desired, ranging from pixel-level predictions of AGB density to estimates of AGB stocks for the full domain. The lidar informed coregionalization models consistently outperformed their counterpart lidar-free models in terms of point-level predictive performance and total AGB precision. Additionally, the inclusion of Landsat-derived forest cover as a model covariate further improved estimation precision in regions with lower lidar sampling intensity. Our findings also demonstrate that model-based approaches that do not explicitly account for residual spatial dependence can grossly underestimate uncertainty, resulting in falsely precise estimates of AGB. On the other hand, in a geostatistical setting, residual spatial structure can be appropriately modeled within a Bayesian hierarchical framework to obtain statistically defensible measurements of uncertainty for AGB estimates.

Coregionalization+Tree Cover Model

$$\begin{bmatrix} y(\boldsymbol{s}) \\ z(\boldsymbol{s}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_{0y} + \beta_{1y}x(\boldsymbol{s}) \\ \beta_{0z} + \beta_{1z}x(\boldsymbol{s}) \end{bmatrix} + \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} w_y(\boldsymbol{s}) \\ w_z(\boldsymbol{s}) \end{bmatrix}, \begin{bmatrix} \tau_y^2 & 0 \\ 0 & \tau_z^2 \end{bmatrix} \right)$$

$$w_y(\boldsymbol{s}) \sim GP(\mathbf{0}, \rho(||\boldsymbol{s} - \boldsymbol{s}'||; \phi_y))$$
$$w_z(\boldsymbol{s}) \sim GP(\mathbf{0}, \rho(||\boldsymbol{s} - \boldsymbol{s}'||; \phi_z))$$

$$\boldsymbol{K} = \boldsymbol{A}^T\boldsymbol{A} \sim IW(\nu, \boldsymbol{\Sigma})$$

$y(\boldsymbol{s})$ = AGB (squareroot transformed)
$z(\boldsymbol{s})$ = 90th percentile height derived from lidar
$x(\boldsymbol{s})$ = Landsat-derived percent tree cover

Other Candidate Models

$$\textit{null}: y(\boldsymbol{s}) = \beta_{0y} + \epsilon_y(\boldsymbol{s}); \quad \epsilon_y(\boldsymbol{s}) \sim \mathcal{N}(0, \tau_y^2)$$

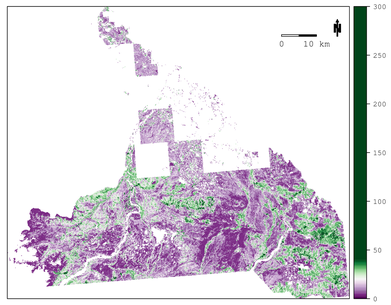
$$\textit{Tree Cover}: y(\boldsymbol{s}) = \beta_{0y} + \beta_{1y}x(\boldsymbol{s}) + \epsilon_y(\boldsymbol{s})$$

$$\textit{Spatial}: y(\boldsymbol{s}) = \beta_{0y} + \sigma_y w_y(\boldsymbol{s}) + \epsilon_y(\boldsymbol{s})$$

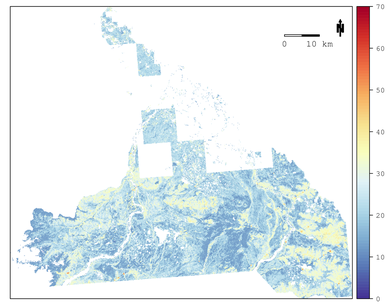
$$\textit{Spatial+Tree Cover}: y(\boldsymbol{s}) = \beta_{0y} + \beta_{1y}x(\boldsymbol{s}) + \sigma_y w_y(\boldsymbol{s}) + \epsilon_y(\boldsymbol{s})$$

$$\textit{Coregionalization}: \begin{bmatrix} y(\boldsymbol{s}) \\ z(\boldsymbol{s}) \end{bmatrix} = \begin{bmatrix} \beta_{0y} \\ \beta_{0z} \end{bmatrix} + \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} w_y(\boldsymbol{s}) \\ w_z(\boldsymbol{s}) \end{bmatrix} + \begin{bmatrix} \epsilon_y(\boldsymbol{s}) \\ \epsilon_z(\boldsymbol{s}) \end{bmatrix}$$

Tetlin National Wildlife Refuge



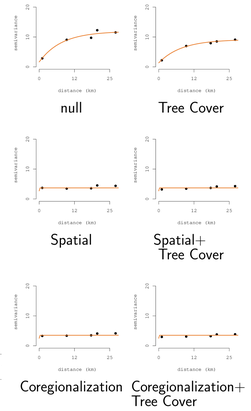
AGB predictions (Mg/ha)



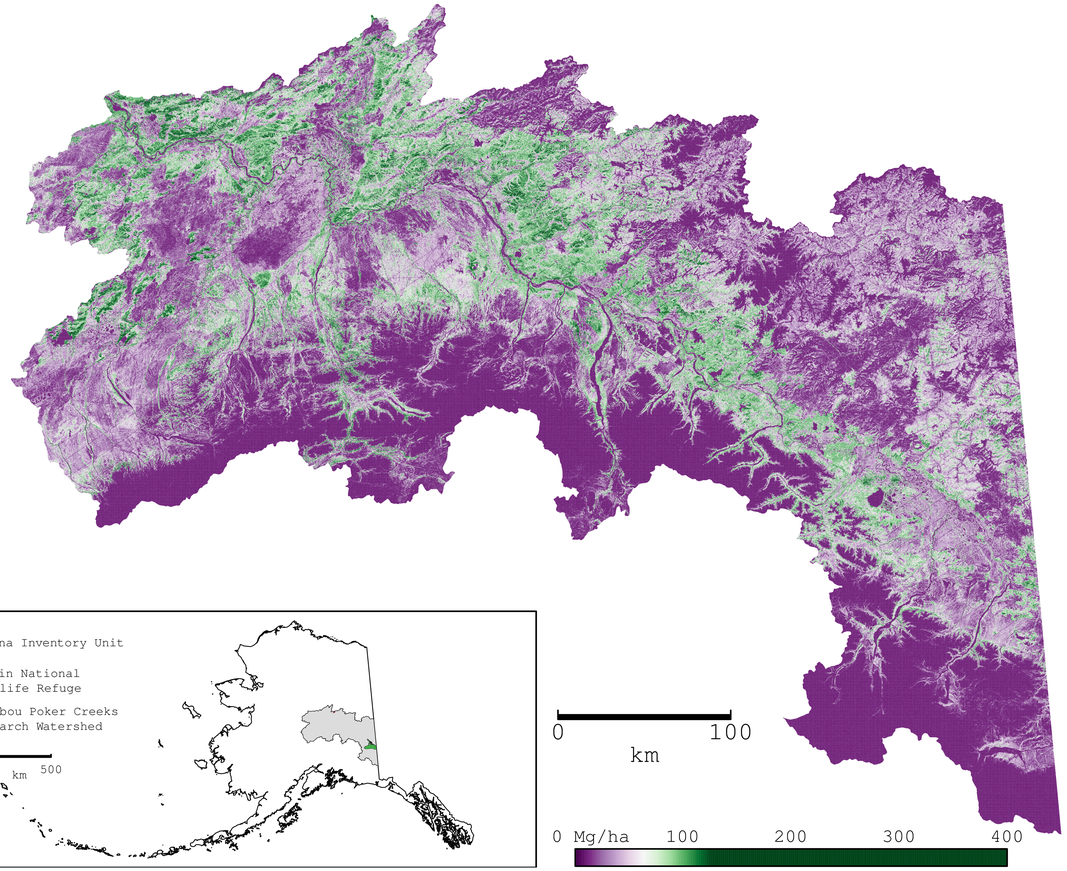
AGB standard errors (Mg/ha)

		Null	Tree Cover	Spatial	Spatial+Tree Cover	Coregionalization	Coregionalization+Tree Cover
Prediction Acc. and Total Est.	Est (Tg)	8.65 (7.09, 11.01)	8.13 (6.72, 10.13)	8.70 (6.19, 12.43)	8.00 (5.69, 10.72)	9.08 (7.66, 11.18)	8.38 (7.18, 9.68)
	SE (Tg)	1.01	0.86	1.61	1.20	0.95	0.65
	RSE	11.68%	10.59%	18.53%	14.94%	10.44%	7.73%
	RMSE (Mg/ha)	51.89	44.70	37.37	38.01	34.03	30.89

Table above shows candidate model total AGB estimates with uncertainty for Tetlin National Wildlife Refuge.
Est = Estimated total AGB, *SE* = standard error, *RSE* = relative standard error, *RMSE* = 10-fold holdout root mean squared error



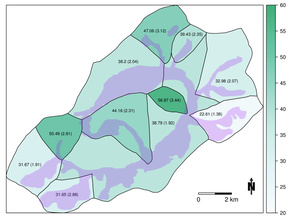
Residual variograms for the candidates (above) show that *Null* and *Tree Cover* errors are spatially autocorrelated. Spatial autocorrelation leads to underestimated *SE*s.

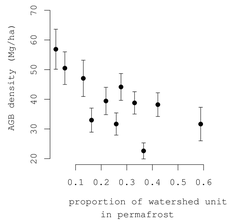


Results & Discussion

- Coregionalization frameworks can be used to couple sampled lidar and ground data with Landsat products to predict and map AGB.
- Access to the full posterior predictive distribution, a by-product of Bayesian inference, allows for generation of total forest biomass estimates for arbitrary areal units with error.
- Spatial random effects can absorb extraneous spatial autocorrelation, allowing us to generate valid model-based confidence intervals
- Nearest neighbor Gaussian process priors allow for scaling up of Bayesian spatial models to very large areas.
- Results suggest coregionalization frameworks may be useful for combining Forest Inventory and Analysis (FIA) and Global Ecosystem Dynamics Investigation (GEDI) lidar to generate national-scale forest biomass maps with associated uncertainty.

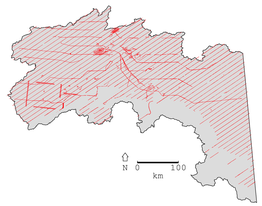
Caribou-Poker Creeks Research Watershed

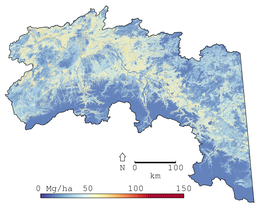




Above left figure maps watershed unit-level mean AGB (Mg/ha) with *SE*s in parentheses. Black polygon boundaries delineate watershed units. Translucent purple polygon identifies area covered in permafrost. The above right figure shows a scatter plot highlighting the relationship between the proportion of permafrost in a watershed unit and AGB density.

Tanana Valley Predictions with NNGP





Central map depicts aboveground biomass density predicted using a coregionalization model implementing Nearest neighbor Gaussian process (NNGP) spatial random effects for the Tanana Inventory Unit (TIU) at a 30 meter resolution. Above left figure shows lidar coverage over the TIU in red. Above right figure shows pixel-level posterior predictive standard errors.