REGRESSION KRIGING DOES NOT IMPROVE LANDSAT-BASED PREDICTIONS OF FOREST BIOMASS ACROSS NEW YORK STATE

Lucas Johnson

Ph.D. Student

Graduate Program in Environmental Science,

State University of New York College of Environmental Science and Forestry,

ljohns11@esf.edu

Abstract

This is my abstract and I am testing for double spacing. Because double spacing seems to be annoying.

Introduction

Forest mapping and monitoring is becoming increasingly important as federal, state, and global agencies look towards natural solutions to mitigate a warming climate and the myriad resulting challenges. Field sampling programs, like the United States Department of Agriculture's Forest Inventory and Analysis program (FIA) (Gray et al. 2012), provide unbiased estimates of forest structure over large areas, but lack the fine spatial resolution to understand and manage forests at relevant scales. Thus, high-resolution forest mapping is needed to inform decision-makers where forest resources should be managed or preserved. New York State in particular has mandated that they reach net-zero emissions across the entire economy by the year 2050, and they are actively researching carbon benefits that their forests, which dominate 60% of the statewide landscape, can offer.

Since the United States Geological Survey (USGS) opened the Landsat data archive in 2008, there has been an explosion of terrestrial monitoring approaches built with these freely available global datasets (Wulder et al. 2012; Banskota et al. 2014). Landsat offers the longest history of publicly available remote sensing data, from 1972 to present day, and has repeat observations for the same location roughly twice a month. These data are available everywhere, and Landsat missions are well supported meaning that any methods developed now can likely applied for future monitoring. Given the availability, moderate resolution (30m), temporal density, and global spatial coverage of Landsat data, many researchers have explored their utility for mapping and monitoring forest conditions over large scales and long time periods. Landsat forestry applications include forest cover or species mapping, disturbance mapping, and biophysical (e.g. basal area, biomass, and canopy height) mapping (Banskota et al. 2014).

While the benefits of Landsat imagery in terms of coverage and availability are well documented, it is also well known that they cannot predict forest structure as accurately as active remote sensing data like LiDAR (Huang et al. 2019; Hurtt et al. 2019; Chen and McRoberts 2016). Landsat-based models inherently rely on measures of spectral reflectance or 'greenness' to predict forest structure. This is challenging in that vegetation near the ground floor can and does appear quite green, and at a certain point more mature forests reach a greenness saturation point after which biomass, basal area, or height growth may still proceed without changes in spectral reflectance. This inherent limitation in Landsat's ability to model forest structure leads to less accurate spatial predictions of biomass and carbon which can have large impacts on greenhouse gas budgets that rely on

1

mapped predictions. Landsat-based model predictions have been shown to be more accurate when aggregated to larger scales (Riemann et al. 2010), however accurate predictions at smaller scales will open the door to more targeted management (e.g. private land parcel, individual forest stands) for improved forest management.

Regression kriging, or kriging of model residuals, is an approach that has been previously documented to improve spatial predictions of forest structure which rely on both optical and radar satellite imagery (A. Hudak et al. 2002; Meng, Cieszewski, and Madden 2009; Tsui et al. 2013). In this report I assessed the benefit of regression kriging to improve Landsat-based model predictions of forest aboveground biomass (AGB) across New York State (NYS). I examined first order effects effects of model residuals, fit variograms to the residuals, and spatially interpolated them using ordinary kriging. I created enhanced predictions at a set of holdout FIA plots by adding the kriged residual prediction to the original prediction. More accurate spatial predictions of AGB can help NYS target areas for increased forest management with the goal of improved carbon sequestration and avoided carbon emissions to help achieve netzero emissions in NYS by 2050.

Literature Review

Several previous studies have explored the efficacy of spatial interpolation for mapping components of forest structure like mean biomass per unit area, canopy height, and basal area. Two of such studies compared aspatial regression and machine learning methods to spatial interpolation methods in the form of ordinary kriging, co-kriging, and regression kriging (A. Hudak et al. 2002; Freeman and Moisen 2006). A. Hudak et al. (2002) showed that aspatial methods, models based on Landsat-derived predictors, better maintained the pattern of vegetation across the study area while the spatial methods were less biased when producing maps of forest canopy height in a 200 km^2 research forest in western Oregon. A. Hudak et al. (2002) categorized regression kriging as an 'integrated' approach in that the primary models relying on Landsat information are aspatial, and then the spatial relationships among the model residuals are used to improve the predictions. The integrated approach was better than the strictly spatial or aspatial approaches in that predictions had low bias and while the maps captured the spatial pattern of the vegetation. Freeman and Moisen (2006) produced similar comparisons between spatial and aspatial methods, but over a much larger study area covering the entire Rocky Mountain region (RMR). This region was broken up into 18 zones based on ecological similarity (Homer and Gallant 2001) to reduce the effects of trend across the RMR. Despite these efforts, they found that their aspatial model built with MODIS-derived predictors, was much better than the aspatial methods, and regression kriging did little to improve the aspatial model predictions. They attribute these outcomes to the fact that the aspatial model already incorporated much of the spatial pattern in the landscape through the MODIS and environmental predictor layers built into the model.

Additionally, two other studies compared various kriging approaches to each other. Tsui et al. (2013) used synthetic aperture radar (SAR) as auxiliary information to estimate AGB across a 25 km^2 area in Vancouver Island, Canada, while Meng, Cieszewski, and Madden (2009) used Landsat imagery as auxiliary information to estimate pine basal area across 20 counties in Georgia, United States. Tsui et al. (2013) showed that regression kriging was more accurate in terms of root mean squared error (RMSE) and mean absolute error (MAE) than all other kriging approaches (ordinary kriging, co-kriging, regression co-kriging). Meng, Cieszewski, and Madden (2009) showed that regression kriging was superior to ordinary kriging and co-kriging in terms of R^2.

Three out of the four studies here showed that regression kriging produced the best results among spatial and aspatial approaches, however each of these three studies operated on much smaller scales (25 km^2, 250 km^2, ~35,000 km^2) than the ~141,000 km^2 in NYS. The Freeman and Moisen (2006) study, the one study listed here where regression kriging was not the best approach, operated on a much larger area than NYS, but broke the region up into smaller units. Additionally, this study noted that the scale of variation in the mountainous RMR was smaller than the distance between reference FIA plots leading to a high nugget effect for the spatial approaches.

Study Area

The study area for this report was all of NYS (Figure 1), spanning 125,411,232 km^2, and 62 counties. The topography across the state varies from 0m above sea level to roughly 1,650m above sea level in the Adirondack region in the northern portion of the state. Roughly 60% of the state is forested, and this forested land is dominated by the maple, beech, birch forest type (53%), and is mostly privately owned (76%) (DEC, n.d.).
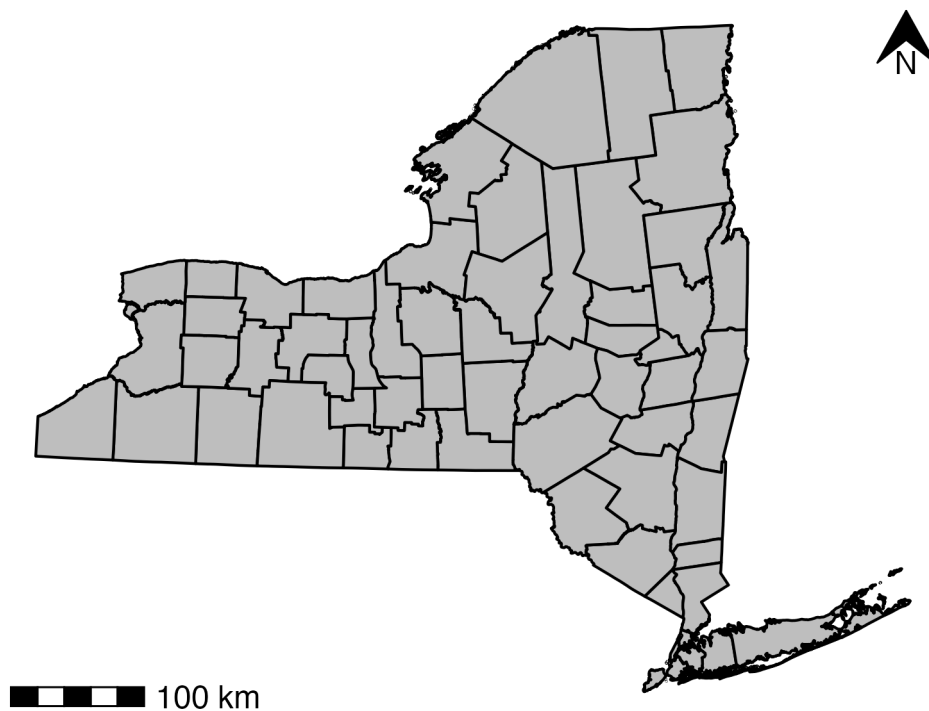

Figure 1: New York State county shoreline map.

Datasets

Table 1: Plot distance summary in kilometers; Min and Max in km; Intensity in plots per thousand km^2

Partition	Min Distance	Max Distance	Intensity
Train	2.08	631.45	3.01
Test	2.46	608.43	1.19

Estimates of AGB for all trees measuring $\geq$ 12.7 cm (5 in) diameter at breast height were produced as part of the USDA FIA program (Gray et al. 2012), with true plot centroid locations obtained under agreement with the USDA. Estimates were recorded in pounds, then summed at each plot and area-normalized to units of megagrams per hectare (Mg ha^{-1}). The plots are sampled on a hexagonal grid, with random offsets within each hexagon, such that one plot is sampled roughly every 2,400 ha (~6000 acres), however only one fifth of the plots are sampled in each year (Bechtold and Patterson 2005). For the purposes of this study, plots inventoried in 2019, the most recent inventory that is publicly available, were used to limit processing time, and to produce temporally coherent results. This field data was partitioned into a roughly 70% training dataset (426 plots), while the remaining 30% of the plots were set aside as a testing dataset (168 plots). The min, max, and average distances between plots are recorded in Table 1.

The modelling approach used to produce model residuals for this study closely followed the approach developed by A. T. Hudak et al. (2020). A set of 20 predictors were derived from Landsat analysis ready data (Dwyer et al. 2018), Landtrendr derived disturbance and temporal segmentation information (Kennedy et al. 2018), a global forest canopy height layer (Simard et al. 2011), topographic data (Mahoney 2021), climate data (Daly et al. 2008), and land cover classifications (Brown et al. 2020; Zhu and Woodcock 2014). Three machine learning (ML) models were fit to a randomly selected 70% of the training dataset. Random forest models (ranger, Wright and Ziegler (2017)), stochastic gradient boosting machines (lightgbm, Ke et al. (2021)), and support vector machines (kernlab, Karatzoglou et al. (2004)) were developed. With these 3 component models, we developed a "stacked" linear model ensemble model (hereafter LINMOD) in effort to reduce the generalization error of our component models (Wolpert 1992). LINMOD was developed by regressing the component model predictions for the 30% of the training data not used to train the component models against the observed values.

LINMOD was then used to make predictions for 30m pixels across the entire state. To produce plot-level residuals, pixel predictions were summarized at the training plot locations as well as the testing plot locations by taking the weighted average of pixels that intersected each plot. Residuals were computed by subtracting reference AGB value from the model prediction such that positive residuals represent overpredictions and negative residuals represent underpredictions.

Methods and Results

First I examined the first order effects of the data to identify any global trends or directional patterns in the data. Before attempting to visualize the data spatiall, I produced a histogram of the training data residuals (Figure 2). The histogram of the residuals indicated that the data was roughly normally distributed and centered on 0, which is perhaps unsurprising given that the the atttributes are model residuals which ideally have a mean of 0. The data is slightly skewed

left, with a few large negative residuals near 200 Mg ha^{-}1. For the sake of mapping firest order effects, I interpreted these residuals as outliers, and the color scale range was restricted to $[-max(residual), max(residual)]$ such that these large negative outliers were included, but the color scale was not shifted drastically due to their inclusion.

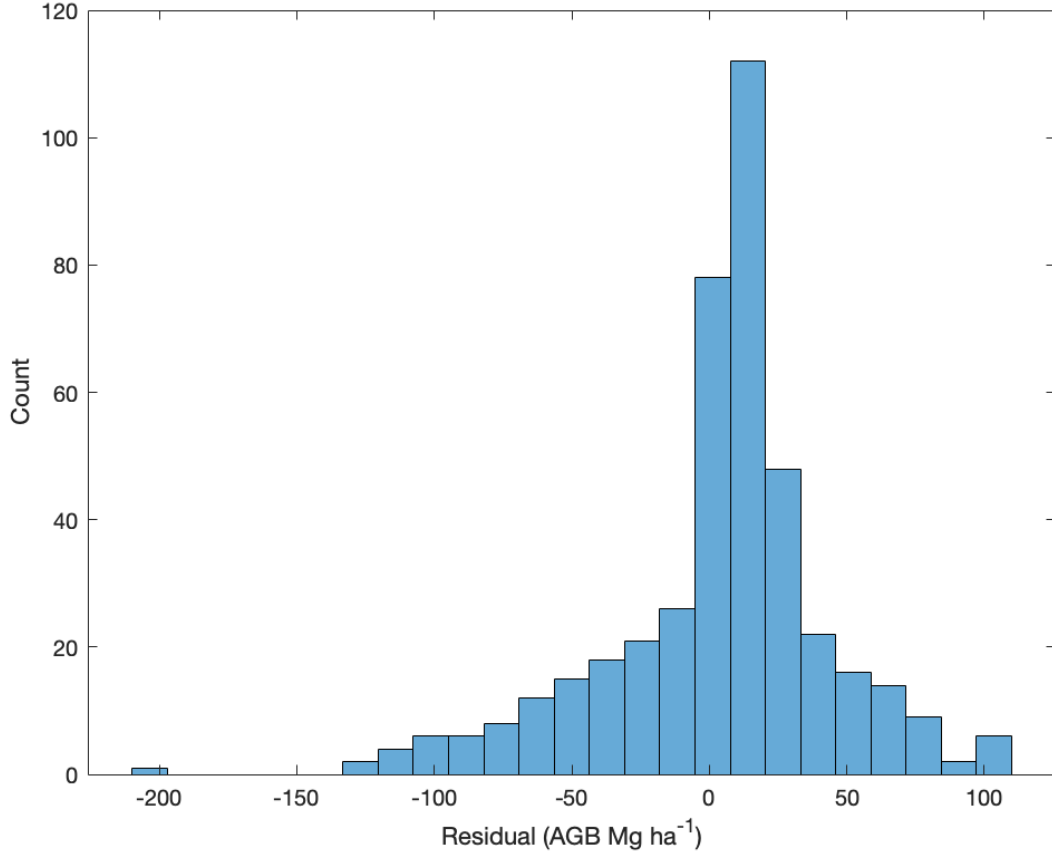Figure 2: Histogram for initial model residuals at training data locations. 25 bins used.

To visualize the first order effects spatially, I plotted a continuous point kernel function which estimates local means where attributes are weighted by their distance from the estimation location. The continuous point kernel estimates are computed as follows:

$$\hat{\mu}_\tau(s) = \sum_{i=1}^{n} w_i(s)y_i \tag{1}$$

Where $\hat{\mu}_\tau(s)$ is the average estimate for location s, n is the number of points within a distance τ of location s, w_i is the weight for point i and y_i is the attribute value for point i. w_i is computed as follows:

$$w_i(s) = \frac{k(\frac{(s-s_i)}{\tau})}{\sum_{j=1}^{n} k(\frac{(s-s_j)}{\tau})} \tag{2}$$

5

Where k is computed as follows:

$$k = \frac{3}{\pi\tau^2}(1 - \frac{h^2}{\tau^2})^2 \tag{3}$$

Where h is the distance between two locations. I chose a bandwidth of 250km and uniformly spaced kernels 5km apart to produce a smooth surface with acceptable resolution. For the color scale I chose even intervals with $1 + 3.3log(n)$ classes, centered on 0. The range was trimmed as mentioned above.

The plotted continuous point kernel surface (Figure 3) does not show any identifiable trends or directional components. Rather, the spatial arrangement of residuals seems random, with under-predictions and overpredictions not following any strong spatial pattern. This can be interpreted in two different ways. First, we might conclude that the environmental and spatial variables used as predictors in the model were able to capture any large scale trends in biomass across the state. Alternatively, if we assume that biomass is "randomly" distributed across the state, then since these residuals are partly a function of the measured biomass on the ground, it would follow that the residuals themselves are randomly distributed. We do know that there are large concentrations of forest in the Adirondack and Catskill regions, however most of the state is dominated by patchy forest cover.

Next, I explored variograms to identify the second order effects of the data with the goal of leveraging the spatial dependence of the residual values to do spatial interpolation. Specifically, variograms help identify the way the attribute deviations covary across the state. The results of the continuous point kernel analysis led to the use of a single omnidirectional variogram for the entire region. The variogram estimates for a given lag are computed as follows:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{h=s_i-s_j} (y_i - y_j)^2 \tag{4}$$

Where h represents a lag distance between two locations s_i and s_j, n is the number of points that have a distance smaller than lag h, and y_i, y_j are the attribute values at locations i and j respectively.

First I plotted the variogram estimates using the training data and 500 equally spaced lag distances (Figure 4). Additionally I plotted the variogram cloud (Figure 5) where instead of the average pairwise relationship at each lag, each individual pairwise relationship is plotted at each lag. The variogram cloud helps make sense of the odd spread (very high and very low) of variogram estimates in Figure 4 at large distances. The variogram cloud indicates that there are simply very few values at such long distances which create noisy variogram estimates.

Using the estimated variogram plot as reference I manually fit a variogram model to the estimated variogram points.

The two variogram models that I tested were a spherical model defined as:

$$\hat{\gamma}(h) = \begin{cases} s^2(\frac{2h}{2r} - \frac{h^3}{2r^3}) : h \leq r) \\ s^2 : otherwise \end{cases} \tag{5}$$

And an exponential model defined as:

$$\hat{\gamma}(h) = \begin{cases} a + ((s-a)(1 - e^{-3h/r})) : h > 0 \\ 0 : h == 0 \end{cases} \tag{6}$$
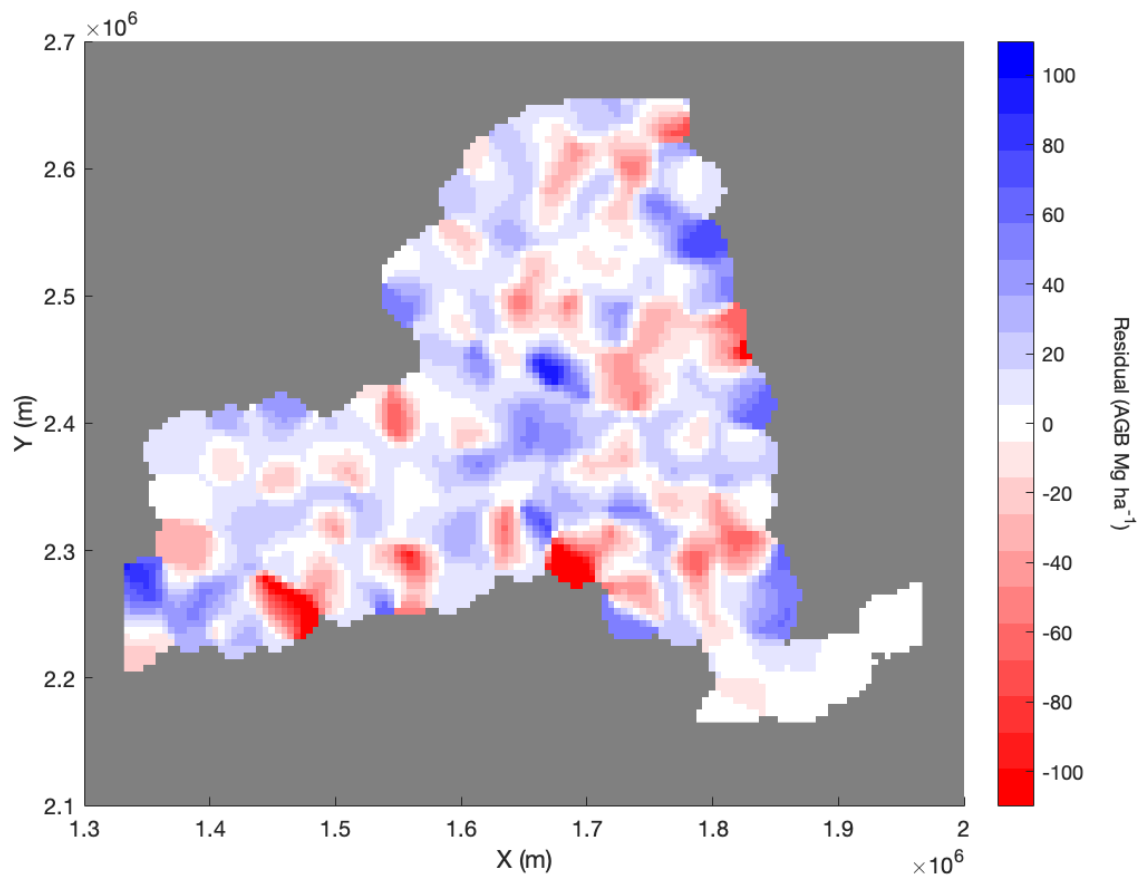
6

Figure 3: Continuous point kernel estimates for training data; Bandwidth = 25 km; Kernels uniformly spaced 5 km apart.
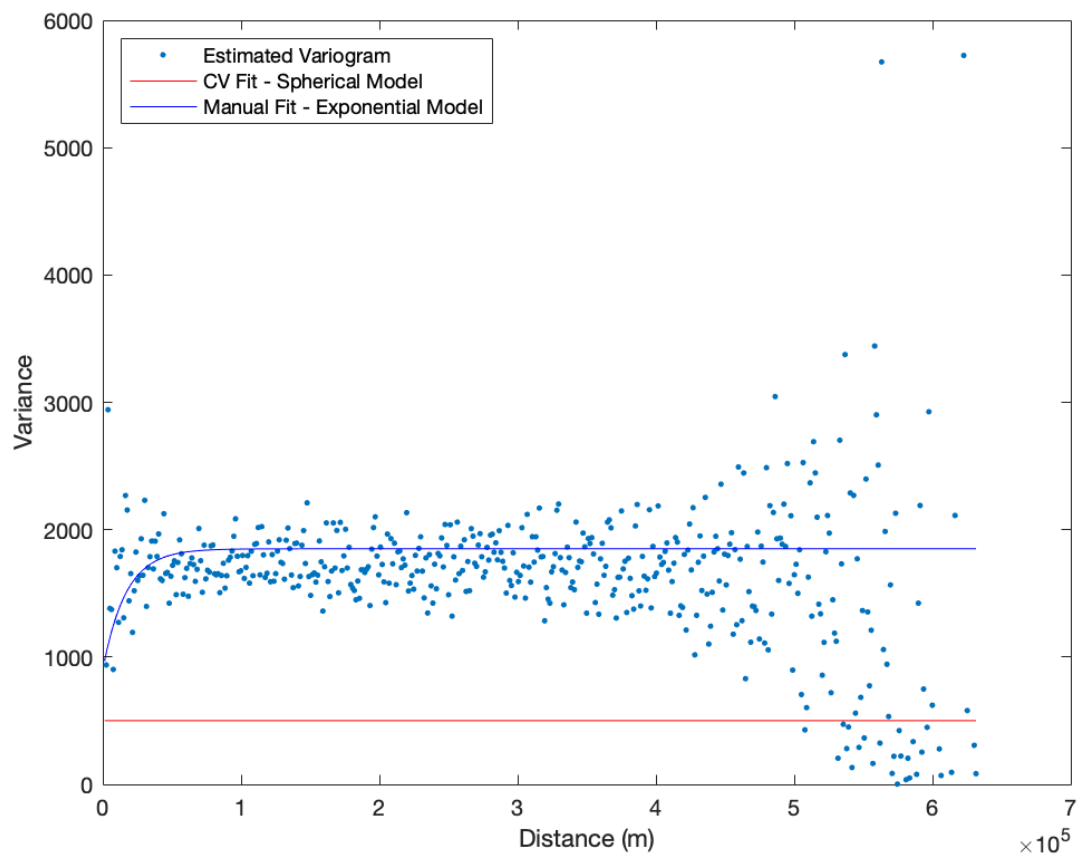
Figure 4: Estimated variogram plot from training data, with CV fit variogram and manual fit variogram overlaid.
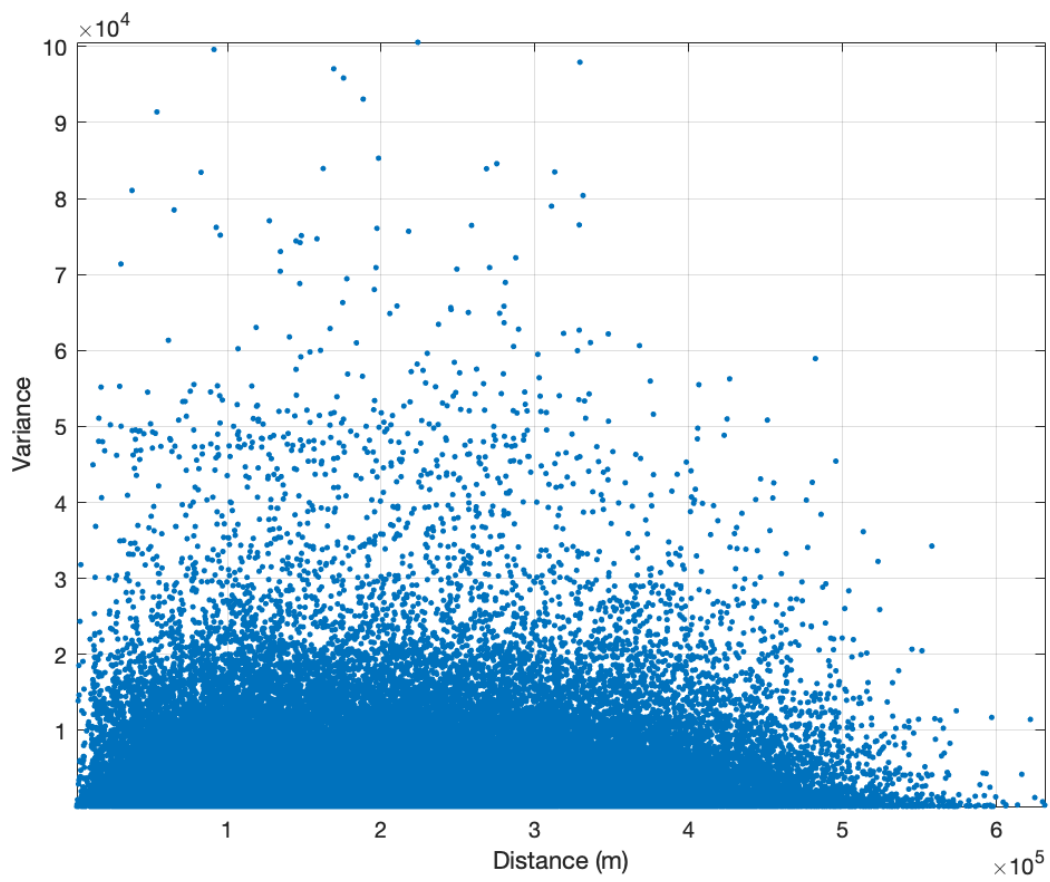
Figure 5: Variogram cloud from training data.

Table 2: Variogram parameters selected through CV fitting and manual fitting; Range in km.

Fit	Model Type	Sill	Range	Nugget
CV	Spherical	500	60	500
Manual	Exponential	2,200	60	900

Table 3: Variogram parameter search ranges (Min to Max by Step); 1,440 unique combinations tested; model type 0 corresponds to the spherical model and model type 1 corresponds to the exponential model.

Parameter	Min	Max	Step
Model Type	0	1	1
Nugget	500	2,000	100
Range	15,000	75,000	15,000
Sill	500	2,500	250

where h is the the lag distance between points, a is the nugget, r is the range, and s is the sill.

First, I manually fit a variogram model to the variogram estimates. The parameters chosen for the manually fit variogram model can be seen in Table 2, and the fit is exhibited by the blue curve in Figure 4. Additionally, I performed a standard grid search across the four variogram parameters (sill, range, nugget, model type) and assessed each using k-fold cross-validation with 5 folds. The assessment was performed using ordinary kriging (described subsequently), and the performance was assessed by computing the root mean squared error (RMSE; calculation described subsequently) of the kriged predictions of residuals compared against the actual residuals. For each unique combination of variogram parameters, the training data was split into k folds. For the kth iteration of the cross validation, the kth fold of the data was held out, and the k-1 other folds were used for training. In other words the k-1 other folds were used to compute the covariance matrix used for making kriging predictions. Predictions are made for the kth fold and those predictions are compared to the actual values. The RMSEs produced for each of the k iterations is averaged to produce a single RMSE value for each unique set of variogram parameters. I used the plotted variogram estimates (Figure 4) to determine suitable ranges for each of the variogram parameters which are described in Table 3. These combinations were sorted by their associated RMSE values, and the combination with the lowest RMSE was selected. This best combination of paramters is described in Table 2, and the fit is shown by the red line in Figure 4.

Interestingly the best CV iteration produced a "nugget variogram," where the nugget and sill are equal, indicating that the attributes do not covary in space. This fit does not account for the seemingly present rise near-range relationships as visually apparent in the variogram estimates and as identified by the manually fit variogram.

I used ordinary kriging to produce prediction surfaces across the state, and to produce point predictions at the testing data locations. Kriging estimates for a location $s0$ were computed as follows:

$$\hat{Y}(s_0) = \sum_{i=1}^{n} w_i Y(s_i) \tag{7}$$

Where $Y(s_i)$ is the value for sampled (member of training data set) location i and w_i is a corresponding weight computed as follows:

$$w(s) = C_+^{-1} c_+(s) \tag{8}$$

Where C_+^{-1} is the inverse covariance matrix computed with the provided variogram model and the sampled locations with an additional column and row of all ones appended. The last value in the diagonal (first is top left, last is bottom right) is set to 0. These modifications enforce that weights w sum to 1. c_+ is a vector containing covariance between the estimated location s and the sampled points, with a 1 appended to enforce weights w sum to 1. The variance of each kriging prediction is computed as follows:

$$\sigma_e^2 = \sigma^2 - c_+^T(s) C_+^{-1} c_+(s) \tag{9}$$

Where σ^2 is the sill parameter.

The kriging surface produced with the manually fit variogram is displayed in Figure 6 and the associated variance surface is displayed in Figure 7. The CV fit variogram was not used to produce prediction surfaces since the CV fit variogram would only produces one prediction due to the nugget being equal to the sill (Table 2. The kriging predictions of model residuals (Figure 6) are relatively conservative compared to the training data residuals (Figure 2) with predictied residuals in the range of -80 Mg to 80 Mg ha$^-$1 and much of the surface showing near zero predictions. This is likely due to the relatively small difference between the sill and the nugget. The variance surface (Figure 7) shows how quickly the uncertainty of the kriged predictions grows as we move away from plot locations, as we can assume the dots or pockets of lower variance are plot locations.

Both the CV fit variogram and the manually fit variogram were used to make predictions at the testing data locations. The residuals predicted at these locations were subtracted from the original model predictions at these locations to create new "improved" estimates. The new estiamtes, as well as the original estimates, were compared to the FIA AGB values. RMSE (Equation (10)), mean bias error (MBE; Equation (11)), and the coefficient of determination (R^2, (12)) were computed for the three sets of predictions (original, CV fit predictions, and manual predictions). These three performance metrics were computed as follows:

$$\text{RMSE} = \sqrt{(\frac{1}{n}) \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{10}$$

$$\text{MBE} = (\frac{1}{n}) \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{11}$$

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{12}$$

Where n is the number of FIA plots included in the data set, $\hat{y}_i$ is the predicted value, y_i the
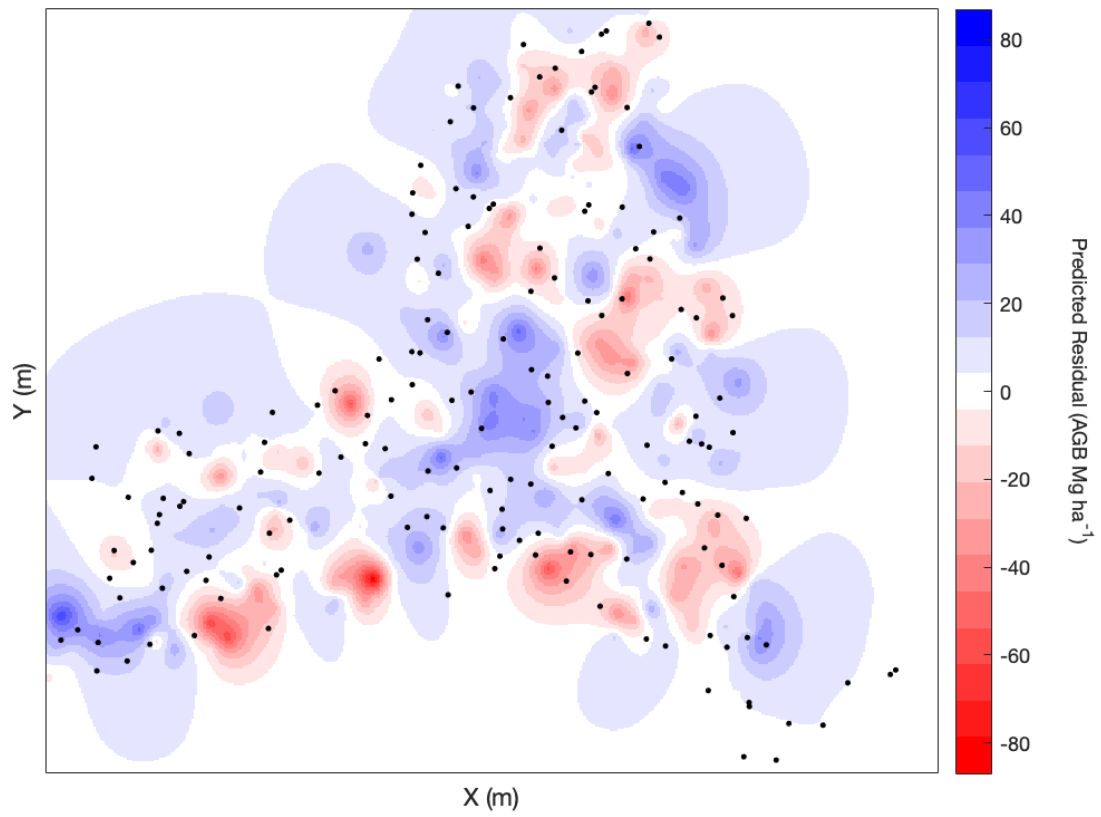
Figure 6: Kriging prediction surface for manual fit variogram with test plot locations overlaid. X and Y axis marks removed to preserve the confidentiality of the FIA plot locations.
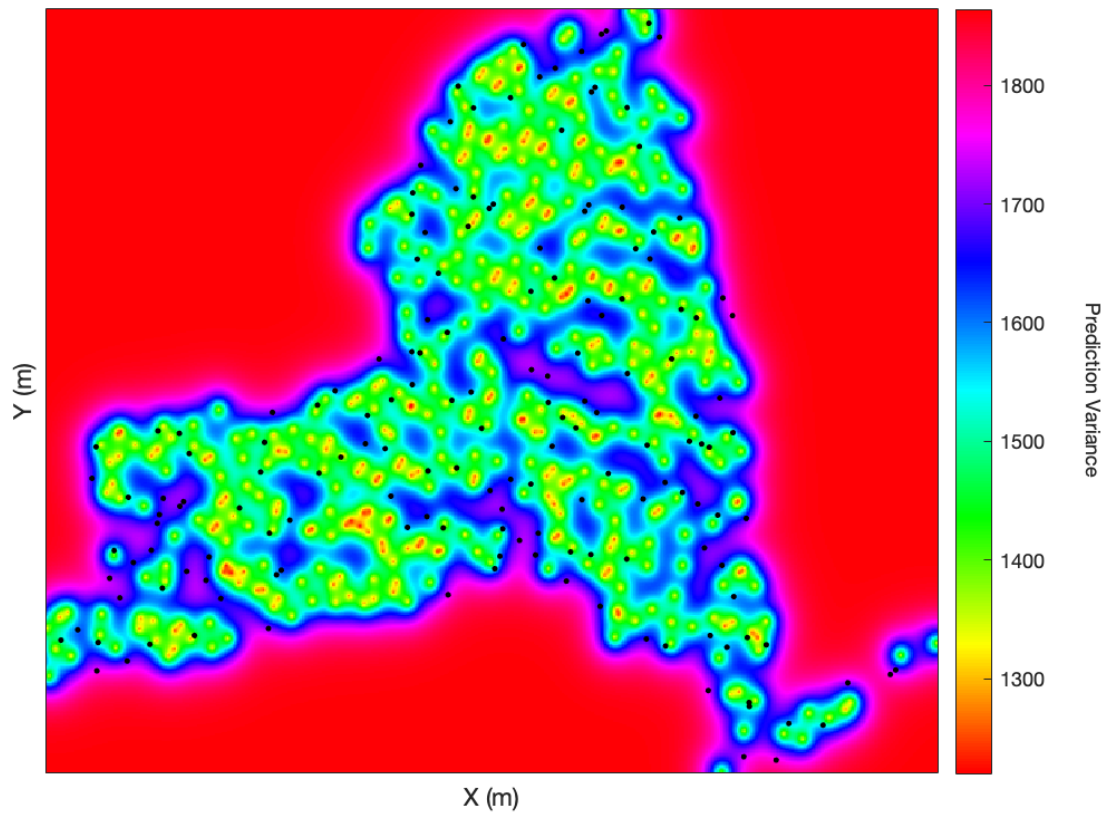
Figure 7: Kriging variance surface for manual fit variogram with test plot locations overlaid. X and Y axis marks removed to preserve the confidentiality of the FIA plot locations.

Table 4: Performance metrics for model accuracy at testing locations before and after prediction improvement via residual kriging; RMSE, MBE in AGB Mg ha^{-1}.

Predictions	RMSE	MBE	R^2
Original	42.69	-2.24	0.76
CV Fit	43.02	-5.63	0.75
Manual Fit	43.78	-5.33	0.75

measured value at the corresponding location, and $\bar{y}$ the mean value from measurements or observations.

The original model predictions were better than both the CV fit and manually fit approaches across all three metrics, however differences in RMSE and R^2 were marginal (Table 4). The CV fit appraoch produced marginally more accurate predictions than the manual fit approach in terms of RMSE, but was slightly more negatively biased (MBE). Additionally, each of the three sets of predictions were plotted against the FIA reference values and a 1 to 1 line was overlaid (Figure 8). In most cases, all three predictions are very similar which further reinforces the comparison of accuracy metrics. One noticable pattern is that the manual fit predictions are some of largest predictions were made by the manually fit model, which corresponds to the marginally less negative bias in the manual fit model as compared to the CV fit model.

Conclusions

Here's the big takeaway.

Future Work

I should also…

References

Banskota, Asim, Nilam Kayastha, Michael J. Falkowski, Michael A. Wulder, Robert E. Froese, and Joanne C. White. 2014. "Forest Monitoring Using Landsat Time Series Data: A Review." *Canadian Journal of Remote Sensing* 40 (5): 362–84. https://doi.org/10.1080/07038992.2014. 987376.

Bechtold, William A, and Paul L Patterson. 2005. *The Enhanced Forest Inventory and Analysis Program–National Sampling Design and Estimation Procedures*. Vol. 80. USDA Forest Service, Southern Research Station.

Brown, Jesslyn F., Heather J. Tollerud, Christopher P. Barber, Qiang Zhou, John L. Dwyer, James E. Vogelmann, Thomas R. Loveland, et al. 2020. "Lessons Learned Implementing an Operational Continuous United States National Land Change Monitoring Capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) Approach." *Remote Sensing of Environment* 238: 111356. https://doi.org/10.1016/j.rse.2019.111356.

Chen, Qi, and Ronald McRoberts. 2016. "Statewide Mapping and Estimation of Vegetation Aboveground Biomass Using Airborne Lidar." In *2016 IEEE International Geoscience and Remote*
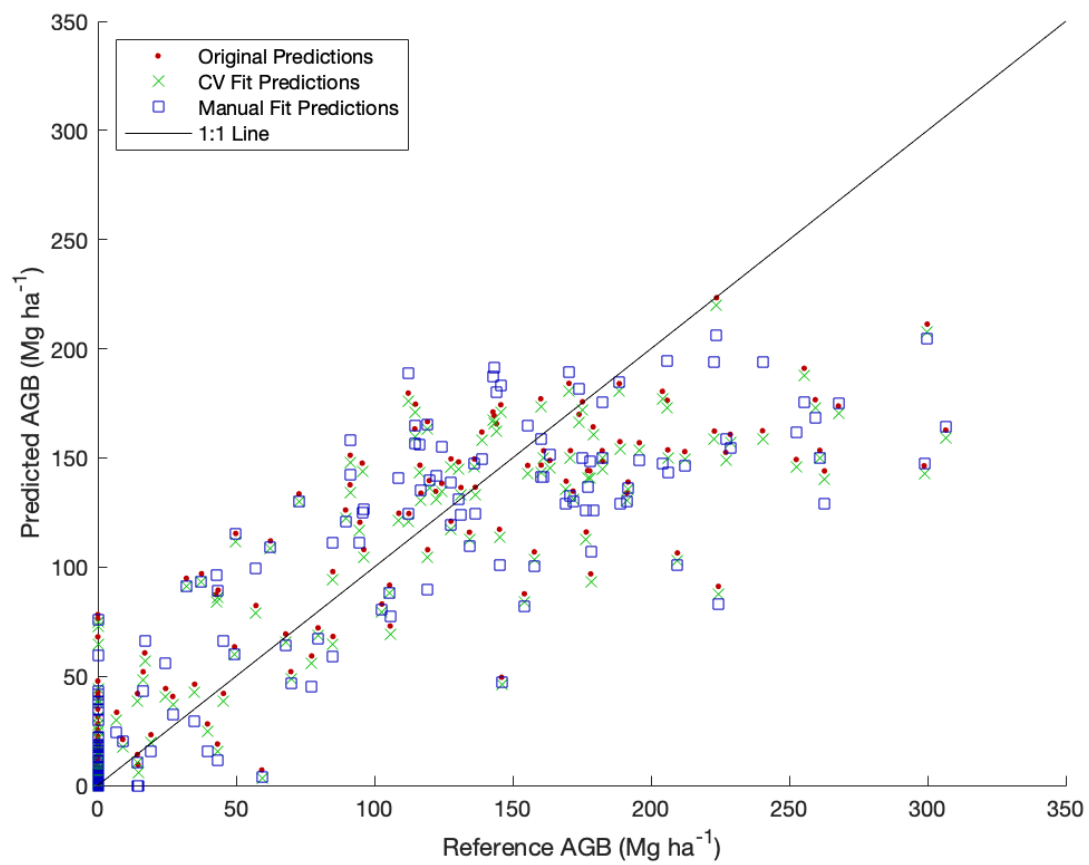
Figure 8: Predicted vs reference scatter plot of original model predictions at test data locations.

Sensing Symposium (IGARSS). IEEE. https://doi.org/10.1109/igarss.2016.7730157.

Daly, Christopher, Michael Halbleib, Joseph I. Smith, Wayne P. Gibson, Matthew K. Doggett, George H. Taylor, Jan Curtis, and Phillip P. Pasteris. 2008. "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States." *International Journal of Climatology* 28 (15): 2031–64. https://doi.org/10.1002/joc.1688.

DEC. n.d. "Forests." *Forests - NYS Dept. Of Environmental Conservation*. NYS Dept. of Environmental Conservation. https://www.dec.ny.gov/lands/309.html.

Dwyer, John L., David P. Roy, Brian Sauer, Calli B. Jenkerson, Hankui K. Zhang, and Leo Lymburner. 2018. "Analysis Ready Data: Enabling Analysis of the Landsat Archive." *Remote Sensing* 10 (9). https://doi.org/10.3390/rs10091363.

Freeman, Elizabeth A., and Gretchen G. Moisen. 2006. "Evaluating Kriging as a Tool to Improve Moderate Resolution Maps of Forest Biomass." *Environmental Monitoring and Assessment* 128 (1-3): 395–410. https://doi.org/10.1007/s10661-006-9322-6.

Gray, Andrew N, Thomas J Brandeis, John D Shaw, William H McWilliams, and Patrick Miles. 2012. "Forest Inventory and Analysis Database of the United States of America (FIA)." *Biodiversity and Ecology* 4: 225–31. https://doi.org/10.7809/b-e.00079.

Homer, Collin, and A Gallant. 2001. "Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping." *Unpublished US Geologic Survey Report. Http://Landcover. Usgs. Gov/Pdf/Homer. Pdf.(1 August 2008)*.

Huang, Wenli, Katelyn Dolan, Anu Swatantran, Kristofer Johnson, Hao Tang, Jarlath O'Neil-Dunne, Ralph Dubayah, and George Hurtt. 2019. "High-Resolution Mapping of Aboveground Biomass for Forest Carbon Monitoring System in the Tri-State Region of Maryland, Pennsylvania and Delaware, USA." *Environmental Research Letters* 14 (9): 095002. https://doi.org/10.1088/1748-9326/ab2917.

Hudak, Andrew T, Patrick A Fekety, Van R Kane, Robert E Kennedy, Steven K Filippelli, Michael J Falkowski, Wade T Tinkham, et al. 2020. "A Carbon Monitoring System for Mapping Regional, Annual Aboveground Biomass Across the Northwestern USA." *Environmental Research Letters* 15 (9): 095003. https://doi.org/10.1088/1748-9326/ab93f9.

Hudak, AT, Michael Lefsky, Warren Cohen, and Mercedes Berterretche. 2002. "Integration of Lidar and Landsat ETM+ Data for Estimating and Mapping Forest Canopy Height." *Remote Sensing of Environment* 82 (October). https://doi.org/10.1016/S0034-4257(02)00056-1.

Hurtt, G, M Zhao, R Sahajpal, A Armstrong, R Birdsey, E Campbell, K Dolan, et al. 2019. "Beyond MRV: High-Resolution Forest Carbon Modeling for Climate Mitigation Planning over Maryland, USA." *Environmental Research Letters* 14 (4): 045013. https://doi.org/10.1088/1748-9326/ab0bbe.

Karatzoglou, Alexandros, Alexandros Smola, Kurt Hornik, and Achim Zeileis. 2004. "Kernlab - an S4 Package for Kernel Methods in r." *Journal of Statistical Software, Articles* 11 (9): 1–20. https://doi.org/10.18637/jss.v011.i09.

Ke, Guolin, Damien Soukhavong, James Lamb, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2021. *Lightgbm: Light Gradient Boosting Machine*. https://CRAN.R-project.org/package=lightgbm.

Kennedy, Robert E, Janet Ohmann, Matt Gregory, Heather Roberts, Zhiqiang Yang, David M Bell, Van Kane, et al. 2018. "An Empirical, Integrated Forest Biomass Monitoring System." *Environmental Research Letters* 13 (2): 025004. https://doi.org/10.1088/1748-9326/aa9d9e.

Mahoney, Michael. 2021. *terrainr: Landscape Visualizations in r and Unity*. https://doi.org/https:

//doi.org/10.5281/zenodo.5142763.

MATLAB. 2021. *Version 9.11.0.1769968 (R2021b)*. Natick, Massachusetts: The MathWorks Inc.

Meng, Qingmin, Chris Cieszewski, and Marguerite Madden. 2009. "Large Area Forest Inventory Using Landsat ETM+: A Geostatistical Approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (1): 27–36. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2008.06.006.

Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An Effective Assessment Protocol for Continuous Geospatial Datasets of Forest Characteristics Using USFS Forest Inventory and Analysis (FIA) Data." *Remote Sensing of Environment* 114 (10): 2337–52. https://doi.org/10.1016/j.rse.2010.05.010.

Simard, Marc, Naiara Pinto, Joshua B Fisher, and Alessandro Baccini. 2011. "Mapping Forest Canopy Height Globally with Spaceborne Lidar." *Journal of Geophysical Research: Biogeosciences* 116 (G4).

Tsui, Olivier W., Nicholas C. Coops, Michael A. Wulder, and Peter L. Marshall. 2013. "Integrating Airborne LiDAR and Space-Borne Radar via Multivariate Kriging to Estimate Above-Ground Biomass." *Remote Sensing of Environment* 139: 340–52. https://doi.org/https://doi.org/10.1016/j.rse.2013.08.012.

Wolpert, David H. 1992. "Stacked Generalization." *Neural Networks* 5 (2): 241–59. https://doi.org/10.1016/S0893-6080(05)80023-1.

Wright, Marvin N., and Andreas Ziegler. 2017. "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r." *Journal of Statistical Software, Articles* 77 (1): 1–17. https://doi.org/10.18637/jss.v077.i01.

Wulder, Michael A., Jeffrey G. Masek, Warren B. Cohen, Thomas R. Loveland, and Curtis E. Woodcock. 2012. "Opening the Archive: How Free Data Has Enabled the Science and Monitoring Promise of Landsat." *Remote Sensing of Environment* 122: 2–10. https://doi.org/https://doi.org/10.1016/j.rse.2012.01.010.

Zhu, Zhe, and Curtis E. Woodcock. 2014. "Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data." *Remote Sensing of Environment* 144: 152–71. https://doi.org/10.1016/j.rse.2014.01.011.