

ESTIMATING FIA PLOT CHARACTERISTICS USING NAIP IMAGERY, FUNCTION MODELING, AND THE RMRS RASTER UTILITY CODING LIBRARY

John S. Hogland & Nathaniel M. Anderson¹

Abstract—Raster modeling is an integral component of spatial analysis. However, conventional raster modeling techniques can require a substantial amount of processing time and storage space, often limiting the types of analyses that can be performed. To address this issue, we have developed Function Modeling. Function Modeling is a new modeling framework that streamlines the raster modeling process by utilizing delayed reading methods. Using this approach, we have successfully characterized the impacts of fuel treatments on soil erosion, estimated basal area, trees, and tons of above ground biomass per acre, identified locations in need of forest management, calculated forest residuals given multiple management prescriptions, and developed forest residual delivery cost models all in a fraction of the time and storage space it would take to perform similar analysis using conventional methods. To facilitate the use of Function Modeling, we have built an object oriented .NET library called RMRS Raster Utility. RMRS Raster Utility is free, readily available, and has an intuitive user interface that directly plugs into Environmental Science Research Institute (ESRI)'s software. In this paper we will discuss the basic concepts behind Function Modeling and present some of our recent findings related to using this technique to estimate FIA plot characteristics from NAIP imagery.

INTRODUCTION

Raster modeling is vital to performing remote sensing and spatial analysis. Combined with classical statistic and machine learning algorithms, it has been used to address a wide range of questions in a broad array of disciplines (e.g. Patenaude et al. 2005; Reynolds-Hogland et al. 2006). However, the traditional workflow used to integrate statistical and machine learning algorithms and process raster models within a geographic information system (GIS) can limit the types of analyses performed and outputs created. In a recent study, Hogland and Anderson (2014) demonstrated that Function Modeling (Hogland et al. 2013) can streamline manipulating raster surfaces, building predictive models, and creating predictive

raster outputs while substantially reducing processing time and digital storage requirements. Using the concept of Function Modeling (FM), they built a publicly available coding library that facilitates a wide range of tabular, spatial, statistical, and machine learning analyses (RMRS 2012a). To simplify the use of these procedures and concepts, they built an intuitive user interface packaged as an Environmental Systems Research Institute (ESRI) toolbar add-in called RMRS Raster Utility that works directly with ESRI's ArcMap versions 10.0 through 10.3 (RMRS 2012b).

One of the key benefits of FM is a reduction in the number of input/output operations performed within complex spatial modeling tasks. From a programming perspective, this can be achieved using strategies such as delayed reading and overloading the functionality of existing classes that support reading data from disk (Van Roy and Haridi 2004). In practice, these coding

¹ Biological Scientist (JSH) and Research Forester Nathaniel M. Anderson (NMA), USDA Forest Service, Rocky Mountain Research Station, 800 E Beckwith Missoula, MT 59801. JSH is corresponding author: to contact, call (406) 239-2138 or e-mail at jshogland@fs.fed.us.

techniques take advantage of the fact that we rarely need to view, process, or use all spatial data at once and that the speed of computer processing units are significantly faster than the transfer speeds of data from disk to computer memory. Taking advantage of these types of efficiencies, we have successfully related Forest Inventory Analysis (FIA) program plot information to fine grained National Agriculture Imagery Program (NAIP) imagery and have built predictive surfaces for the extent of a national forest in Colorado (Hogland et al., 2014) and are now in the process of expanding these types of analyses across multiple forests that include both public and private lands in Montana.

METHODS

The study area is a 100 mile radius around Helena, MT. Within this area, forest community characteristics are estimated by creating statistical and machine learning relationships between FIA plot summaries (the response) and fine grained NAIP imagery, National Elevation Dataset (NED) digital elevation models (DEM), and textural and topographical derivative of those dataset (explanatory variables). Using the spatial locations of the FIA plot, explanatory variable values are extracted and related to the response to create a predictive model. The resulting predictive model is then applied to the explanatory variable values to produce a surface of forest characteristics across the landscape.

We used this approach to estimate basal area per acre (BAA), trees per acre (TPA), and tons of above ground

biomass per acre (AGB) based on commonly used allometric equations (Jenkins et. al. 2003) for all tree species. The sample included 1652 FIA field plots, of which 1097 plots were randomly chosen to calibrate each model and the remaining 555 plots were used to independently evaluate each model. Imagery for the area surrounding Helena was collected, preprocessed, and mosaicked by NAIP at a spatial resolution of 10.76 feet and a geometric accuracy of + 3 pixels (McGlone et al. 2006). FIA plot summaries were related to visually interpreted patterns of 2013 NAIP color infrared imagery (CIR) using a two stage classification and estimation approach (Hogland et al., 2014).

The first stage of this approach results in a probabilistic classification of visually identifiable patterns (Table 1) such as live and dead tree crowns using RMRS Raster Utility's Soft-max neural network algorithm, derivatives of NAIP spectral reflectance, and second order Gray Level Co-occurrence Matrix (GLCM; Haralick 1973) texture values of those spectral derivatives. Training samples for the probabilistic classification consisted of 1000 random locations across the study area. Outputs from the first stage classification were summarized using the focal mean function within the RMRS Raster Utility toolbar for the approximate extent of a 1/10 acre field plot (window size of 21 x 21 NAIP pixels). Raster cells with mean percent tree canopy cover greater than 10 percent were classified as Forested. Forested areas were then used to partition FIA training plots and mask BAA, TPA, and AGB outputs.

Table 1—Visually identifiable patterns and associated labels used to estimate Forested areas.

Name	Description	Label
Green Grass	Vigorously growing grass that has a strong green spectral signature.	GG
Brown Grass	Sensed grass that does not have a strong green spectral signature.	BG
Green Crown	Live crowns for needle leaf tree species	GC
Brown Crown	Dead crowns for needle leaf tree species	BC
Tree Shadow	Shadows created by trees	TS
Green Deciduous	Live crowns for broad leaf tree species	GD
Bare Soil	Bare ground with no vegetation.	BS
Rock Pavement Building	Rock, Pavement, and Buildings	RPB
Snow Bright	Snow highly reflective materials	SB
Water	Water	W

The second stage of our estimation procedure relates FIA plot summaries to spectral, textural, elevation, and topographical derivatives using the spatial coordinates of each FIA plot, a focal window of 78 x 70 cells, and the random forest regression tree tool within the RMRS Raster Utility toolbar. Spectral, textural, elevation, and topographic variables include first order mean and standard deviation textural metrics of NAIP CIR, natural difference vegetation index (NDVI) derived from NAIP imagery, and elevation and topographic derivatives, along with second order horizontal contrast GLCM metrics values of NAIP CIR spectral bands.

RESULTS

The first stage of our approach had a maximum likelihood classification accuracy of 65 percent and accounted for 84.56 percent of the information in the data (). Using the extent of a FIA plot, the focal mean function within RMRS Raster Utility toolbar, and the

probability surfaces of live and dead crowns, we created a Forested mask and attributed each FIA plot as either forested or non-forested (Fig. 1). In all, 660 and 992 FIA plots were located within and outside of Forested areas, respectively, and were used to train or validate the second stage of our classification and estimation process.

The second stage of our procedure generated a series of regression tree models that had a mean out of bag (OOB) root mean square error (RMSE) of 47.26 for BAA, TPA, and AGB (Table 2). Comparing observed plot summaries to estimated BAA, TPA, and AGB for our independent validation sample, we found that RMSE calculated from the independent sample was similar to OOB RMSE of our models. Using our regression tree models, the RMRS Raster Utility tool “Build Raster from Models”, and the explanatory variable surfaces, we created a series of raster surfaces that predicted FIA plot BAA, TPA, and AGB summaries for every 32.81 ft across the study area (Fig. 2).

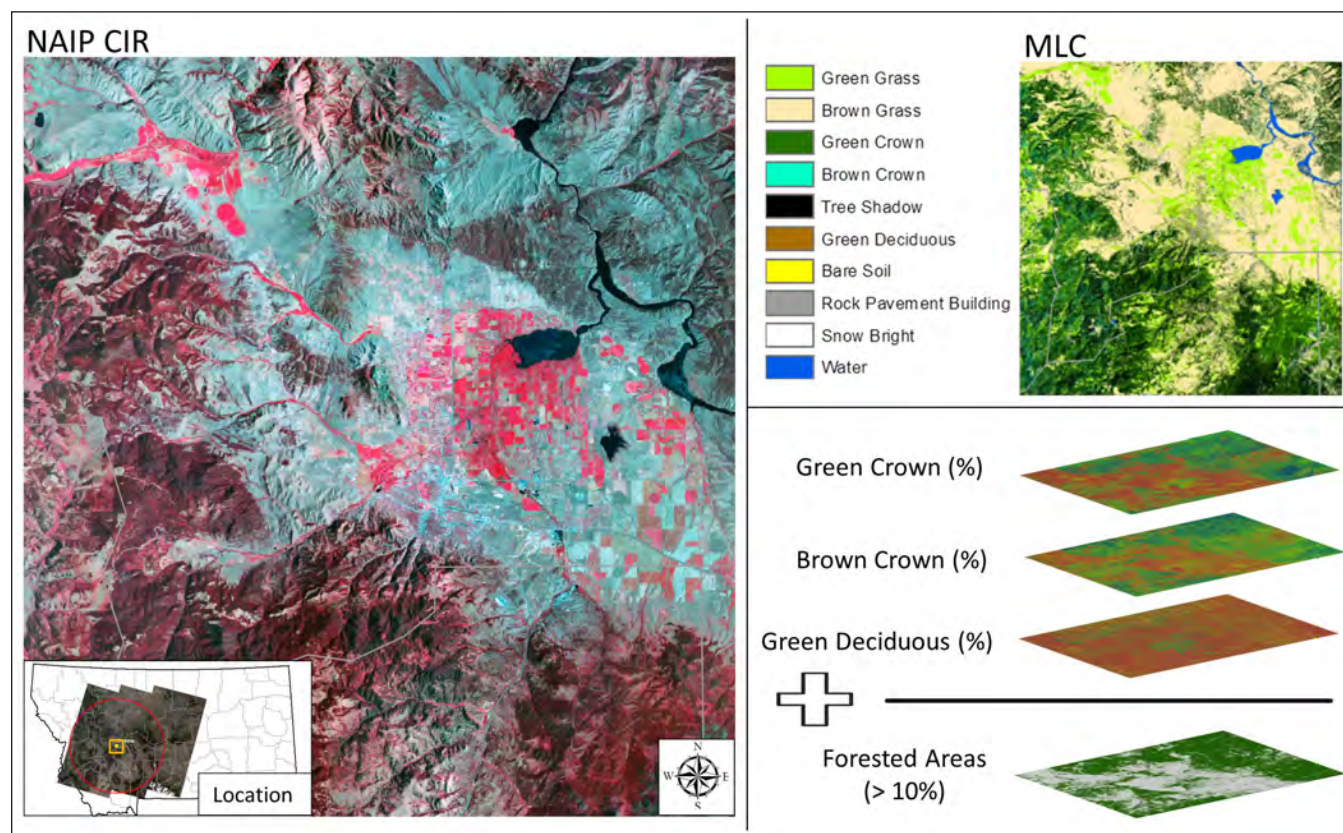


Figure 1—A close up illustration of the outputs from the first stage of our classification and estimation methodology. The extent of the close up covers 2,200 square miles around Helena, Montana and is denoted by the yellow square in the lower left Location map. Most Likely Class (MLC), Green Crown, Brown Crown, and Green Deciduous probabilistic outputs, and Forested Areas raster surfaces were derived from visually identified point locations, a Soft-max Neural Network, and NAIP CIR spectral and textural metrics.

Table 2—Fit statistics for BAA, TPA, and AGB models.

Characteristic	OOB RMSE	OOB R	Validated RMSE	Validated R
BAA	37.23	0.66	36.82	0.71
TPA	82.13	0.68	94.58	0.68
AGB	22.43	0.60	21.07	0.65

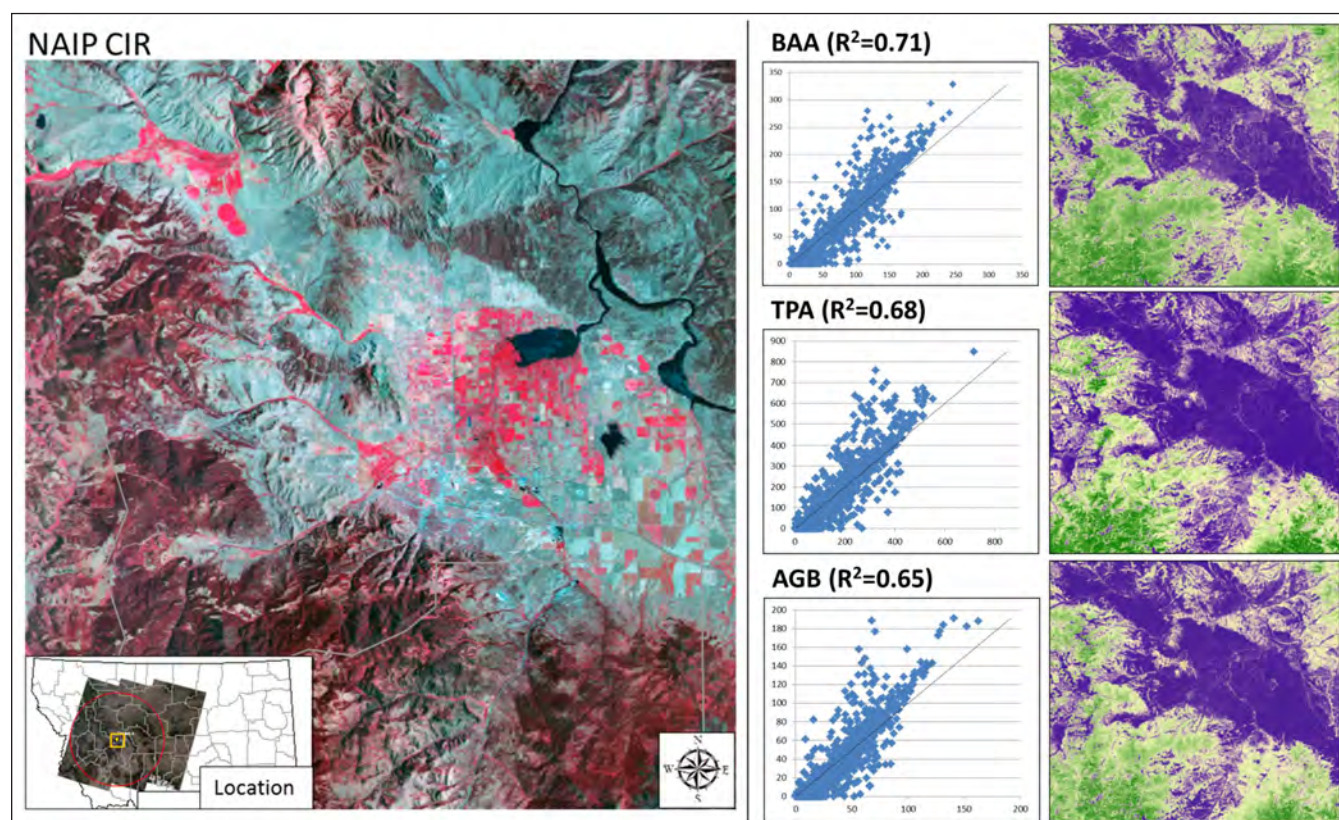


Figure 2—A close up illustration of the outputs from the second stage of our classification and estimation methodology. The extent of the close up covers 2,200 square miles around Helena, Montana and is denoted by the yellow square in the lower left Location map. Basal area per acres (BAA), trees per acre (TPA), and tons of above ground biomass per acre (AGB) were derived from FIA plot data, NAIP imagery, NED DEM, and multiple random forest algorithms. The right portion of the figure depicts the independently assessed relationships between response and explanatory variables and spatially illustrates increases in BAA, TPA, and AGB as color transitions from purple to green.

DISCUSSION

Our two stage classification and estimation approach successfully depicts mean forest characteristics for every 32.81 ft within 100 mile radius of Helena, MT. The relationships in these models are derived in explanatory variable space, as opposed to being spatial aggregations of plot characteristics, and are based on the correlative strength between the explanatory variables and FIA plot characteristics. Given that explanatory variables have been inventoried across the study

area, we can use these relationships to depict forest characteristics at the resolution of each raster cell (Fig. 2). These types of outputs are extremely useful and can be aggregated across multiple geographic boundaries at many different spatial scales without losing accuracy or precision. Using our newly developed RMRS Raster Utility toolbar and coding libraries, these depictions can be easily manipulated within a GIS to address a wide range of management related questions at regional, landscape, and project extents.

ACKNOWLEDGMENTS

This project was supported by the Rocky Mountain Research Station and Agriculture and Food Research Initiative, Biomass Research and Development Initiative, Competitive Grant no. 2010-05325 from the USDA National Institute of Food and Agriculture.

LITERATURE CITED

- Haralick, R.; Shanmugam, K.; Dinstein I. 1973. Texture features for image classification. *IEEE Trans.Syst. Man. Cybern.* 3: 610-621.
- Hogland J.; Anderson N.; Jones G. 2013. Function modeling: improved raster analysis through delayed reading and function raster datasets. *Proceedings, Council on Forest Engineering (COFE) Annual Meeting, 2013 July 8; Missoula MT, USA.* Available: http://www.fs.fed.us/rm/pubs_other/rmrs_2013_hogland_j001.pdf. (Accessed June 6, 2014.)
- Hogland J.; and Anderson N. 2014. Improved analyses using Function Datasets and statistical modeling. *Proceedings, ESRI international users conference Annual Meeting, 2014 July 10-17; San Diego, CA.* Available: http://proceedings.esri.com/library/userconf/proc14/papers/166_182.pdf (Accessed June 25, 2015.).
- Hogland J.; Anderson N.; Wells L.; Chung W. 2014. Estimating forest characteristics using NAIP imagery and ArcObjects. *Proceedings, ESRI international users conference Annual Meeting, 2014 July 10-17; San Diego, CA.* Available: http://proceedings.esri.com/library/userconf/proc14/papers/155_181.pdf (Accessed June 25, 2015.).
- McGlone, C.; Barclay, T.; Freeborn, E.; Greve, C.; Habib, A.; Keating, T.; Lenczowski, R.; Logan, B.; Schenk, T.; Stojic, M.; Voss, A. 2006. ASPRS Report to the U.S. Geological Survey on Digital Orthoimagery. Available: http://nationalmap.gov/report/ASPRS_Report_on_Digital_Orthoimagery.pdf (Accessed June 25, 2015.).
- Patenaude, G.; Milne R.; Dawson, T. 2005. Synthesis of remote sensing approaches for forest carbon estimation Reporting to the Kyoto Protocol. *Environmental Science and Policy* 8(2): 161–178.
- Reynolds-Hogland, M.; Mithchell, M.; Powell, R. 2006. Spatio-temporal availability of soft mast in clearcuts in the Southern Appalachians. *Forest Ecology and Management* 237(1–3): 103-114.
- RMRS (2012a). RMRS Raster Utility SVN Repository. Available: <https://collab.firelab.org/svn/rmrsraster/> (Accessed June 6, 2014.).
- RMRS (2012b). RMRS Raster Utility. Available: <http://www.fs.fed.us/rm/raster-utility/> (Accessed June 6, 2014.).
- Van Roy, P.; Haridi, S. 2004. Concepts, techniques, and models of computer programming, Massachusetts Institute of Technology Press Cambridge, MA, 900p.