

BUILDING A NATIONWIDE 30-METER FOREST PARAMETER DATASET FOR FOREST HEALTH RISK ASSESSMENTS

James R. ELLENWOOD and Frank J. KRIST

FHTET; USDA Forest Service, 2150 Centre Ave., Bldg A, Ste. 331, Fort Collins, CO 80526,
jellenwood@fs.fed.us

This paper was presented at the ForestSat'07 Conference, "Forests, remote sensing and GIS: methods and operational tools", held in Montpellier France, November 5th – 7th 2007.

ABSTRACT

Although useful at a national scale, the 1-km resolution 2006 National Insect and Disease Risk Map (NIDRM) has limited utility at the local and regional levels. The USFS has begun work on a 'next-generation' national risk map at a 30-meter resolution to facilitate local and regional planning efforts. In support of this project, forest parameter datasets of basal area and trees per acre are being developed for more than 175 tree species using ground inventory data and 52 national data layers ranging from three-season Landsat 7 imagery to local soil information. To develop models of forest parameters by tree species, USFS Forest Inventory and Analysis (FIA) data are linked to each national data layer and analyzed within Cubist data mining software (Rulequest 2007). Model outputs are converted to a 30-meter spatial dataset using ERDAS Imagine software. Forest parameter models representing less dominant tree species were not well-correlated. In order to improve model results, a new technique called extirpation was developed in which models are assembled for an individual species from the total parameter values of all remaining species. Early results have been promising based upon field validation.

Keywords: Cubist, Landsat, Forest Parameters, Extirpation

1 INTRODUCTION

The National Insect and Disease Risk Map (NIDRM) is a tool to provide policy makers, USFS officials, and federal and state land managers with a periodic strategic assessment of risk of tree mortality due to major forest insects and diseases (Krist et. al. 2007b). Limitations of the most recent risk assessment were evaluated and opportunities for improvement were identified for the 2010 assessment. One of the key findings of this evaluation was to create base layers depicting forest host parameters at a higher spatial resolution. The purpose of this paper is to document the procedures used to develop these improved forest host parameter layers.

1.1 BACKGROUND

For the 2006 risk assessment, field specialists formulated 188 multi-criterion pest models to assess the risk of mortality on individual tree species over a 15-year period (Krist et. al. 2007a). Each of these models required one or more tree species-specific forest parameters including basal area (over 1" dbh), mean diameter, and tree density (over 1" dbh). These parameters were modeled on a 1-km cell basis, allowing for broad-scale analysis; however, forest pests inherently operate at a local-scale

making local and regional assessments difficult. Therefore, the goal of the 2010 NIDRM is to utilize forest parameter base-layers at 30-meter ground resolution to produce a product that can be used at the local, regional, and national levels.

1.2 AVAILABLE NATIONAL DATASETS

The USGS has recently completed the 2001 National Land Cover Dataset (Homer, 2007) encompassing the continental United States and Alaska. This 30-meter dataset consists of 3 primary products: a landcover layer representing 18 different classes, a forest-canopy cover layer, and an impervious surface layer. In addition to these products, the NLCD project has made available a set of terrain data, and radiometrically corrected and mosaicked three-season Landsat 7 scenes for each of the 78 USGS mapping zones (Figure 1).

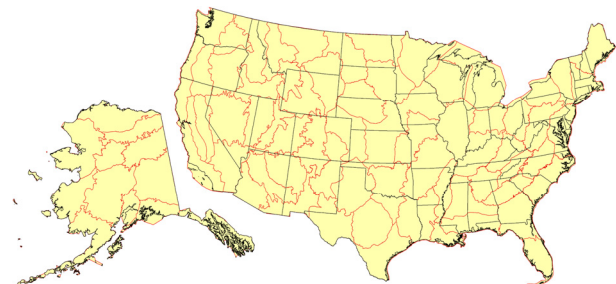


Figure 1. USGS Mapping Zones

The USDA-NRCS is currently coordinating the National Cooperative Soil Survey (NCSS), which is a nationwide partnership of federal, regional, state, and local agencies and institutions. Two products are being developed as part of the NCSS: a broad-scale product known as the US General Soil Map or STATSGO and a local-scale product known as the SSURGO (NRCS, 2007).

Climate data were derived from the University of Montana Numerical Terradynamic Simulation Group (NTSG) which produces the Daymet climate datasets at 1-km resolution (Thornton et. Al. 1997, Daymet 2007).

The USDA Forest Service FIA conducts annual inventories of forested lands for all ownerships in each state across the United States. This dataset provides a nationally consistent measurement of individual trees for determining various forest parameters (Becthold and Patterson, 2005). The FIA data is the only ground survey data used to produce the model parameters for this project.

2 PROCEDURES

2.1 NLCD DERIVATIVES

Since the NLCD project undertook the daunting task of creating radiometrically-corrected, Landsat-7 scenes for each of the 78 USGS mapping zones, the opportunity existed to extract more information from this intermediate product. Each mapping zone contains a mosaicked dataset consisting of reflectance corrected Landsat imagery (6 bands) for 3-seasons plus three tasseled-cap transformed bands comprising a total of 27 image bands. For each season, a date-band thematic image was created. In addition to the thematic date-band, a Julian day image was created as a continuous band so the modeling process can better distinguish the temporal differences between the plot data and the imagery. Including the core NLCD products of land cover, impervious surface, and canopy cover, a total of 36 predictor layers were assembled from the NLCD project.

2.2 SOIL DATA PREPARATION

The SSURGO data is a vector-based dataset developed from 1:12,000 or 1:24,000 orthophotos for each county or soil survey area within a given state. Each individual soil map unit within a survey area contains one or more soil components. To reduce the data complexity, the majority soil component was selected for analysis. Since a multitude of variables are associated with each soil

component, a process was needed to reduce the variables into a meaningful measure for incorporation in the forest parameter modeling effort. A soil drainage index (DI) was developed to allow for the incorporation of a measure of the long-term wetness of a soil, which is a key variable related to tree growth (Schaetzl et. al. 2007).

Originally named the “natural soil wetness index” (Schaetzl 1986), DI is an ordinal measure of the relative amount of water that a soil contains and makes available to plants under normal climatic conditions. The main factors affecting DI are the depth to the water table, soil moisture regime and volume available for rooting, and (lastly) texture. Therefore, the DI is calculated from the soil’s taxonomic subgroup classification in the US system of soil taxonomy, along with its textural family and slope class. DI values range from 0 to 99. The higher the DI the more water a soil can supply to plants. Sites with a DI of 99 are almost continually waterlogged while a soil with a DI value of zero is thin and dry enough to almost be bare rock or raw sand. A DI value was assigned to each majority soil component within the SSURGO and STATSGO databases.

In addition, soil component dominance and diversity were determined for each soil map unit and incorporated in the predictor layers. In areas where SSURGO data is not available, the majority soil components from the STATSGO data were utilized. From these two data sources, a single 30-meter soil layer was created for each USGS map zone.

2.3 CLIMATE DATA PREPARATION

From the Daymet data, eight layers were utilized: frost free days, growing degree days, average annual precipitation, short wave radiation, average annual temperature, average maximum temperature, average minimum temperature, and water vapor pressure. For each Daymet layer, the first principle component of the 12-monthly averages (1980-1997) was rescaled using the ArcGRID SLICE routine with the equal area option to reduce the dataset and to maximize the variation for the modeling process.

An additional layer was developed to represent seasonal moisture by combining the USGS NDVI seasonal metric data, 1989 to 2001 (Swets, 1999), with a summation of the corresponding average monthly precipitation data.

The native resolution of the Daymet layers is 1-kilometer. A 30-meter modeled surface generated from these datasets would result in a product that has large blocks of pixels which effectively reduces the ground resolution of the data. In order to make

the climate data more comparable to other predictor layers, the native raster layers were disaggregated into point datasets and then resampled to a finer resolution utilizing the Arc/Info TOPOGRID function.

2.4 TERRAIN DATA PREPARATION

The terrain data is derived from the 30-meter National Elevation Dataset (NED) utilized in the NLCD production (USGS 2007). In addition to the digital elevation model (DEM) and position index obtained from the NLCD project, slope was recomputed to percent and aspect was recomputed into two phases based upon the sin and cosine, and a slope percent weight (Stage 1976). Slope curvature was also computed and included in the analysis. A total of six continuous terrain variables were generated from this source.

2.5 FIA DATA PREPARATION

Ground data is derived from continuous forest inventory data collected by the USFS Forest Inventory and Analysis (FIA) unit. The standard national FIA plot design consists of 3-subplots centered on a central sub-plot (Bechtold and Patterson 2005). Only the central subplot has geographic coordinates recorded. The locations of the remaining 3-subplots were calculated from the plot design. The subplot installation design orients the plots utilizing magnetic north. Subplot locations were calculated using the subplot design and the magnetic declination at the time the plots were installed. Magnetic declination was determined from the 1995 magnetic declination (Tarr 2000a) and adjusted by the secular variation (Tarr 2000b) based upon the installation date.

Since FIA is a continuous inventory, plot data was selected by individual state inventories that were within 2-3 years of the imagery date. From these cycles, all live trees greater than or equal to 1 inch in diameter were isolated and forest parameters for basal area and trees per area were computed for each individual sub-plot. Plots which were determined to be sampled but with no sample trees were assigned a value of 0 basal area and trees per acre. Plots that were not sampled, designated status code 3 (USFS 2007), as well as plots without geographic coordinates or coordinates that were deemed inaccurate were removed from the sample.

2.6 MODEL BUILDING

A total of 52-predictor layers were sampled utilizing the FIA subplot locations. For continuous variables bilinear sampling was utilized and for thematic variables nearest neighbor sampling was employed.

Bilinear sampling was selected to account for misalignment between pixel and subplot locations. Once a table containing FIA tree records and predictor layer values was assembled, a separate model was generated for basal area and trees per acre for each USGS mapping zone.

Cubist software (Rulequest 2007) was used to fit models for each tree species with a typical mapping zone. Cubist produces a set of rules, each associated with a multivariate linear model. The rule-based output is used to produce a geo-spatial data set using freely distributable code that was integrated with ERDAS Imagine Software (Ruefenacht et. al. 2007). The resultant geo-spatial data is a 30-meter raster dataset depicting a continuous range of either basal area or trees per acre values.

2.7 EXTIRPOLATED MODELS

For most individual species, subplot occurrence is infrequent and difficult to model. The resultant models typically only have a few values. For this reason, a new technique was used on models with r-values of less than 0.50. The forest parameter of interest (basal area, trees per acre) is subtracted from the total for all species and then modeled through cubist, $BA_{t-si} = BA_t - BA_{si}$. A geospatial product is built from the resulting model. This product is then subtracted from the total of all species geospatial product to calculate the original species parameter, $Y_{si} = Y_t - Y_{t-si}$. For ease of discussion, it is proposed to term this procedure as an “extirpolated” model. An example of the normal model versus the extirpolated model is illustrated in the first two panels of Figure 2.

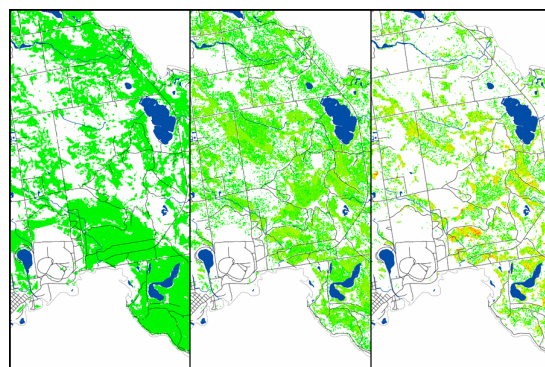


Figure 2. Normal vs extirpolated vs adjusted models

3 RESULTS

The overall purpose of this project is to produce a 30-meter species level forest parameter dataset for

use in the 2010 National Insect and Disease Risk Map. Interim goals for regional projects have been tasked. To date three regional forest parameter modeling projects have been completed and resultant maps have been produced for 26 of the 78 mapping zones as of September 30, 2007.

3.1 CUBIST MODELS

Over 10,000 individual cubist models have been produced in the 26 completed mapping zones. Of these, 5,429 models have been selected for geospatial product generation (Table 1). The strongest models are typically the total basal area with r-values ranging from 0.58 to 0.77. For total trees per acre, r-values range from 0.35 to 0.61. Models for individual tree species rarely exceed r-values of 0.50 and most need the extirpolated model to produce a valid modeled parameter. Extirpolated model r-values typically match the corresponding total parameter model.

Table 1. R-value summary of models

Parameter	Max	Min	Avg	Total
Basal area	0.77	0.58	0.68	27
Trees per acre	0.61	0.35	0.50	27
# of models per zone	304	58	201	5429
# of species per zone	151	28	100	2688

3.2 GEOSPATIAL PRODUCTS

Only about 350 forest parameter surfaces have been computed so far, due to the required computing hours. Some of these models were validated against the original ground samples and additional field datasets during local stand exams. Though most of the geospatial products captured the variable density for a species parameter, the overall parameter variation seemed to be somewhat under-represented, as regressions tend to “smooth” the data around the mean. In order to mitigate this tendency, for each of the forest parameter layers generated, a histogram balance-adjustment is performed to scale the product to match the distribution of FIA plot data (Lister and Lister, 2006). An example is illustrated in the third panel of Figure 2.

3.3 FIELD VALIDATION

Several zones were analyzed to meet immediate needs for deployment of survey strategies for recent insect outbreaks. Species parameter maps were used to narrow sampling areas for the sirex woodwasp in central New York and emerald ash borer in

Michigan, Ohio, and Indiana. Feedback from survey efforts suggests the forest parameter maps did increase survey efficiency and effectiveness.

Southern pine beetle (SPB) is a key threat to the pine forests of the southern United States. Efforts to control the impact of the southern pine beetle require extensive preventative treatments. In order to prioritize treatment areas, a hazard assessment was developed across 11 ecological mapping zones. Forest parameter mapping was essential to this assessment. Figure 3 illustrates Loblolly pine basal area, which is one of 11 host-species that were modeled in the 11 USGS mapping zones. Field validation showed very high correspondence among the dominant pine species. Some of the less common pine species also showed good correspondence where SSURGO soil data and recent FIA data were available.

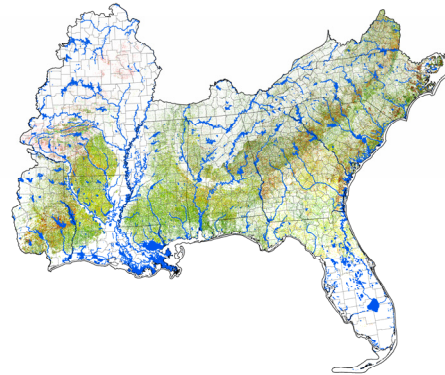


Figure 3. Southern pine basal area

3.4 COMPLETION

For select species within the completed zones, draft forest parameter geospatial products are currently available. Completion of the remaining zones and species is anticipated by 2009.

4 SUMMARY

The final products will provide better representation of forest parameters and support the 'next-generation' risk mapping effort. This will help to bridge the gap between national-scale planning and on-the-ground conditions. The risk map has been a tool through which to communicate national forest health concerns to policy makers. With finer resolution data, it can now be a tool for land management planners to better represent forest health concerns in the forest planning process. Through these mechanisms, better and more effective treatments can be implemented. Caution must be applied when utilizing these products at resolutions smaller than 30-meters, such as would be necessary for project level planning. The 30-meter pixel does not represent individual trees, though there may be utility for application at the stand level by aggregating 30-meter pixels for stand-level summaries. Aggregation for multiple species can be utilized for creating dynamic type maps depending upon end-user needs.

ACKNOWLEDGMENTS

We would like to thank Colin Homer, Elizabeth LaPointe, Dorn Egly, and Randy Schaeztl for all their support in providing us with data.

REFERENCES

- Bechtold, W., and Patterson, P. 2005. The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures. *Gen. Tech. Rep. SRS-80*. Asheville, NC, USDA Forest Service, Southern Research Station. 85 p.
- Daymet, 2007. <http://www.daymet.org>. (accessed 28 Sep. 2007).
- Homer, C, et. al. 2007. *Photogrammetric Engineering & Remote Sensing* Vol. 73, No. 4, April 2007, pp. 337-341.
- Krist, F., Sapio, F., and Tkacz, B. 2007a. A Multi-Criteria Framework for Producing Local, Regional, and National Insect and Disease Risk Maps. *In: Advances in Threat Assessment and their Application to Forest and Rangeland management*, Pye, J.M.; Rauscher, H.M.; Sands, Y.; Lee, D.C.; Beatty, J.S., eds. Gen. Tech. Rep. PNW. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. (In press).
- Krist, F., Sapio, F., and Tkacz, B. 2007b. Mapping Risk from Forest Insects and Diseases. USDA-Forest Service FHTET (In press).
- Lister, A., and Lister, T. 2006. Post-Modeling Histogram Matching of Maps Produced Using Regression Trees. *In: 2004 Proceedings of the Sixth Annual Forest Inventory and Analysis Symposium*, pp. 111-117.
- NRCS, 2007. USDA Natural Resources Conservation Service Soil Survey Staff. U.S. General Soil Map and SSURGO. <http://soildatamart.nrcs.usda.gov>. (accessed 28 Sep. 2007).
- Ruefenacht, B., Finco, M., Nelson, M., Czaplowski, R., Helmer, E., Blackard, J., Holden, G., Lister, A., Salajanu, D., Weyermann, D., Winterberger, K. 2007. Conterminous US and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. *Photogrammetric Engineering and Remote Sensing*. (In press).
- RuleQuest 2007. <http://www.rulequest.com>. (accessed 28 Sep. 2007).
- Schaeztl, R. 1986. Soilscape Analysis of Contrasting Glacial Terrains in Wisconsin. *Annals of the Association of American Geographers* 76(3):414-425.
- Schaeztl, R., Krist, F., and Stanley, K. 2007. The Natural Soil Drainage Index – An ordinal estimate of water availability in soils. *Unpublished manuscript on file at FHTET*.
- Stage, A. 1976. An expression for the effect of aspect, slope, and habitat type on tree growth. *Forest Science*. 2:457-460.
- Swets, D., Reed, B., Rowland, J., and Marko, S., 1999. A weighted least-squares approach to temporal NDVI smoothing, *Proceedings of the 1999 ASPRS Annual Conference*, Portland, Oregon, May 17-21, 1999
- Tarr, A. 2000a, Magnetic Field – Declination Component for the Epoch 1995.0: U.S. Geological Survey, Reston, VA.
- Tarr, A. 2000b, Magnetic Field - Secular Variation of the Declination Component for the Epoch 1995.0: U.S. Geological Survey, Reston, VA.
- Thornton, P., Running, S., and White, M. 1997. Generating surfaces of daily meteorology variables over large regions of complex terrain. *Journal of Hydrology* 190:214-251.
- USFS, 2007. USDA Forest Service, Forest Inventory and Analysis. <http://fia.fs.fed.us/library/field-guides-methods-proc/>, (accessed 28 Sep. 2007).
- USGS, 2007. USDI- USGS Mapping Division. National Elevation Dataset. <http://ned.usgs.gov/Ned/index.asp>, (accessed 28 Sep. 2007).