EXTRAPOLATING INTENSIFIED FOREST INVENTORY DATA TO THE SURROUNDING LANDSCAPE USING LANDSAT

Evan B. Brooks, John W. Coulston, Valerie A. Thomas, and Randolph H. Wynne

Abstract—In 2011, a collection of spatially intensified plots was established on three of the Experimental Forests and Ranges (EFRs) sites with the intent of facilitating FIA program objectives for regional extrapolation. Characteristic coefficients from harmonic regression (HR) analysis of associated Landsat stacks are used as inputs into a conditional random forests model to form predictive models of key forest biophysical parameters for use in making wall-to-wall maps.

In 2011, the FIA program established intensified plots in several of the Experimental Forests and Ranges (EFRs). These plots resemble standard FIA phase 2 plots as described in Bechtold et al. (2005), but they are much denser per unit area, with roughly 50 plots in each EFR. The motivation behind the intensified sampling was to facilitate the FIA mission by having a large number of plots on representative sections of the nation’s forests, enabling the extrapolation of forest biophysical parameter estimates into the areas around the forests.

The Landsat collection of satellite imagery is a common choice for regional extrapolation of point forest data, primarily because it has the spatial resolution to monitor individual stands and has a decades-long record to compare with historical FIA plot measurements. Multitemporal approaches are not subject to issues resulting from the choice of a single image for analysis, but they must contend with missing or poor-quality data due to striping from the scan line corrector failure on Landsat 7 in 2003, as well as clouds, shadows, and other obscuring factors.

Harmonic regression (HR; Brooks et al., 2012) is an algorithm designed to interpolate missing data in large multitemporal image stacks. HR fits curves based on the superposition of sinusoidal curves of different frequencies (harmonics), independently to each pixel. Each HR curve is characterized by a collection of data-driven coefficients. While these coefficients are generally used to generate simulated Landsat-scale images, they have value in their own right. In addition to straightforward applications like land cover/land use classification, the coefficients have also been used to improve the precision of forest parameter estimation (Brooks et al., 2015). Thus, given training data in the form of spatially explicit FIA plot measurements, it is our contention that they may also be used as predictors in a model for forest biophysical variables. Additionally, because the coefficients are not specific to any one image, we expect that once trained, models based on them may be applied to Landsat data from other years as well.

STUDY AREA

Our study focused on three EFRs which formed an approximate line, covering roughly the range of elevation in the southeastern US. (Fig. 1) Specifically, we looked at the Coweeta Hydrologic Laboratory in southwestern North Carolina, the Calhoun Experimental Forest in north-central South Carolina, and the Santee Experimental Forest in South Carolina, near the Atlantic coast. Each of these EFRs has approximately 50 intensified FIA plots, all established so as to avoid overlapping other ongoing experiments in the EFRs. The plots were measured in 2011, the measurements being a slight subset of the biophysical metrics commonly found in FIA.
phase 2 plots. Note that the term “intensified” refers to the increased spatial density of the plots, not to the degree of measurements taken within the plots. Table 1 lists some of the variables measured in the intensified plots.

Table 1—Selected biophysical variables and parameters in the intensified FIA plot data.

<table>
<thead>
<tr>
<th>Condition table variables</th>
<th>Tree table variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owner, forest type, stand</td>
<td>Species, species group,</td>
</tr>
<tr>
<td>age, site index, stocking,</td>
<td>DBH, height, age, volume,</td>
</tr>
<tr>
<td>disturbance type</td>
<td>biomass</td>
</tr>
</tbody>
</table>

In order to extrapolate the information contained within the EFRs’ intensified plots to the broader landscape, we used Landsat thematic mapper (TM) and enhanced thematic mapper plus (ETM+) images, covering the transect with data from WRS-2 path/row combinations 18/36, 17/36, 17/37, and 16/37. We acquired every available image from 2010 and 2011, regardless of image quality, choosing images that were already processed to L1T standards and also corrected to surface reflectance via the Landsat Ecosystem Disturbance Adaptive Processing System. (LEDAPS, Masek et al. 2006)

We then obtained spatial subsets corresponding to the three EFR boundaries for model training, using FMask (Zhu and Woodcock, 2012) from the post-LEDAPS product to filter out the majority of cloud and shadow-related pixels. Subsequently, we filled in the missing data gaps within the resulting images with window regression. (WR, de Oliveira et al., 2014)

Because the spatial size of the FIA plots is larger than a single 30m Landsat pixel, we took the additional step of computing 3x3 pixel window averages, by layer, to ensure that each pixel reflected the immediate neighborhood that would comprise an FIA plot. We then used these stacks as inputs into HR, thus obtaining characteristic coefficient rasters for each subset.

We treated each of the seven spectral bands separately, obtaining a collection of coefficient rasters for each band in addition to a raster for normalized difference vegetation index. (NDVI, Tucker, 1979) For each band and index, we fitted the data to a two-harmonic curve, obtaining five coefficients (constant, sin(t), sin(2t), cos(t), and cos(2t)) in each case, resulting in 40 coefficient layers total for each EFR. Preprocessing was done using R (R Core Team, 2014), with emphasis on the spatial.tools package and its dependencies. (Greenberg, 2014)

**METHODS**

We treated the HR coefficients as predictors of the measured values from the intensified plots in the EFRs boundaries. Thus, we first joined the plot measurements to the associated pixels in the HR coefficient stacks by the spatial location of the plot,
in each case using the pixel which corresponded to the recorded location. Due to the correlated nature of many of the original Landsat spectral bands and the inclusion of NDVI-based coefficients as potential predictors, we used conditional random forests from the party package as the basis for our model-fitting. (Strobl et al., 2008)

After a final predictive model is derived, we will apply that model to the HR coefficients obtained by applying HR across the full extent of the study area. Where possible, the resulting predictions will be compared with FIA Phase 2 plot data from the study area.

RESULTS
Currently, processing and preliminary model-fitting are complete for the EFRs subset stacks. These results show R² values for quantitative variables such as height, Carbon above ground, and age on the order of 55 to 64 percent. Similarly, the misclassification rate for the species group is 22.3 percent. These values seem promising when one considers that the only predictors used were products derived from multitemporal satellite data. Further comparison with FIA Phase 2 plots across the study area is planned, pending calculation of HR coefficients across the region.

DISCUSSION
The spatially dense nature of the intensified plots made computation of the HR coefficients simpler. While we will utilize broad coverage from all four input scenes, the ability to crop out the subsets around the EFRs made the processing and model training much more efficient. This in its own right is computationally valuable, and when coupling this fact with the public availability of the exact spatial coordinates of the plots, it makes the intensified plots in the EFRs convenient for this sort of extrapolation effort.

While the HR coefficients were trained on the 2010-2011 period, the coefficients themselves are simply characteristic of the corresponding curves. If such curves represent commonly occurring phenologies, then it reasonable to assume that one may use the models from this study with HR coefficients from different years. This possibility makes the intensified plots that much more potentially valuable.

Complications may arise, however, from the choice of intensified plot locations within the EFRs. The locations of intensified plots were chosen to avoid intersecting with other experiments currently being conducted in the EFRs. As a result, the models we built were fitted to forests that were undisturbed after the establishment of the EFRs. Accordingly, we expect comparisons of model predictions with the more general Phase 2 plots to have a considerable amount of disagreement. In order to extend the effectiveness of the models, additional intensified plots, covering a broader range of treatments, would be helpful.

ACKNOWLEDGMENTS
This work was supported by the USDA Forest Service Cooperative Agreement with Virginia Tech (Grant No. 10-CA-11330145-158). It was also supported by the Landsat Science Team (USGS contract number G12PC00073), the Pine Integrated Network: Education, Mitigation, and Adaptation Project (PINEMAP, Coordinated Agricultural Project funded in 2011 by the USDA National Institute of Food and Agriculture), the McIntire-Stennis Cooperative Forestry Research program (USDA CSREES, Project No. VA-136614), and the Department of Forest Resources and Environmental Conservation at Virginia Tech.
LITERATURE CITED


