ERRORS IN TERRAIN-BASED MODEL PREDICTIONS CAUSED BY ALTERED FOREST INVENTORY PLOT LOCATIONS IN THE SOUTHERN APPALACHIAN MOUNTAINS, USA

HUEI-JIN WANG¹, S. P. PRISLEY¹, P. J. RADTKE¹, J. COULSTON²

¹Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA 24061, USA
²USDA Forest Service, Southern Research Station FIA, Knoxville, TN 37919, USA

ABSTRACT. Forest modeling applications that cover large geographic areas can benefit from the use of widely-held knowledge about relationships between forest attributes and topographic variables. A noteworthy example involves the coupling of field survey data from the Forest Inventory Analysis (FIA) program of USDA Forest Service with digital elevation model (DEM) data in attempts to explain how topographic characteristics influence forest productivity, vegetation composition, fire behavior, and other phenomena. Because U.S. federal law prohibits the release of actual FIA plot coordinates, only altered coordinates are released to the public. Here, terrain-based variables derived from a 10 m DEM using actual FIA plot locations were compared to those from altered plot locations in a region of the Southern Appalachian Mountains of western North Carolina, USA. Variables examined included simple terrain attributes such as percent slope and azimuth of aspect, and composite attributes such as terrain shape index, flow accumulation, slope position and forest site quality index. Results showed little correspondence between variables from altered plot locations compared to those derived using actual locations. Further, FIA field measurements of percent slope and azimuth of aspect showed little correspondence with corresponding DEM-derived estimates from the actual plot coordinates. In order to effectively link FIA plot data with DEM-derived topographic variables in mountainous regions like the Southern Appalachians, access to actual plot coordinates or terrain variables derived from them may be required.

Keywords: uncertainty; errors in measurements; GIS; FSQI.

1 INTRODUCTION

Forest scientists have known for decades that strong linkages exist between the topographic features of land-]scapes and forest stand attributes including productivity, species composition, fire behavior, and other phenomena. Many efforts have attempted to quantify these linkages to better model and predict forest productivity for assessment and management (Bolstad et al. 2001; Davis and Goetz 1990; Franklin 1995). The availability of advanced spatial analysis software and consistent, reliable, national coverage of digital elevation model (DEM) data has facilitated our ability to characterize and quantify topographic conditions at locations where productivity or other forest attributes are measured. With regard to forest characteristics across large geographic areas, it is advantageous to use broad-scale inventory data coupled with DEM-derived topographic variables for developing models to estimate site productivity (Bolstad et al. 1998; Deng et al. 2007; Iverson et al. 1997).

In conducting the national forest inventory of the United States of America (USA), the Forest Inventory Analysis (FIA) program of the USDA Forest Service collects field plot data and information about forest resources of the USA (Smith 2002). In compliance with the U.S. Food Security Act of 1985 modified by the year 2000 Consolidated Appropriations Bill, publicly available FIA data do not include actual plot coordinates (LaPoint 2005; McRoberts et al. 2005). Plot locations are altered by adding a random error of less than 1.6 km to the locations of all plots and swapping the locations of a small proportion of similar plots (Guldin et al. 2006; Lister et al. 2005). Terminology for perturbing and swapping has been referred to differently in different contexts (Coulston et al. 2006b; Lister et al. 2005). We’ve adopted the terminology used by McRoberts et al. (2005) in referring to the methods of perturbing and swapping that the FIA program uses to alter plot co-
ordinates prior to releasing them to the public domain. Further, we refer to perturbed and swapped plot locations as “altered” in contrast with the unaltered positions which we consider to be “actual.”

While aimed at balancing landowner privacy with data users’ needs to link spatial information to FIA field data, the effects of altering coordinates may be significant when public-domain plot locations are used to predict terrain-based attributes. McRoberts et al. (2005) suggested that effects of using altered plot locations were non-negligible when forested areas of interest were less than 321,700 ha. Coulston et al. (2006a; 2006b) showed that perturbed and swapped coordinates were adequate for predicting plot attributes at coarse spatial resolutions when linking to ancillary geospatial data, but not necessarily at fine resolutions. They concluded that predictions of forest-related attributes at fine spatial scales using altered plot locations may be inaccurate.

Raster grid sizes on the order of 10 m (10 m × 10 m) to 30 m (30 m × 30 m) are commonly adopted in many applications that assign DEM-derived terrain attributes to field plot data. Such grid sizes are primarily based on DEM data sets readily available from public sources. Fine resolution (≤ 30 m) data may be warranted in landscape-level modeling applications because terrain-based variation is likely underestimated using low-resolution terrain data (Bolstad et al. 1998). Topographic variables derived from DEMs have been applied to terrain-based estimation procedures of forest site productivity, accounting for slope, aspect, and water-movement relevant variables such as terrain shape, downhill flow accumulation, and slope position (Iverson et al. 1997; McNab 1993; Meiners et al. 1984; Stage 1976). When DEM-derived topographic variables are paired with productivity measures observed on field plots, a concern arises that altered plot coordinates may cause substantially poorer model performance (Coulston et al. 2006b).

Effects of topographic variables on forest productivity in the Southern Appalachian Mountains of the USA have been known for decades (McNab 1993; Whittaker 1966). Because of the strength of relationships between topography and productivity in the Southern Appalachian Mountains, the region provides an ideal setting for an assessment of the effects of altering plot locations on geospatial modeling. Previous work has identified elevation and/or terrain shape as main determinants of basal area and forest composition (Bolstad et al. 1998). Aspect and slope position – closely related to differences in microclimate, soil moisture, and soil depth – are significantly associated with site quality for upland oaks (Carmean 1970). Tree height is known to be correlated with a terrain shape index in southern Appalachian deciduous forests (McNab 1989), as is leaf area index to terrain position (Bolstad et al. 2001). Meiners et al. (1984) developed and tested a model of productivity-related soil properties based on simple terrain variables including slope, aspect, and slope position.

The objective of this study was to quantify the magnitudes of errors in forestry-related terrain-based model predictions for FIA plot locations when altered coordinates were used in place of actual ones. The goals included the examination of several simple terrain variables including slope and aspect, along with some model-based attributes including the forest site quality index (FSQI) of Meiners et al. (1984). A further goal involved the assessment of the potential for interchanging FIA field-based terrain measurements with DEM-derived estimates. Favorable comparison of field measurements with DEM-derived values from actual plot locations was seen as a potential means to overcome any limitations of working with the same attributes derived from publicly available altered plot locations.

2 Methods

2.1 Study Area This study was conducted using data from the FIA survey area that covers the mountainous region of western North Carolina, USA (Figure 1). This area comprises 21 counties covering approximately 17,870 km². Most of the study area is in the Blue Ridge physiographic province of the Southern Appalachian Mountains (Fenneman and Johnson 1946). Elevations in the study area range from 266 to 2,033 m. Forests are dominated by mixed upland hardwood communities, with some oak-yellow poplar and oak-hickory communities (Forest Inventory and Analysis 2009). Chestnut oak (Quercus prinus L.) occurs frequently across a broad range of elevations and terrain shapes in the region, and it is particularly dominant on xeric upland sites such as ridges and upper slopes (Bolstad et al. 1998). Mesophytic species such as yellow poplar (Liriodendron tulipifera L.) typically dominate the valleys and moist slopes (McNab 1993; Whittaker 1956). Northern red oak (Quercus rubra L.), white oak (Q. Alba L.), and hickory (Carya spp.) exhibit dominance across a wide range of sites having moisture regimes that range between xeric and mesic (Whittaker 1956).

2.2 FIA Data FIA field data from North Carolina measured between 1998 and 2002 were obtained from the USDA Forest Service through a memorandum of understanding as part of the FIA Privacy Policy Study Group. Both altered (perturbed and swapped) and actual coordinates for each plot were obtained to facilitate the comparisons of interest. Data from a total of 1,022 field plots were used in the analysis. Due to a lack of field-measured terrain data, a subset of 767 plots was
used to provide field measurements of percent slope and azimuth of aspect.

2.3 Terrain Data Digital elevation data were obtained from the US Geological Survey “seamless” website (http://seamless.usgs.gov, 9/1/2008) for 10 m grid size DEM coverage of the study area. From this dataset, additional terrain datasets were developed using spatial analysis functions available in ArcGIS 9.2 software (Environmental Systems Research Institute, Inc., Redlands, California). Percent slope was computed using the ArcGIS slope tool, which applies the Horn (1981) algorithm using elevations at eight adjacent cells. Azimuth of aspect was calculated using the default ArcGIS algorithms. To avoid difficulties in working with 0° - 360° azimuth values, aspect was also transformed using a cosine transformation that assigned values of 1 for due north, -1 for due south, and 0 for due east and west.

Several composite terrain variables were calculated from the DEM data, including terrain shape index, flow accumulation, and slope position. These three variables can be used to identify landforms such as valleys, sideslopes, and ridges known to have effects on forest composition and productivity in the Southern Appalachian Mountains. Terrain shape index (McNab 1989) was computed as the difference between the elevation at a cell and the average elevation of a circular neighborhood with a radius of 37 m, corresponding to the coverage area of a FIA field plot. Negative terrain shape indices indicate convex upper slopes while positive values indicate concave lower slopes. The flow accumulation algorithm from ArcGIS was used to compute the number of grid cells sending surface runoff downslope to the cell being evaluated – assuming surface impermeabil-
ity. Flow accumulation values are generally zero along ridges and large near stream channels. Slope position was computed as an index bounded between 0 (valley) to 1 (ridge) by dividing the flow path distance from FIA plot center downhill to the nearest stream by the flow path distance from the nearest ridge to stream using the flow length tool in ArcGIS.

2.4 Terrain-based Model of Forest Productivity

The FSQI developed by Meiners et al. (1984) was used to predict site productivity from DEM-derived attributes of each FIA plot. The index assigns ordinal ranks corresponding to relative site quality for each of three attributes, slope, aspect, and slope position, which reflect moisture availability in Southern Appalachian forest soils. The FSQI is then computed as the unweighted sum of the ordinal values. Slope was ranked into one of five integer classes with 1 corresponding to steepest (> 60%) slopes where forest productivity is typically low, and 5 corresponding to least steep (< 15%) slopes where productivity is typically high in Southern Appalachian forests. Aspect was ranked similarly into classes ranging from 1 to 6, with southerly and westerly aspects having low values to reflect the relatively high evaporative demands associated with low forest productivity. Slope positions were ranked from 1 to 5 for classes representing, in ascending order, the following classes: shoulders (DEM-derived slope position between 0.75 - 0.9); backslopes (0.25 – 0.75); summits (0.9 – 1.0); footslopes (0.1 – 0.25); and toeslopes, terraces, or floodplains (Meiners et al. 1984).

2.5 Analysis Approach

Data listing actual and altered plot coordinates were projected to a suitable coordinate system so they could be overlaid on DEM-derived terrain layers in a GIS. Cell values from the DEM-based terrain layers were subsequently assigned to each plot location and compared between actual and altered locations. Scatterplots were used to evaluate whether or not linear relationships existed between DEM-derived terrain variables from actual versus altered plot locations. Pearson product-moment correlation, denoted \( r \), was used to measure the strength of the linear relationship between DEM-derived terrain variables at actual and altered locations, and Spearman rank correlation \( r_s \) was used to account for nonlinear relationships. Statistical significance was determined by Fisher’s Z transformation and a large-sample approximation to the normal distribution (Best and Roberts 1975; Fisher 1915; Higgins 2004).

Plot FSQI values were assigned into four classes: (1) xeric sites, FSQI = 3-6; (2) subxeric sites, FSQI = 7-10; (3) submesic sites, FSQI = 11-13; (4) mesic sites, FSQI = 14-16. Categorical FSQI site index classes for actual and altered plot locations were compared in a contingency table. The kappa statistic was used to examine the strength of FSQI agreement between actual and altered plot locations, adjusting for the chance agreement due to randomness (Stehman 1997). Using the kappa statistic, the validity of substituting actual with altered plot locations was evaluated with the strength of agreement ranging from +1 (perfect agreement) to 0 (poor agreement or agreement by chance) (Landis and Koch 1977).

3 Results

3.1 Actual versus Altered Plot Coordinates

Perturbing and swapping plot locations noticeably affected DEM-derived topographic attributes. Correlation coefficients for terrain shape index and flow accumulation did not show significant evidence \( (\alpha = 0.05) \) of positive correlation between actual and altered locations (Table 1). Percent slope, slope position, and cosine-transformed aspect values from actual and altered plot locations were weakly correlated by either correlation measure \( r \) or \( r_s \) (Table 1). Elevation values were the most consistent of the variables examined between actual and altered plot locations (Table 1, Table 1). Regardless of the strength or weakness of correlation, differences between DEM-derived variables obtained at actual and altered plot locations had wide ranges, with correspondingly large sums of squared errors (Table 1).

An example of differences between actual and altered plot locations was generated to illustrate typical discrepancies observed in data when a plot’s location was altered (Figure 2). The example simulates a perturbation of 670 m between actual and altered plot locations in a mountainous section of the study area. Elevations of 1043 m and 785 m were obtained from the DEM at the actual and perturbed plot coordinates, respectively. Percent slopes of 99% (actual) and 2% (perturbed) were calculated from the DEM, along with aspect angles of 88˚ (actual) and 275˚ (perturbed).

3.2 FSQI Model Agreement

FSQI consisted of scores from 3 (most xeric) to 16 (most mesic). Only 15.4% of plots had the same scores for actual and altered locations. The Pearson product-moment correlation showed a weak but statistically significant positive correlation \( r = 0.167 \) \( (p < 0.01) \). When aggregated to four classes, FSQI exhibited 48% overall agreement (Figure 3); however, the strength of agreement was slight \( (\text{Kappa} = 0.04) \) when excluding the agreement by chance.

3.3 Field-Measured versus DEM-Derived Variables

Inconsistencies were noted when comparing FIA field-observed slope and aspect to DEM-derived values
Table 1: Pearson product-moment correlation (r), Spearman rank correlation (r_s), and root mean squared error (RMSE) for comparison of the effects of actual versus altered plot locations in calculating DEM-derived variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pearson r</th>
<th>p-value</th>
<th>Spearman r_s</th>
<th>p-value</th>
<th>RMSE†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation (m)</td>
<td>0.928</td>
<td>&lt; 0.001</td>
<td>0.930</td>
<td>&lt; 0.001</td>
<td>113.06</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>0.383</td>
<td>&lt; 0.001</td>
<td>0.395</td>
<td>&lt; 0.001</td>
<td>23.14</td>
</tr>
<tr>
<td>Cosine of aspect</td>
<td>0.093</td>
<td>0.003</td>
<td>0.080</td>
<td>0.010</td>
<td>0.95</td>
</tr>
<tr>
<td>Terrain shape index</td>
<td>0.036</td>
<td>0.247</td>
<td>0.019</td>
<td>0.547</td>
<td>4.37</td>
</tr>
<tr>
<td>Flow accumulation</td>
<td>-0.006</td>
<td>0.860</td>
<td>0.036</td>
<td>0.249</td>
<td>2812.12</td>
</tr>
<tr>
<td>Slope position</td>
<td>0.201</td>
<td>&lt; 0.001</td>
<td>0.084</td>
<td>0.008</td>
<td>0.18</td>
</tr>
</tbody>
</table>

†same units as variable. Refer to Table 1 for ranges and scales of observed data.

4 Discussion and Conclusions

Differences are substantial when perturbed and swapped plot locations are used in place of actual plot locations to derive terrain variables from a 10 m DEM in the Southern Appalachian Mountains, USA. It is widely known that microscale variation in landform and terrain affects temperature and soil moisture, thus influencing plant growth, especially in mountainous regions (Baily 2009; Roise and Betters 1981; Stage and Salas 2007). Applying this knowledge to the results obtained here, perturbed and swapped plot locations may not prove useful when modeling forest productivity, species composition, or other attributes related to terrain variables because the fine-scale information from 10 m DEMs does not generally apply equally at both actual and altered plot locations. Although the FSQI model has been applied in forest productivity research of upland oak forests in the Southern Appalachians (Bolstad et al. 1998; Ross et al. 1986), little evidence was found here to support that FIA altered plot locations could be used to calculate FSQI in lieu of actual plot locations.

The number of terrain based, GIS derived variables currently used in ecological and hydrological applications is large and growing (Horsch 2003; Tenenbaum et al. 2006). Plot locations released to the public by the FIA program are altered by perturbation and swapping as required by law; therefore, obtaining fine scale topographic variables for FIA field plots is problematic. One possibility that may serve end-users’ needs is for the FIA database to incorporate some DEM-derived values from actual plot locations in the publicly released data sets. As noted by Liknes et al. (2005), such an approach must be evaluated first for the possibility that it would com-

Table 2: The contingency table of forest site quality index (FSQI) for actual versus altered plot coordinates, accounting for DEM-derived percent slope, azimuth of aspect and slope position.

<table>
<thead>
<tr>
<th>Actual locations</th>
<th>FSQI 3-6</th>
<th>7-10</th>
<th>11-13</th>
<th>14-16</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-6</td>
<td>15</td>
<td>42</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>7-10</td>
<td>51</td>
<td>371</td>
<td>183</td>
<td>13</td>
</tr>
<tr>
<td>11-13</td>
<td>12</td>
<td>184</td>
<td>98</td>
<td>10</td>
</tr>
<tr>
<td>14-16</td>
<td>3</td>
<td>11</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percent slope</th>
<th>Error (%)</th>
<th>Plots (%)</th>
<th>Azimuth of aspect</th>
<th>Error (˚)</th>
<th>Plots (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>±5</td>
<td>31.2</td>
<td>±5</td>
<td>12.3</td>
<td>±5</td>
<td>12.3</td>
</tr>
<tr>
<td>±10</td>
<td>52.8</td>
<td>±10</td>
<td>21.3</td>
<td>±10</td>
<td>21.3</td>
</tr>
<tr>
<td>±15</td>
<td>68.8</td>
<td>±15</td>
<td>32.5</td>
<td>±15</td>
<td>32.5</td>
</tr>
<tr>
<td>±20</td>
<td>79.5</td>
<td>±20</td>
<td>39.9</td>
<td>±20</td>
<td>39.9</td>
</tr>
<tr>
<td>±25</td>
<td>87.2</td>
<td>±25</td>
<td>45.8</td>
<td>±25</td>
<td>45.8</td>
</tr>
<tr>
<td>±30</td>
<td>91.4</td>
<td>±30</td>
<td>50.7</td>
<td>±30</td>
<td>50.7</td>
</tr>
</tbody>
</table>
Figure 2: Scatterplots of terrain variables for actual and altered plot locations: elevation, percent slope, cosine of aspect, terrain shape index, flow accumulation, and slope position (sample size = 1022).

promise the confidentiality of FIA plot locations.

The potential for interchanging FIA field terrain measurements with DEM-derived estimates is somewhat unfavorable for either developing or applying topographic-based models as well. According to the FIA (2007) field manual, allowable tolerances for percent slope and azimuth of aspect errors were $\pm 10\%$ and $\pm 10^\circ$, respectively, on at least 90% of plots measured. Even though actual plot locations were used, DEM-derived slope and aspect did not meet these accuracy standards (Figure 4). The algorithms used in GIS-based terrain analysis do not measure slope and aspect in the same ways as they are measured in FIA field procedures, so this comparison should not be taken as evidence that FIA standards for field measurement accuracy are not being met. We make the comparison here to inform readers that field measurements of terrain variables in the FIA database are not interchangeable with DEM-derived values in the Southern Appalachian Mountains.

Opportunities for spatial modeling by linking FIA data to ancillary geospatial data have increased along with the availability of geospatial data, e.g., DEMs (Davis et al. 2007; Tirpak et al. 2009; Zarnetske et al. 2007). These opportunities will continue to increase as more geospatial information at higher resolution becomes available. Regardless of the need for and use of accurate ancillary spatial datasets, the usefulness of such datasets is related to the accuracy of the plot locations. Although errors in actual locations were not considered in this analysis, these errors create additional uncertainty for the types of analyses presented here (McRoberts 2010). For example, the FIA program typically uses recreation-grade GPS units to collect coordinates of each inventory plot. These types of GPS units have under-canopy positional accuracies of approximately 7.5 m and a maximum error of 21.5 m (Bolstad
et al. 2005; Forest Inventory and Analysis 2007). Given the positional uncertainties, the FIA program might consider reporting accuracy of the actual locations, when available, to those clients who have access within a confidentiality agreement to work with actual plot coordinates. This information is particularly important when conducting research to develop models based on fine resolution spatial data such as DEMs and remote sensing imagery. In keeping pace with the increased resolution and spatial precision of publicly available spatial data sets, researchers and end-users must be aware of the limitations that perturbing and swapping of FIA field plot coordinates impose on applications involving fine-scale geospatial analyses.

ACKNOWLEDGEMENTS

This research was supported by the USDA Forest Service, Southern Research Station under Agreement No. SRS 05-CA-11330134-251. The authors gratefully acknowledge the support of the Southern Research Station, specifically Cathryn Greenberg and David Loftis. We also acknowledge the valuable assistance given by Bill Burkman and Sam Lambert of the SRS FIA unit, and Claudia Cotton while she was a graduate student in the Virginia Tech Department of Forest Resources and Environmental Conservation.

REFERENCES


Figure 4: Comparison of FIA field measurements and DEM-derived values using actual plot locations: Pearson product-moment correlation $r = 0.61$ for percent slope; Spearman rank correlation $r_s = 0.54$ for cosine of aspect (sample size = 767).

McWilliams (eds). USDA, Forest Service, Denver, CO.


Forest Inventory and Analysis. 2007. Forest inventory and analysis national core field guide - Volume 1: field data collection procedures for phase II plots version 4.0. USDA Forest Service (ed). Forest Inventory and Analysis National Office.

Forest Inventory and Analysis. 2009. FIA DataMart - FIA-ADB version 4.0. USDA Forest Service, Forest Inventory and Analysis National Program, Arlington VA.


Higgins, J.J. 2004. Introduction to Modern Nonpara-


