EVALUATING THE COMpatibility OF AMERICAN AND MEXICAN NATIONAL FOREST INVENTORY DATA

Todd A. Schroeder, Sean P. Healey, and Gretchen G. Moisen

ABSTRACT

The international border region between the United States and Mexico represents a point of discontinuity in forest policy, land use management and resource utilization practices. These differences along with physical barriers which separate the two countries can interact to alter the structure and functioning of forest vegetation. One valuable source of information for analyzing potential effects of management on forest attributes is National Forest Inventory (NFI) data. Both Mexico and the United States have systematically designed NFI programs, the U.S. Forest Service Forest Inventory and Analysis (FIA) program and the Comisión Nacional Forestal (CONAFOR) Inventario Nacional Forestal y de Suelos (INFyS). However, data from NFIs are seldom harmonized with respect to reporting units, field procedures and estimation methods. Here we evaluate two important aspects of NFI data compatibility using seamless geospatial data. First, to gauge plot measurement and location accuracy we compared the elevations recorded in each country’s NFI database with those taken from an independently derived digital elevation model (DEM). Second, basal area compatibility was determined by means of analysis of covariance (ANCOVA) using a seasonal time series of normalized difference vegetation index (NDVI) data from Landsat. The results showed that both countries have good location and measurement accuracy in relation to DEM elevations and in the majority of cases, statistically similar estimates of basal area per unit of NDVI. Despite finding a high level of plot data compatibility, our study uncovered key differences in inventory stratification between the two countries which prevented further statistical comparison of oak woodland stand densities. Suggestions for improving local and regional scale analysis compatibility of American and Mexican NFI data are provided.

INTRODUCTION

In response to interest concerning the effects of global climate change there is a growing need for information on the health, status, and biodiversity of the world’s forest resources. In many countries, the current condition of forests is often estimated with data collected by national forest inventory (NFI) programs. NFIs typically collect detailed tree and stand measurements across a statistically designed, systematic layout of field plots. Although timber assessment has traditionally been a focus of many NFIs (Scott and Grove, 2001), measurement of forest attributes relating to ecosystem functioning and health is increasing. NFI data is frequently called upon to generate continental- and global-scale information on biological diversity, ecosystem health and forest carbon pools. However, data collected by independent NFIs is seldom harmonized (i.e., in agreement) with respect to reporting units, field procedures, and estimation methods (Winter et al., 2008). Resulting discrepancies can produce large uncertainties when multiple NFIs are used to estimate attributes such as forest area and biomass change (Schoene, 2002; Cienciala et al., 2008).

As multinational NFI data represents a critical source of global information on greenhouse gases (e.g., United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol (1997)) and sustainability (e.g., Food and Agriculture Organization (FAO), Global Forest Resources Assessment (FRA, 2006)), promoting the harmonization of NFI definitions and measurement protocols will help reduce uncertainty, and facilitate the comparison of estimates across international boundaries.

Until recently, reporting efforts in North America have been hampered by the lack of systematically collected field data over much of the continent. Although the United States has been conducting a statistically based NFI since the late-1920’s (Shaw, 2006), neither Canada nor Mexico had, until recently adopted systematically implemented national programs (Canada see Gillis, 2001 and Gillis et al., 2005; Mexico see Sandoval et al., 2008). Given different histories of the three countries, efforts to harmonize terminology and field measurement protocols are only beginning to take shape. Nonetheless, as Mexican and Canadian NFI data begin to come on-line, new methods will be needed to determine the extent to which plot data from the three North American NFIs are compatible for continental scale reporting. Here our objective is to evaluate the inter-compatibility of plot data collected in borderland oak woodland forests by the U.S. Department of Agriculture’s Forest Inventory and Analysis (FIA) program and in Mexico by Comisión Nacional Forestal (CONAFOR).

Focusing on oak woodland on both sides of the Arizona/Sonora border, we evaluate two important aspects of data
compatibility through the analysis of seamless elevation and spectral geospatial data sets. First, to gauge a sense of plot measurement and location accuracy we compare the elevations recorded in each country’s NFI database with those taken from an independently derived digital elevation model (DEM). Second, we use a seasonal time series of normalized difference vegetation index (NDVI) images derived from Landsat satellite data to assess the consistency of the relationship between plot and satellite forest measurements across the border.

**METHODS**

**STUDY AREA**

Lying equidistant between the United States and Mexico, the study area is the 74,655 km² Madrean Archipelago ecoregion (Omernik level III, CEC 1997; Fig. 1). The forests in this border region of southeastern Arizona (United States) and northeastern Sonora (Mexico) contain some of the most diverse temperate forest ecosystems in the world. The mountains here straddle two major faunal realms (Neotropic/Holarctic) and two climatic zones (Subtropical/Temperate). The confluence of these zones interacts with complex mountain topography to support high levels of endemic biodiversity (Coblentz and Riitters, 2004). The forests, which primarily reside on a series of disconnected mountain ranges, are surrounded by vast “seas” of desert vegetation. These valley seas inhibit new species colonization which serves to isolate the higher elevation “island” biotic communities (Warshall, 1994). The forest composition displays an altitudinal gradient; open oak woodlands are found at lower elevations, which cede to closed canopy pine and fir dominated forests as elevation increases.

Oak woodland forests were selected for this analysis because they represent the largest area of forested land within the ecoregion. Focusing on this large forested area ensured that a sufficient number of NFI plots from each country were available for analysis and, minimized the effect of different inventory stratification procedures used by the two countries. The oak woodland forest areas were defined by a geographic information system (GIS) coverage of biotic communities assembled by the U.S. Forest Service at a scale of 1:1,000,000 (Brown and Lowe, 1982). In the United States, the oak woodland forest type covers 16 percent of the landscape and captures 36 percent of the FIA inventory plots collected in the ecoregion (Arizona plots only, New Mexico plots are not included in this analysis). In Mexico, the oak woodland forest type covers 32 percent of the landscape and captures 75 percent of the collected CONAFOR inventory plots in the ecoregion (Sonora plots only, Chihuahua plots are not included in this analysis).

**DATA**

**FIA**

FIA data are collected on a nationally consistent hexagonal sampling frame where at least one plot is randomly selected within each 6,000 acre hexagon (Bechtold and Scott, 2005). Each plot consists of four fixed-radius circular subplots, which taken together represent an area approximately 1 acre in size. Data collected on each FIA plot includes land use, tree measurements (e.g., species, height, and diameter) as well as other tree and site related forest attributes.

For this study, we queried the FIA database to obtain the annual inventory data collected in Arizona between 2001 and 2007. Using the geographic coordinate locations of the plots, a GIS overlay operation was used to identify the Arizona plots falling within the oak woodland boundary. The measured live tree data from these plots (n = 117) was then used to calculate basal area. Most of the trees sampled in this region are defined by FIA as woodland species which are measured for diameter at the root collar (DRC) near ground line. Thus, to calculate basal area we first converted DRC to diameter at breast height (DBH) using Eq. 1 (Chojnacky and Rogers, 1999),

\[
DBH = \beta_0 + \beta_1 \text{DRC} + \beta_2 \text{stm} + \beta_3 \text{Pied} + \beta_4 \text{DRC}_p + \beta_5 \text{Quga} + \beta_6 \text{DRC}_q
\]

where DBH is diameter at 1.3m above groundline, DRC is diameter at root collar, stm is 1 for trees with 1 stem at DRC or 0 otherwise, Pied is 1 for pinyon pine species and 0 otherwise, Quga is 1 for oak species and 0 otherwise, DRC is DRC for pinyon pine species, and DRC is DRC for oak species. Constants for the β terms (in inches) are \(\beta_0 = -2.6843, \beta_1 = 1.0222, \beta_2 = 0.7433, \beta_3 = 0.7469, \beta_4 = -0.0399, \beta_5 = 1.2244, \text{ and } \beta_6 = -0.0689\). Equation 1 was formulated using 224 trees sampled in western Colorado for Pinus edulis, Utah juniper (Junipers osteosperma) and Gambel oak (Quercus gambelii). Here we applied the equations at the genus level (e.g., all oak species were converted to DBH using Quga in Eq. 1), which accounted for nearly 85 percent of the measured trees in the study area. The remaining trees were converted to DBH using the closest available matching equation (e.g., deciduous species were converted using the Quga equation, conifer species using the Pied equation). Although this extrapolation involves applying the equation outside of the range and species in which it was initially developed, it currently represents the best available option for converting FIA data from DRC-to-DBH.
After converting from DRC to DBH, basal area per tree was calculated for each measured live tree $\geq 5$ inches using Eq. 2,

$$BA \ (ft^2) = 0.005454 \times DBH^2 \quad [2]$$

where $BA$ is basal area in $ft^2$ and DBH is in inches. Basal area per tree was multiplied by trees per acre (TPA) and condition proportion (COND_PROP) variables in the FIA database, and then summed across each plot to yield per plot estimates of basal area in $ft^2/ac$. The $ft^2/ac$ estimates were then multiplied by 0.2296 to get basal area in $m^2/ha$.

**CONAFOR**

CONAFOR data are also collected on a nationally consistent sampling grid which consists of more than 24,000 plots covering all vegetation types. The grid spacing of plots depends on vegetation type (e.g., 5x5 km grid for temperate and high tropical forests, 10x10 km for shrub lands and low tropical forests, and 20x20 km for arid regions) which is taken from a national land use and vegetation cover map derived from Landsat data. Similar to FIA, CONAFOR data are collected on four circular subplots which cover an area approximately 1 acre in size. Data collected include topography, land use and disturbance as well as tree species and diameter measurements among others. For more information on the enhanced Mexican national forest inventory program see Sandoval et al. (2008).

Plot data for the Mexican state of Sonora were spatially queried in a GIS system to select the plots contained within the geographic extent of the oak woodland boundary. The measured live trees $\geq 12.7$ cm DBH (or 5 inches, same minimum used for FIA) from the selected plots ($n = 142$) were used to calculate basal area per tree using Eq. 3,

$$BA \ (m^2) = 0.00007854 \times DBH^2 \quad [3]$$

where $BA$ is basal area in $m^2$ and DBH is in cm. The CONAFOR tree data is collected at DBH approximately 1.3 m above ground line, therefore no DRC conversion was necessary. The Mexican inventory data does not contain expansion factors. In order to obtain basal area on a per hectare basis, we used only the plots which contained 4 measured subplots. Because the plots have a fixed radius, this allowed the use of a constant 6.25 area expansion factor. Basal area per tree was multiplied by this constant expansion factor, then summed across each plot to yield per plot estimates of basal area in $m^2/ha$.

**SRTM DEM**

To help evaluate the location and measurement accuracy of the NFI plots (described below) we obtained digital elevation data from the Consultative Group for International Agriculture Research – Consortium for Spatial Information (CGIAR-CSI; http://srtm.cgiar.org/). Based on the unfinished 3 arc second data originally released by the National Aeronautics and Space Administration (NASA), the CGIAR-CSI version-4 data used here have been hydrologically corrected with a gap-filling algorithm to produce a smooth continuous raster surface at 90 m spatial resolution. The data were downloaded in separate 1’ x 1’ degree grid tiles, which were mosaiced together in ArcInfo Grid to produce seamless coverage of the study area. Once mosaiced, the study area elevation grid was reprojected from geographic coordinates to UTM projection with WGS 84 datum.

**DATA COMPATIBILITY TESTS**

**PLOT LOCATION**

One important indicator of data compatibility is that NFI plots are located where they are supposed to be in geographic space and that they accurately reflect the topography of the landscape. In general, if plots are properly located and measured, then we should be able to use each plot’s geographic coordinates to derive independent estimates of topographic variables (e.g., elevation from a DEM) which closely match the records found in each country NFI database. To test this idea we compare independent estimates of elevation extracted from an SRTM DEM with those found in each countries NFI database (FIA $n = 117$, CONAFOR $n = 142$). SRTM data was extracted for each NFI plot using the mean of a $3 \times 3$ window placed over plot center (for both FIA and CONAFOR we used actual plot coordinates, not publically available). It is possible that the level of agreement (based on $R^2$) of the two countries will differ because plot elevations in the CONAFOR data are taken from field measurements, whereas in FIA they are either taken from...
field measurements, DEM or topographic map. Minor agreement differences aside, if the plots are reasonably located in geographic space, and in the case of CONAFOR are accurately measured, the plots should fall on or close to the 1:1 line when viewed in a two dimensional scatter plot. This test is intended only as a check for errors which might bring into question the general reliability of the location and measurement of the NFI plots, and is not intended to be a precise quantitative assessment of elevation accuracy.

**BASAL AREA ESTIMATION**

Barring differences in precipitation and back-ground effects, it is to be expected that NDVI (a satellite based measure of green leaf area) will increase as basal area increases. If the basal area estimates derived for each country are compatible, we should find no statistical difference between their fitted relationships with NDVI. To test this hypothesis we conducted an analysis of covariance (ANCOVA). The analysis was restricted to the range of basal area measured by both countries (i.e., 16.77 m²/ha). In addition to capping the range of basal area, the Landsat images do not cover the full extent of the study area, thus the number of plots available for the ANCOVA analysis (FIA n = 74, CONAFOR n = 121) is less than was used for the plot location and measurement test described above. For the plots qualifying for the analysis, NDVI was extracted from each of the six seasonal images using the mean of a 3x3 window placed over plot center (for FIA and CONAFOR we used actual plot coordinates, not publically available). We then tested the null hypothesis that the slopes of each countries fitted lines were equal using a standard F test. If the slopes are found equal, then each countries fitted mean basal area is “adjusted” according to the overall mean of NDVI. The null hypothesis of equal adjusted means is then tested with a second F test. If we do not reject the null hypothesis of equal adjusted means (i.e., p-value >= 0.05) then we can conclude that per unit NDVI, the sample plots collected on both sides of the border have statistically similar estimates of basal area.

**RESULTS**

**PLOT LOCATION TEST**

Scatter plots comparing the SRTM elevations and elevations from the NFI data revealed good agreement for FIA as indicated by all of the plots falling along the 1:1 line (Fig. 2). Although the majority of CONAFOR plots also fell on or near the 1:1 line, we did find seven plots (indicated by dashed oval and arrow in Fig. 2) which were not; all but one of these plots had recorded elevations almost exactly 1,000 m above the SRTM measurements. Given the small percentage of plots affected (4.9 percent) and the systematic nature of these deviations, it is likely these errors were the result of data entry mistakes rather than plot location or measurement inaccuracies. Removing the seven outliers from the CONAFOR data we found that elevations from both NFI data sets were in similar agreement with the independent SRTM elevations (FIA R² = 0.99, CONAFOR R² = 0.97). Although the CONAFOR data displays higher residual variance (Figure 2), the R² results verified that both FIA and CONAFOR plots were placed on the landscape with sufficient accuracy that the topographic descriptors published in each database could be accurately reproduced using independent data.

**BASAL AREA COMPATIBILITY TEST**

The ANCOVA results revealed that the fitted lines for both countries were statistically similar for all six NDVI image dates (Table 1). While the fitted lines were not necessarily parallel (Figure 2), they were similar enough that the null hypothesis of equal slopes could not be rejected. The test of equal adjusted means revealed that for four of the six image dates the null hypothesis could not be rejected (Table 1). This indicates that once canopy conditions represented by NDVI were accounted for, the adjusted mean basal areas of the two countries were, in the majority of cases, not significantly different. Although two of the image dates (4/24/2004, 6/25/2003) produced results which were close to rejecting the null hypothesis, the small average difference in adjusted mean basal area (0.2 m²/ha) across the six image dates supports the conclusion that the basal area estimates from the two countries are similar enough to be deemed compatible.

**DISCUSSION**

In this study we evaluated the compatibility of American (FIA) and Mexican (CONAFOR) NFI data using sample plots collected across an area of borderland oak woodland forest. Given the discontinuous nature of the NFI data, the evaluation of compatibility focused on the analysis of geospatial data sets which seamlessly and consistently spanned the area of data collection. To gauge location and measurement compatibility we compared the similarity of each countries plot responses to topographic (i.e., elevation) and spectral based NDVI data.

As verification that the NFI plots were located on the landscape with sufficient spatial accuracy for joint analysis, we compared independently derived SRTM DEM data with elevations recorded in each countries NFI database. The test identified seven CONAFOR plots which had considerable deviation in recorded elevations. As these errors were systematic in nature they were most likely the result of data entry mistakes. Comparison with freely available SRTM data may in the future be an efficient quality control measure for NFI elevation data. Plots identified as erroneous can be re-inspected to verify coordinate, elevation measurement and data entry integrity. Aside from the identified outliers, we found all of the FIA and CONAFOR NFI elevations to be in good agreement with the SRTM data. This provided
evidence that the FIA and CONAFOR plots were reasonably located on the landscape and that elevation was accurately measured or estimated by each inventory program. In general, this test provided an indirect verification of plot location compatibility, as well as an effective means for identifying plots which might have potential misalignment or measurement errors.

Although the inventory programs have similar plot layout designs, differences in data collection protocols exist which must be accounted for before undertaking a more thorough assessment of data measurement compatibility. Here, efforts were taken to harmonize the calculation of basal area in order to assess the compatibility of stand density estimates derived from the two NFI data sets. Harmonization efforts included applying published equations to convert FIA DRC to DBH, using the same minimum DBH cutoff and converting basal area estimates to like units (m²/ha). Although the DRC-to-DBH equations (Chojnacky and Rogers, 1999) used here were extrapolated well beyond their geographic and ecological boundaries, the adjustment was a critical step in harmonizing the NFI data.

In this study FIA basal area was reduced by roughly 32 percent after the DRC-to-DBH conversion was applied. This is similar in magnitude to the 10 to 25 percent reduction in basal area reported by Chojnacky and Rogers (1999) for ponderosa pine forests in the Gila National Forest, New Mexico. It should be noted that basal area reported in the FIA database (e.g., variable BALIVE) is calculated for woodland species without converting DRC-to-DBH. In addition, reported diameters for woodland species in the FIA database (e.g., variable DIA) are actually DRC measurements, thus when calculating basal area, volume or biomass with equations that require DBH as input (e.g., Jenkins et al., 2003) DRC-to-DBH conversion is required. Given the considerable difference between DRC-corrected and uncorrected basal area, future work should focus on improving the necessary equations required to make this critical adjustment. These improvements would also stand to benefit future studies which use FIA data to estimate carbon and biomass for woodland species. As CONAFOR measurements are taken at DBH, the conversion of FIA DRC-to-DBH was an important step in developing harmonized estimates of basal area for the two countries.

To test the compatibility of the basal area estimates we used NDVI data from Landsat. The reasoning behind this test comes from the fact that canopy conditions of forests in this region are open and highly variable. For example, a stand with the same unit basal area could have a relatively open canopy structure consisting of a few large but scattered trees or a closed canopy structure consisting of several small but densely clustered trees. Although in this situation basal area is the same, the different canopy conditions result in very different measures of NDVI. Thus, if the relationship between basal area and NDVI were drastically different for each country, this might suggest that inventory stratification or systematic data collection differences might be affecting the compatibility of the basal area estimates. The ANCOVA analysis showed that the fitted relationships between basal area and NDVI were similar for each country, thus offering evidence that per unit NDVI the basal area estimates were in the majority of cases, statistically compatible.

Overall, both independent tests based on the analysis of seamless geospatial data indicated a high degree of plot level compatibility between American and Mexican NFI data. Given the high level of plot data compatibility we hoped to proceed with a joint analysis of the two NFI data sets with the purpose of investigating the ecological impacts of divergent management and land use practices on stand density in borderland oak woodland forests. Jointly analyzing the NFI data in this context could take two different approaches. One approach might be to use the plot data from both countries to analyze basal area distributions using tests of central tendency (e.g., looking for statistical differences in population means or medians). A second approach might collectively use the NFI plots from each country to derive statistical estimates of basal area for a particular region of interest (e.g., the Madrean archipelago ecoregion or the oak woodland forest type).

To perform these types of joint analyses requires accounting for differences in inventory stratification which exist between the two countries. FIA’s sample grid extends with equal intensity to all lands, but only forested plots (as identified though photos or in the field) are surveyed in detail. In contrast, CONAFOR pre-stratifies their sample grid according to a land use map produced by the Mexican Agency INEGI (Instituto Nacional de Estadistica, Geografia e Informatica). Detailed plot measurements are taken on all lands with forests more heavily sampled than other wooded lands and deserts. While the Mexican plot data contain stratum identifiers, thus allowing calculation of average conditions by stratum, the stratification map itself is not publically available. Without this map, it is impossible to determine weights for a complex analysis unit such as the oak woodland forest type within the Madrean archipelago ecoregion. Publication of the Mexican land cover map, or at least development of factors communicating the area represented by each plot, would greatly increase the inventory’s usefulness in local and cross-border analyses.

**CONCLUSION**

We determined that plot data from the two inventories are compatible: plots from both countries were accurately geo-located, and the relationship between measured basal area
and satellite imagery was consistent across the border. The following recommendations may be identified following our work.

ACKNOWLEDGEMENTS

The authors wish to thank Rigoberto Palafox Rivas, Carmen L.M. Tovar and Vanessa S. Mascorro of CONAFOR for their important contributions to this study.

| CONAFOR | 1. Systematically check archived plot elevations against freely available global terrain data.  
2. Attach stratum-adjusted area equivalents (expansion factors) to plot data. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FIA</td>
<td>1. Develop systematic conversion from root collar diameter to diameter at breast height to improve regional basal area variables.</td>
</tr>
</tbody>
</table>

LITERATURE CITED


Table 1 – ANCOVA results for basal area compatibility test

<table>
<thead>
<tr>
<th>Image Date</th>
<th>F</th>
<th>p-value</th>
<th>FIA</th>
<th>CONAFOR</th>
<th>Diff</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/24/2004</td>
<td>1.06</td>
<td>0.306</td>
<td>6.3</td>
<td>5.4</td>
<td>0.9</td>
<td>3.81</td>
<td>0.052</td>
</tr>
<tr>
<td>5/8/2003</td>
<td>1.74</td>
<td>0.189</td>
<td>5.5</td>
<td>5.9</td>
<td>-0.5</td>
<td>0.84</td>
<td>0.360</td>
</tr>
<tr>
<td>5/13/2005</td>
<td>0.03</td>
<td>0.856</td>
<td>6.3</td>
<td>5.4</td>
<td>1.0</td>
<td>4.51</td>
<td>0.035</td>
</tr>
<tr>
<td>6/11/2004</td>
<td>1.68</td>
<td>0.197</td>
<td>5.5</td>
<td>5.9</td>
<td>-0.3</td>
<td>0.47</td>
<td>0.493</td>
</tr>
<tr>
<td>6/25/2003</td>
<td>0.47</td>
<td>0.492</td>
<td>5.1</td>
<td>6.2</td>
<td>-1.1</td>
<td>3.56</td>
<td>0.061</td>
</tr>
<tr>
<td>7/16/2005</td>
<td>0.04</td>
<td>0.845</td>
<td>6.5</td>
<td>5.2</td>
<td>1.3</td>
<td>7.55</td>
<td>0.007</td>
</tr>
</tbody>
</table>

*significant tests in bold

Figure 1 – The Madrean Archipelago study area (red outline) showing location of oak woodland forest (yellow) and Landsat path/rows (blue outline).
Figure 2—Scatter plots of SRTM DEM elevation versus NFI database elevation for FIA (•) and CONAFOR (•).