Analyzing Landsat Time-series Data Across Adjacent Path/Rows and Across Multiple Cycles of FIA: Lessons Learned in Southern Missouri

Mark D. Nelson[1], Sean P. Healey[2], W. Keith Moser[1], Mark H. Hansen[1], Warren Cohen[3], Mark A. Hatfield[1], Nancy Thomas[4], and Jeff G. Masek[5]

ABSTRACT: *The North American Forest Dynamics (NAFD) Program is assessing disturbance and regrowth in the forests of the continent. These forest dynamics are interpreted from per-pixel estimates of forest biomass, which are produced for a time series of Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced TM Plus images. Image data are combined with sample plot data from the Forest Inventory and Analysis (FIA) program using Random Forests, a tree-based estimation method implemented here in the R statistical environment. The NAFD approach is based on a sample of image Path/Rows, resulting in most images being disjunct from and independent of other images in the sample. Increases in sample intensity and needs for assessing forest dynamics over geographic extents larger than a single image are leading to increased frequency of adjacent, overlapping images in the sample. We assessed the consistency of estimates of forest biomass and classification of forest/nonforest in southern Missouri, USA, across space and time, for adjacent images in Path 25/Row 34 and Path 24/Row 34, and for coincident images in Path 25/Row 34 acquired in 2000 and 2007. Results were consistent across space and time, implying consistency of both Landsat and FIA data, and supporting the NAFD image sample strategy and subsequent augmentation with overlapping images.*

KEYWORDS: North American Forest Dynamics, NAFD, Landsat, consistency analyses, forest inventory, FIA

[1] U.S. Forest Service Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108
[2] U.S. Forest Service Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401
[3] U.S. Forest Service Pacific Northwest Research Station, Forestry Sciences Laboratory, 3200 SW Jefferson Way, Corvallis, OR 97331
[4] Department of Geography, 2181 LeFrak Hall, University of Maryland, College Park, MD 20742
[5] Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, MD 20771

In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. 2009. 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 1 CD.

Introduction

The North American Forest Dynamics (NAFD) project is an effort to map forest disturbance and regrowth across the continent (Goward et al. 2008). The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture Forest Service recently entered into close collaboration with NAFD, which is funded through the North American Carbon Program.

For NAFD, satellite image time-series data are carefully cross-normalised via state-of-the-art radiometric and geometric image processing procedures (Canty et al. 2004, Masek et al. 2008). The normalised image data are combined with forest inventory data, and a “state model differencing” approach is used to create a spectral model of a biophysical variable, e.g., forest biomass. The model is date-independent because images are selected for near anniversary dates and are cross-normalised; thus, relationships observed between the inventory data and contemporaneous imagery are assumed to hold throughout the time series. A single model is applied to every image in the time series, and estimates of change are obtained as differences in predictions between years. Relatively low prediction errors for state model differencing have been reported in Healey et al. (2006).

NAFD was designed for producing estimates from a sample of Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Path/Rows. This approach results in many images multi-temporal image stacks being disjunct and independent from other images in the sample. However, augmentation of the sample and needs for assessing forest dynamics over geographic extents larger than a single image are leading to more and more situations where newly added Path/Rows overlap pre-existing images in the sample.

Classification consistency, determined by comparing overlapping portions of individual Landsat Path/Row scenes, can be used as an indicator of classification quality (Cihlar et al. 2003). Relative accuracy assessments, or ‘confidence overlays’ complement conventional accuracy assessments that use "ground truth" data (Guindon and Edmonds 2002). In addition to assessing classification consistency, image overlap regions have been used to characterize the accuracy of landscape metrics (Brown et al.2000) and systematic surface reflectance and leaf area index (Butson and Fernandes 2004).

The question is, how consistent are independently modelled NAFD predictions of forest biomass within overlapping portions of adjacent Path/Rows, or between years within the same Path/Row? This study attempts to answer this question for an area of interest in southern Missouri, USA. We created predictive models using NAFD image data and assessed their performance using as prediction and validation input, imagery from: 1) The same image as the original model; 2) An adjacent, overlapping Path/Row image from the same year, e.g., performance across space; and 3) The same Path/Row, but from a different year, having a similar anniversary date, e.g., performance across time.

Methods

Study Area

We evaluated relationships in three areas of interest within south-central Missouri, USA, generally encompassing the Mark Twain National Forest and surrounding forest land (Fig. 1): 1) The western, nonoverlapping portion of Landsat World Reference System (WRS2) Path 25, Row 34 (P25R34), labeled "A"; 2) The eastern, nonoverlapping portion of Path 24, Row 34 (P24R34), labeled "B"; and 3) The area of overlap between P25R34 and P24R34, labeled "O" (Fig. 2). The geographic extent of overlap in O contains image pixels from both P25R34 and P24R34, labeled “AO” and “BO”, respectively. For FIA plots AO = O = BO.

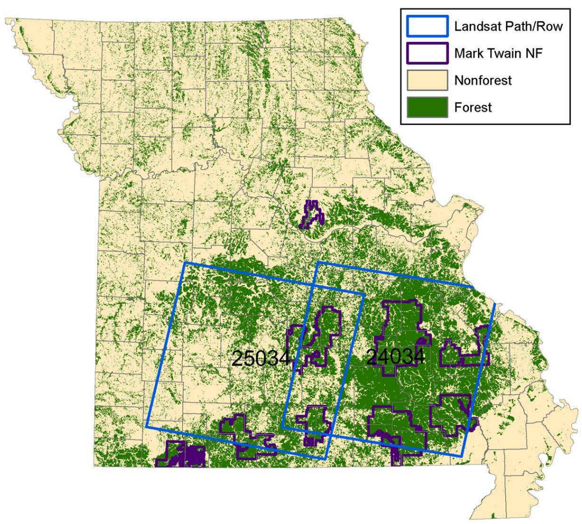

Figure 1: Landsat Path 25 / Row 34 and Path 24 / Row 34, southern Missouri, USA.

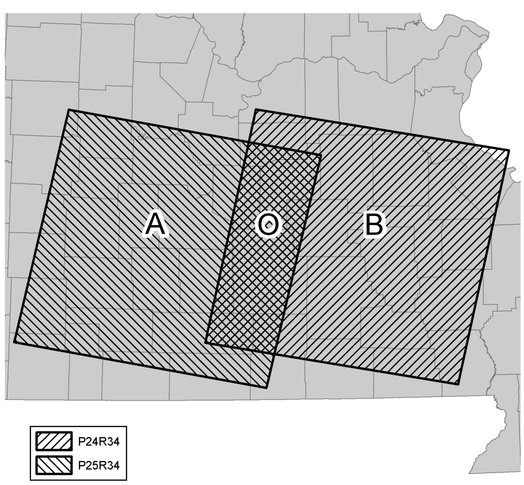

Figure 2: Study areas: A – western, nonoverlapping portion of P25R34, B – eastern, non-overlapping portion of P24R34, and O – extent of overlap between P25R34 and P24R34, southern Missouri, USA.

Data

FIA: FIA defines forest land as lands currently or formerly supporting a minimum level of tree stocking (10 percent) and not developed for a nonforest use such as agriculture, residential, or industrial use. Forest land includes commercial timberland, some pastured land with trees, forest plantations, unproductive forested land, and reserved, noncommercial forested land. FIA's definition of forest land also requires a minimum area of 1 acre and minimum continuous canopy width of 120 ft (U.S. Forest Service 2003). FIA sample plots follow a nationally consistent design comprised of four fixed-radius circular subplots, selected from a nationally consistent hexagonal sampling frame with at least one plot selected for each 6,000-acre hexagon (Bechtold and Scott 2005, Reams et al. 2005). On each FIA plot, land use (e.g., proportion forest cover), tree (e.g., species, height, and diameter) and other site variables are collected.

The FIA database was queried to obtain inventory field plot data collected between 1999 and 2007 within all Missouri counties intersecting P25R34 and P24R34. Geographic information system (GIS) data layers of inventory plot center locations were created based on global positioning system (GPS) coordinates obtained during field data collection. GPS coordinates were collected and maintained in North American Datum of 1983. The sample of FIA plots was further constrained to retain only those plots located within the geographic extent of P25R34 or P24R34. FIA plots measured during Missouri cycle 5 (1999-2003) were used for analyses across space (n = 2320), with 751 plots within A, 1278 plots within B and 291 plots within O (Figure 2). P25R34 plots measured during cycle 5 and remeasured during cycle 6 (2004-2007) were used for analyses across time (n=735). Per-plot estimates were produced for 1) Proportion forest land; and 2) Total gross biomass oven dry weight (pounds per acre) on forest land, based on trees 1.0 inch or larger, including all tops and limbs, but excluding foliage (DRYBIOT in FIA database, hereinafter: 'biomass').

Satellite Imagery: For comparisons across space, a TM image was obtained for P25R34, dated 29 August 2000; and an ETM+ image was obtained for P24R34, dated 30 August 2000 (Fig. 3). Comparisons across time were conducted within P25R34, using the 29 August 2000 TM image and a TM image from 2 September 2007. All three images were converted to surface reflectance by the NAFD Program (Goward et al. 2008) using NASA's Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2008). However, the cross-normalization procedure was omitted for this particular study. Resulting images had 28.5-m spatial resolution and UTM projection, with North

American Datum of 1983. Clouds and shadows, which covered a minimal fraction of the study area, were excluded from further analyses.

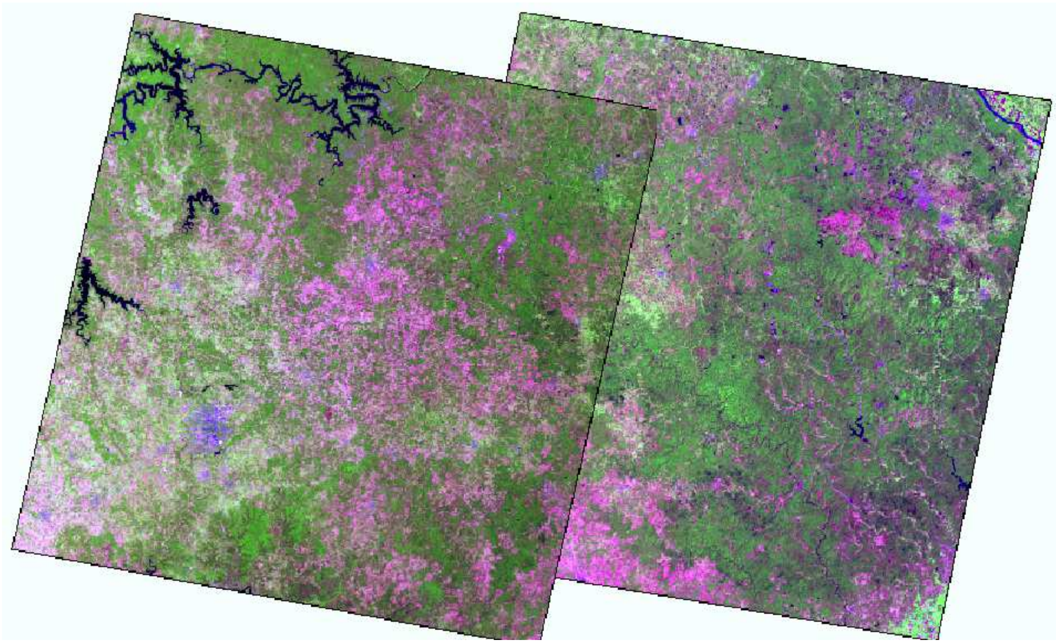

Figure 3: Landsat images of P25R34 (left) and P24R34 (right), late August 2000, southern Missouri, USA.

Modeling

NAFD protocols include methods for making spatially explicit (map-based) estimates of changes in aboveground forest biomass (Healey et al. 2007). Per-pixel predictions of biomass and forest probability were modeled using Random Forests. This work was done with R statistical software (R Development Core Team 2008) using an implementation of Random Forests developed by Liaw and Wiener (2002) and adapted by Freeman and Frescino (2008). For this approach, 2000 trees were created, with each tree using a different bootstrap sample of the data, and the best among a random subset of predictors selected for splitting each node. These trees were assembled into a 'forest' and each tree provides a 'vote' on the final, composite tree. Pixels with predictions of forest probability of 0.5 or greater were labeled forest class; all other pixels were labeled nonforest.

Across Space (within year 2000): To ensure an adequate number of validation plots in O, which is smaller in area than A or B, we chose a lower density of training plots and thus a higher density of validation plots. This ratio was reversed for A and B, where geographic extent and numbers of FIA plots were larger. Sixty percent of the FIA plots within A and 40 percent of the FIA plots within O were selected at random as model data for training P25R34. The remaining 40 percent of plots within A and 60 percent of plots within O were retained for validation analyses. A model for P25R34 was applied to: 1) Spectral data from AO (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from BO (i.e., image pixels from the adjacent image - P24R34). The same set of FIA validation plots from O were used to assess both applications of the model.

The procedure was duplicated exactly with data from scene B to create predicted maps of biomass and forest proportion in area O using imagery from both A and B, thus allowing for comparison of predictions over the same area made with 2 different imagery sources. Within O, the same model and validation plots were used for both P25R34 and P24R34 models and validation tests.

Trading Space for Time (within P25R34): For the year 2000 model, 60 percent of the FIA plots from Missouri cycle 5 (1999-2003) were selected at random for model development and the remaining 40 percent of plots (n ~ 300) were retained for model validation. Similarly, for the year 2007 model, 60 percent of the FIA plots from Missouri cycle 6 (2004-2007) were selected at random for model development and the remaining 40 percent of plots (n ~ 300) were retained for model validation.

A model based on 2000 imagery was applied to: 1) Spectral data from 2000 (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from 2007. Similarly, a model based on 2007 imagery was applied to: 1) Spectral data from 2007 (i.e., image pixels from the same image for which the model was developed); and 2) Spectral data from 2000.

Validation

Site-specific (per-pixel) validation tests were conducted to assess overall classification accuracy of forest/nonforest (f/nf) classes, and root mean square error (RMSE) and coefficient of determination (R-squared) for predictions of biomass. Results of these tests were compared to determine effects on model performance of substituting model image input across space and across time.

When comparing models across space and across time, a naming convention is employed whereby the first term refers to the image source for model development, and the second term refers to the imagery for which models are implemented. Across-space comparisons are termed A-B or B-A, and across-time comparisons are termed 2000-2007, or 2007-2000. Model validation is termed A-A, B-B, 2000-2000, or 2007-2007 when using the same image source for model development and testing.

Results

Across Space

Overall accuracy of the f/nf models ranged from 92 to 94 percent (Fig. 4). Biomass predictions had R-squared values of 0.72 to 0.74, and RMSE (percent of mean) values of 0.56 to 0.60 (Fig. 5). In some cases, substituting imagery across space resulted in slightly higher accuracies (Figs. 4 and 5).

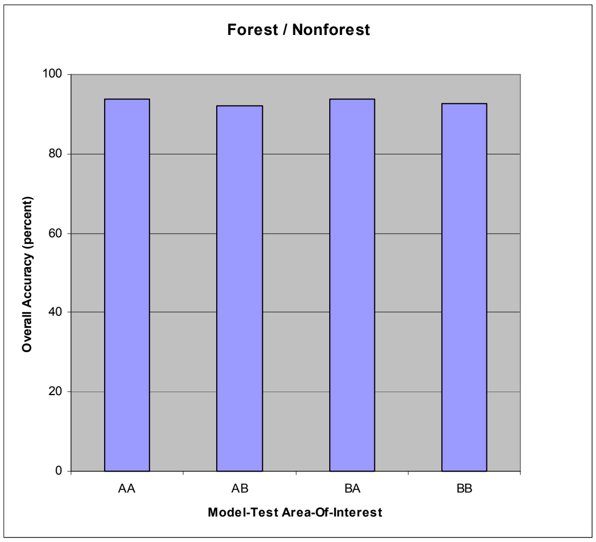

Figure 4: Overall classification accuracy of forest/nonforest classifications, 2000, southern Missouri, USA.

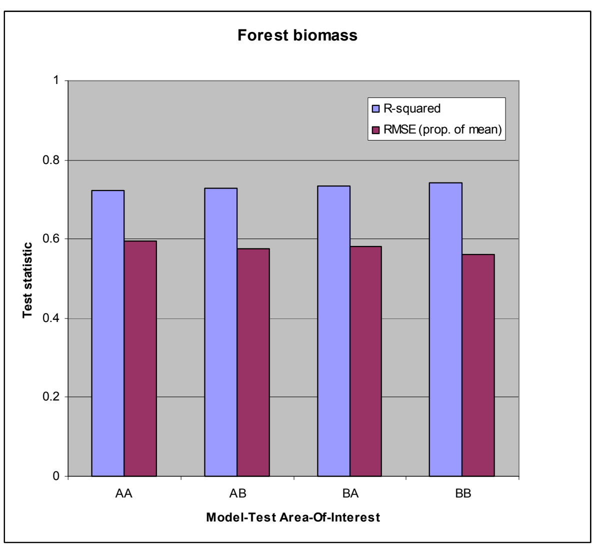

Figure 5: Error metrics for predictions of forest biomass, 2000, southern Missouri, USA.

Across Time

Overall accuracies of forest/nonforest classifications ranged from 88 to 90 percent (Figure 6). Biomass predictions had R-squared values of 0.74 to 0.79, and RMSE (percent of mean) values of 0.64 to 0.72 (Figure 7). Differences between the models built in 2000 and 2007 were negligible when it came to predicting the 2000 and 2007 test data. In some cases, the "off year" gave fractionally better predictions than models produced in the same year (Figs. 6 and 7).

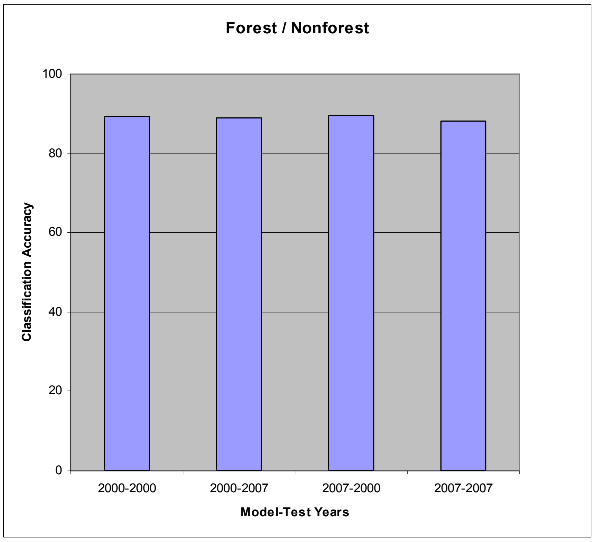

Figure 6: Overall classification accuracy of forest/nonforest classifications, P25R34, southern Missouri, USA.

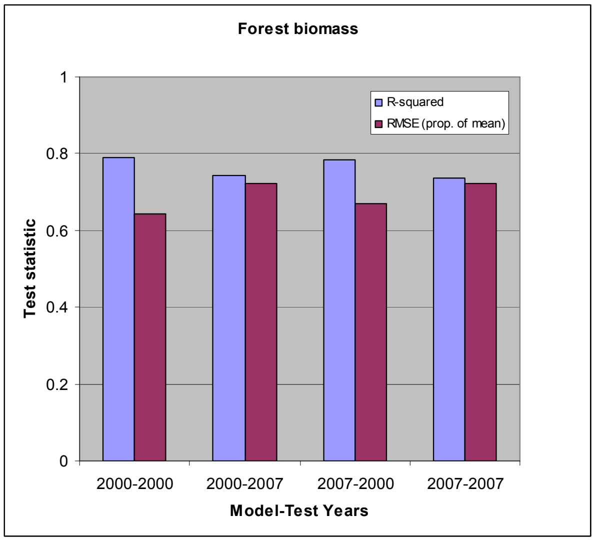

Figure 7: Error metrics for predictions of forest biomass, P25R34, southern Missouri, USA.

Discussion

Using NAFD procedures, forest/nonforest classifications and biomass predictions exhibited minimal effects across space (i.e., WRS Path/Row [Path 25 vs. Path 24]), or across time (i.e., FIA inventory cycle [cycle 5 vs. cycle 6]). While satellite sensor effects were not tested explicitly, results from Landsat 5 TM and Landsat 7 ETM+ images were consistent with each other. Consistency of results implies a consistency of both Landsat and FIA data across space and time.

The images used in this study had almost exact anniversary dates (within 2 days). Thus, sun angle effects, which usually introduce a major source of error, were minimized. Phenology, which can vary from year to year, was assumed to be similar for image anniversary dates from 2000 and 2007. Image cross-normalization, which typically is conducted during LEDAPS processing, was not performed for this study. The minimal differences across space and time observed in this study suggest that cross-normalization may be unnecessary under some circumstances.

At least a portion of the small differences observed in this study may be explainable as mostly random, since the Random Forests modeling process makes a certain number of trees (2000, in this case) from random bootstrap samples and random subsets of predictors.

A companion paper in these proceedings (Moser et al. 2008) discusses utility of NAFD products for assessing increasing eastern redcedar forest area in southern Missouri between 1985 and 2007. Assumptions of across-time consistency in Moser et al. (2008) are supported by this study.

References

Bechtold, W.A.; Scott, C.T. 2005. The forest inventory and analysis plot design. In: Patterson, Paul L., ed. The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture Forest Service: 27-42.

Brown, D.G.; Duh, J.; Drzyzga, S.A. 2000. Estimating error in an analysis of forest fragmentation change using North American Landscape Characterization (NALC) data. Remote Sensing of Environment. 71: 106-117.

Butson, C. R.; Fernandes, R. A. 2004. A consistency analysis of surface reflectance and leaf area index retrieval from overlapping clear-sky Landsat ETM+ imagery. Remote Sensing of Environment. 89: 369-380.

Canty, M.J.; Nielsen, A.A..; Schmidt, M. 2004. Automatic radiometric normalization of multitemporal satellite imagery. Remote Sensing of Environment. 91: 441-451.

Cihlar, J.; Latifovic, R.; Beaubien, J.; Trishchenko, A.; Chen, J.; Fedosejevs, G. 2003. National scale forest information extraction from coarse resolution satellite data, part 1. In: Wulder, M.A.; Franklin, S.E., eds. Remote sensing of forest environments: concepts and case studies. Boston: Kluwer: 337-357.

Freeman, E.; Frescino, T.A. 2008. Modelmap: creates random forest and stochastic gradient boosting models, and applies them to GIS '.img' files to build detailed prediction maps. Unpublished package (version 1.1) for R statistical software. Ogden, Utah.

Goward, S.N.; Masek, J. G.; Cohen, W. B.; Moisen, G. G.; Collatz, G. J.; Healey, S.; Houghton, R.A.; Huang, C.; Kennedy, R.; Law, B.; Powell, S.; Turner, D.; Wulder, M. A. 2008. Forest disturbance and North American carbon flux. Eos, Transactions American Geophysical Union. 89: 105-116.

Guindon, B.; Edmonds, C.M. 2002. Large-area land-cover mapping through scene-based classification compositing. Photogrammetric Engineering & Remote Sensing. 68: 589-596.

Healey, S.P.; Yang, Z.; Cohen, W.B.; Pierce, D.J. 2006. Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. Remote Sensing of Environment. 101: 115-126.

Healey, S.P.; Moisen, G.; Masek, J.; Cohen, W.; Goward, S.; Powell, S.; Nelson, M.; Jacobs, D.; Lister, A.; Kennedy, R.; Shaw, J. 2007. Measurement of forest disturbance and re-growth with Landsat and FIA data: anticipated benefits from Forest Inventory and Analysis' collaboration with the National Aeronautics and Space Administration and university partners. In: McRoberts, R.E.; Reams, G. A.; Van Deusen, P.C.; McWilliams, W. H., eds., Proceedings of the seventh annual forest inventory and analysis symposium; 2005 October 3-6; Portland, ME. Gen. Tech. Report WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 171-178.

Liaw, A.; Wiener, M.. 2002. Classification and regression by randomForest. R News. 2/3: 18-22.

Masek, J.G.;Huang, C.; Wolfe, R.; Cohen, W.; Hall, F.; Kutler, J.; Nelson, P.. 2008. North American forest disturbance mapped from a decadal Landsat record. Remote Sensing of Environment. 112: 2914-2926.

Moser, W.K. ; Nelson, M.D.; Healey, S.P.; Hansen, M.H.; Cohen, W.B. 2008. Estimation of biomass growth of eastern redcedar in southern Missouri: trends in afforestation. In: McWilliams, W.; Moisen, G.; Czaplewski, R., eds. Forest inventory and analysis symposium; 2008 October 21-23; Park City, UT. Proc. RMRS-P-56CD, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

R Development Core Team. 2008. R: A language and environment for statistical computing. Computer program, URL http://www.R-project.org. Vienna, Austria.

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. 2005. The forest inventory and analysis sampling frame. In: Patterson, Paul L., ed.

The enhanced Forest Inventory and Analysis Program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture Forest Service: 11-26.

U.S. Forest Service. 2003. Forest inventory and analysis national core field guide, volume 1: field data collection procedures for phase 2 plots, version 2.0. Washington, DC: U.S. Department of Agriculture, Forest Service. 281 p.