FUSION OF MULTISOURCE DATA TO DERIVE A HIGH RESOLUTION FOREST INVENTORY IN GEORGIA

by

ROGER CHARLES LOWE III

(Under the direction of Chris J. Cieszewski)

ABSTRACT

Timber inventories are designed to give a description of the forest at

- a determined spatial scale, for
- a specific area, with
- a certain level of (un)certainty in mind.

Furthermore, while their intended uses may differ, the underlying reason all inventories are made is to collect information. How that information is used, on the other hand, is often much more complicated. It is a statistical reality that data from different sources with differing statistical and sampling characteristics cannot be pooled together for the purpose of deriving a new unbiased estimator. This means, although there is abundant information about our environment (i.e. atmospheric conditions, soil composition, nearly 40 years of satellite imagery, and a wealth of site specific studies sampling for various data) one cannot assimilate these data and produce a new, unbiased estimate for the variable and area of interest. To address this issue, I present three case studies relating to the use of seemingly unrelated and incompatible data for the derivation and application of a high-resolution inventory for the detailed analysis of fiber supply and policy analysis within spatially explicit, stand level areas.

In the first study, I fill the data gaps present in the SLC-off Landsat 7 Enhanced Thematic Mapper Plus satellite imagery using the nearest neighbor methods applied to multi-temporal Landsat 5 Thematic Mapper data. The second is an application of modeling forest variables across a series of Landsat imagery for the small-area assessment of streamside management zones and road beautifying buffers. The third study describes the development of a high-resolution forest inventory for the state of Georgia from data that is traditionally not used jointly for predictions. The final inventory retains the statistical integrity of the large-area USDA Forest Inventory and Analysis while maintaining the local accuracy of the small-area timber inventories from our industry partners.

INDEX WORDS: Georgia, forest inventory, FIA, Landsat, mean balancing, SMZ

FUSION OF MULTISOURCE DATA TO DERIVE A HIGH RESOLUTION FOREST INVENTORY IN GEORGIA

by

ROGER CHARLES LOWE III

BS, The University of Georgia, 1996

MS, The University of Georgia, 2002

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

FUSION OF MULTISOURCE DATA TO DERIVE A HIGH RESOLUTION FOREST INVENTORY IN GEORGIA

by

ROGER CHARLES LOWE III

Approved:

Major Professor: Chris J. Cieszewski

Committee: Bruce Borders
Marguerite Madden
Nathan Nibbelink
Kim Iles

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

To the most wonderful wife and daughters I could have ever asked for.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee chair, Dr. Chris Cieszewski. I could not have completed this dissertation without his guidance and encouragement throughout. I would like to thank my committee members, Dr. Bruce Borders, Dr. Marguerite Madden, Dr. Nate Nibbelink, and Dr. Kim Iles, for their support. In addition, I would like to express my appreciation to our industrial partners and the folks at the Southern Research Station for their collaboration.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Forest-related industries are an important contributor to Georgia's economy. In 2011, the forest products sector contributed almost $15.1 billion dollars in revenue to the state (Georgia Forestry Commission 2008). That represented a 4% increase over the previous year and is noteworthy considering the lingering economic slow-down. The number of jobs supported by the forest products sector increased 6.8% over the same period. In 2011, it supported over 46,000 jobs. Additionally, developing markets within the state are expected to have a major impact. It is anticipated that the developing bioenergy industry will grow to a point where it contributes an approximated $27.7 billion to the economy (Georgia Forestry Commission 2008).

For a state that depends so much on its natural resources for their continued economic competitiveness both locally and abroad, the ability to frequently assess the land base in a timely manner and provide reliable and accurate information about our forests, the forested ecosystems and surrounding areas is paramount. This is especially important where forests tend to be fast growing and changing, highly fragmented in area and ownership, and there are multiple demands for their use, such as those in Georgia. Competition for the use of our forests due to population growth and migration, development, and climate change, to name a few, are projected to increase (RPA2010). These issues are compounded since the stressors, their magnitude, focus, and their effect on the current and future ecosystems will vary depending on geography. Thus, our responses to them will require regional and local strategies to address these management issues.

Detailed information about our natural resources is needed before we can begin to address these complex issues in a well-informed and fact-based manner. The United States Department of Agriculture (USDA) Forest Service Forest Inventory and Analysis (FIA) program collects forest information and produces regular reports on the condition of forests on the bases of continuous

inventory throughout the country. The inventory provides reliable, unbiased estimates suitable for reporting across large areas (Johnson et al. 2003; Williams and Schreuder 1995). The FIA data is used in Georgia in various large area inventory based analysis ranging from carbon studies to tree mortality analysis (Van Deusen 2010; Meng and Cieszewski 2006). However, since the FIA was not designed to record local knowledge about the landscape that is required for mapping smaller areas, there is a compelling need for a higher-resolution forest inventory for small area predictions and mapping; although, admittedly there have been numerous efforts of using the FIA data with satellite imagery (Blackard et al. 2008; Walker et al. 2007; Wang, Lu, and Haithcoat 2007; Sivanpillai et al. 2005; Wayman et al. 2000; Chojnacky 1998). A more suitable source for this information are the local agencies familiar with these locales whose intimate knowledge is needed for their management. Forest product industry and other large area forest owners typically maintain their own private inventories that are more detail oriented and suitable for small area, stand level, forest management. Some of this kind of information may be even available for public use or for indirect use by research institutions for enhancing the broad large area inventories providing a suitable methodology for an unbiased compilation of multi-source data information.

The FIA program provides annual state inventories based on a grid of approximately 10,000 sample plots in the intensity of about 6,000 acres per plot (McRoberts, Holden, Nelson, Liknes, & Gormanson, 2005). It is responsible for establishing and maintaining a current inventory of our Nation's more than 747 million acres of timberland. The inventory includes all public and private forestlands, wilderness areas, National Parks, defense installations, and National Forests in all 50 states, territories, and possessions of the U.S. Additionally, they are required to analyze and report on the status and trends of our forests. Each year, they are required to produce reports for each state consisting of a core set of tables addressing the status of our resources. Every five years, they are to produce a complete analytical report for each state containing the

- current status of the forest for the last 5 years;
- trends in forest status and condition over the preceding twenty years;

- timber product and output;
- an analysis of the probable forces causing the observed conditions, and
- a projection of the likely trends in key resource attributes over the next twenty years.

The inventory process has three phases. A combination of aerial photography and remotely sensed imagery is used to determine the proportion of forested and non-forest area and to stratify it as such in phase 1. Forest measurement plots are installed in phase 2. The plots are comprised of four one twenty-fourth acre subplots and are located and measured regardless of current or future intended use, ownership, or management policy. A core set of tree-related information regarding tree size, species, and overall condition are measured on each plot. Additional plot specific information like forest type, the GPS location, and an assessment of previous land uses are logged too. Phase 3 plots are installed on a subset of the phase 2 plots. They are surveyed for information about the forest ecosystem, function, condition, and health. Ozone bio-indicators, lichen community samples, soil measurements, crown condition classes, down woody debris and fuel measurements, and diversity and structure measurements, among others, are made in this phase.

The yearly and five-year reports created for their governmental and industrial stakeholders and the public at large contain detailed estimates of volume, biomass, area of forested land, and volume and area of timber removals. Summaries are tabled and graphed by species, ownership, political boundaries like county and state. When possible and where appropriate, assessments of ecological and environmental issues are also identified. The final two products they generate are an updated database of the raw and summarized data and maps showing the distribution of various forest attributes.

Though the FIA program it is chiefly designed for large area assessments (Woudenberg et al. 2010), scientists have been developing various inventory related estimates using these data. Some examples are mapping tree species (Zhang et al. 2009), estimating the amount of forest biomass on the national (Blackard et al. 2008) and regional (Chopping et al. 2011; Zhang et al. 2007) scales, and assessing carbon and fire consumption (e.g., Chen et al. 2011). Research on enhancing the

uses of the FIA data with the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 5 Thematic Mapper (TM) is even more frequent than using MODIS due to their higher spatial resolution and ease of data acquisition. Examples of national scale studies are discussed in Chen et al. (2011), Finley et al. (2008), and He et al. (2011), while examples of regional studies in this area can be found in He et al. (1998), Kellndorfer et al. (2010), Lowe et al. (2009), Sivanpillai et al. (2005), Zhang et al. (2009), and Zheng et al. (2007, 2008).

For effective forest management and planning, it is necessary to make fiber supply assessments within areas smaller than what the FIA inventory was inherently designed. Therefore, there is a critical need for research into technologies that facilitate reliable county, sub-county, and even stand-level resource assessments based on multi-source data that complement the FIA information. The use of remote sensing technology is an obvious resource that can be leveraged with the FIA information to create a more comprehensive data resource. Other data sources that may be used to further enhance the use of this data include water resources related and transportation data (e.g., Lowe et al. 2009), elevation models, land cover GAP data.

Much of the interest in satellite remote sensing as it relates to forest inventories lies with the potential of reducing costs and increasing the consistency, precision and accuracy, and timeliness of the data collected and reported. Lashbrook et al. (2001) realized in their inventory of white pine in eastern Ohio, that a TM-based inventory required less labor and time than traditional inventories while yielding estimates with standard errors substantially below those of existing ones. Researchers from the USDA Forest Service's North Central FIA region have estimated that ground-based sample sizes would have to be increased by at least a factor of 5 to achieve the same level of precision as obtained with their TM-based forest/nonforest post-stratification (McRoberts et al. 2002). Others have shown it has a resolution appropriate for vegetation mapping (Evans 1994) and is a data source from which acceptable estimates over large areas are possible (M. Trotter et al. 1997; Czaplewski 1998). The concept of combining the FIA data with satellite imagery is becoming ever more relevant and much new research has been conducted in the recent years on error estimation and statistically efficient processing of these combined information technologies

(Finley et al. 2008; Healey et al. 2011; Magnussen et al. 2009; McRoberts 2010b; McRoberts et al. 2011; Wang et al. 2011).

Many studies have related remotely sensed data to biophysical variables using traditional supervised or unsupervised classification (Dymond, Mladenoff et al. 2002; Liu, Takamura et al. 2002; Lu, Mausel et al. 2004), nonparametric nearest neighbor (Reese, Nilsson et al. 2002; Makela and Pekkarinen 2004; Gu, Dai et al. 2006), and regression approaches (Hyyppa, Hyyppa et al. 2000; Salvador 2000; Healey, Yang et al. 2006; Sivanpillai, Smith et al. 2006). Classification accuracies at or above 80% for general schemes such as ”forested” and ”non-forested” are common (Rack 2000, Salajanu 2001, De Colstoun 2003; Haapanen 2004; Sivanpillai 2006). However, it is apparent that as the species component becomes more complex, their accuracies tend to decrease quickly (Luther 2006, Salajanu 2001, Reese 2003). Estimates of forest biomass follow a similar trend in which estimates at the pixel level tend to have a higher error rate than those at aggregate levels (Franco-Lopez 2001, Katila 2001, Makela 2004, Reese 2002, Trotter 1997, Tomppo 2004). Abundant research relating satellite image spectral response and forest age or seral stage has been carried out in the Pacific Northwest and Canada (Horler and Ahern 1986; Cohen, Spies et al. 1995; Jakubauskas 1996; Pax-Lenney, Woodcock et al. 2001; Song, Woodcock et al. 2002; Wilson and Sader 2002; Franklin, Hall et al. 2003; Wulder, Skakun et al. 2004; Zhang, Pavlic et al. 2004; Hall, Skakun et al. 2006; Song, Schroeder et al. 2007) and in the tropical regions of the world (Nilson and Peterson 1994; Kimes, Nelson et al. 1998; Nelson, Kimes et al. 2000; Sader, Hayes et al. 2001; Vieira, de Almeida et al. 2003) where forests are long-lived. Among the common findings from the studies are:

1. a decreasing Landsat spectral response value as the forest ages,
2. the relationship between spectral response and age stronger in younger stands,
3. spectral response saturates at the time of crown closure, after which single-image age determination is unlikely,
4. TM bands 5 and 7 are critical in locating clearcuts, and

5. spectral response in young stands is affected by the surrounding ground cover while canopy shadowing is a factor after crown closure.

A handful of studies with similar results have been carried out in the southeastern part of the U.S. (Coleman, Gudapati et al. 1990; Sivanpillai, Smith et al. 2006). Both single- and multi-date image analysis techniques have been implemented to assess the changing landscape. Single-date analyses rely on the spectral differences among various features on the landscape at one point in time (i.e. young & old stands, or different cover types within a scene). As was pointed out by several authors (Nilson and Peterson 1994; Olsson 1994; Song, Woodcock et al. 2002; Song, Schroeder et al. 2007), these relationships are not necessarily linear and may vary greatly. Nelson and Kimes (Nelson, Kimes et al. 2000), after analyzing successional stages of forests in Brazil, went as far as stating that "information concerning clearing history (of the land) cannot reliably be deduced using single-date TM imagery" and was seconded by others (Song, Woodcock et al. 2002). Though there have been related reports of moderate successes (Wulder, Skakun et al. 2004), the results are often specific to the data used in the analysis and not transferable to other sites (Coleman, Gudapati et al. 1990).

Multi-date image analysis of the forested landscape relies on the comparison of response values as they change throughout each individual image and across time (as seen on different images). Wilson and Sader (Wilson and Sader 2002) in their analysis of clearcuts in British Columbia found 79% to 96% of the significant changes in land cover in their study area. They also found an inverse relationship between their classification accuracy and the time period between imagery. Similar to single-date analyses, sensitivity in reflectance to canopy closure across the multi-temporal dataset has been noted (Song, Woodcock et al. 2002; Song, Schroeder et al. 2007).

Lu et.al (2004) conducted a comprehensive review of the many change detection techniques. They listed several factors that may affect the results of a multi-date analysis. Included among them is the calibration or normalization between multi-temporal images and the use of anniversary or very near anniversary acquisition dates. In many operational instances though, fully addressing these issues is not possible. Atmospheric conditions, especially during the summer months in the

South, may preclude the acquisition of cloud free anniversary-date or even same season imagery. There are, for instance, only 5 out of 15 available path 17, row 38 TM images classified as having 0% cloud cover between January 1, 2007 and September 10, 2007. Six of the images have 50% or more.

Overall, there have been great many studies using the FIA data with satellite imagery for the estimation of various natural resource characteristics both in the data analysis domain to obtain estimates and in research into methodologies and evaluations of their outcomes. However, there has not been much research into mixing disparate forest inventory data. Such different inventories are typically not compatible with each other and combining those results in biased estimates. The idea of combining inventory information from various incompatible sources without a resulting bias (Cieszewski et al. 2003, Iles 2009) is a novel and unorthodox idea. It is a nontrivial challenge, but it offers unprecedented potential in increasing the information detail beyond earlier capabilities, because it explores formerly untapped resources and possibilities. The total-balancing concept is enabling inventory compilations that are based on multi-source data with incompatible variances.

The purpose of the research presented here was to demonstrate multi-source data fusion based on the FIA data, ETM+ and TM satellite imagery, and various other forest inventory related information to derive an enhanced high-resolution forest inventory for Georgia. Such information would be suitable for forest management uses in sub-county, stand-level, spatially explicit forest management and related forest resource analysis and planning. As examples of this type of data and their uses, I present here case studies relating to the derivation and applications of a pixel-level inventory for the detailed analysis of fiber supply and policy analysis within spatially explicit stand level areas.

I describe three examples of studies, each progressing in complexity, for improving the FIA inventory data and the level of detail they describe for different locations in the state of Georgia. The first study, presented in Chapter 2 of this dissertation, addresses a generic application problem due to missing data. A faulty SLC sensor on the ETM+ spacecraft results in missing data for approximately 20% of its area. However, due to the spatial pattern of these faulty pixels, only

a small portion of an image's 8.5 million acres is suitable for evaluation across the continuous landscape. The study addresses the issue using K-nearest neighbor (KNN) methods applied to multi-temporal satellite imagery.

The second study, presented in Chapter 3, describes research on a high-resolution delineation problem. The study demonstrates the process of modeling forest variables across multiple segmented TM satellite images for small area estimates. This example describes the derivation of high-resolution inventories throughout the state for the detailed analysis of streamside management zones (SMZ) and road beautifying buffers (RBB) that are used in sustainability analyses. The estimation includes an assessment of both the areas and the inventory characteristics of small areas corresponding to two sizes of buffers. The problem was solved using an appraisal of inventory characteristics based on the FIA inventory data, satellite imagery data, and other information necessary for reliable determination of the streams and roads, such as the stream data, transportation data, and elevation data.

Finally, the third study, presented in Chapter 4, demonstrates the fusion of satellite imagery and Forest Service's inventory to generate a broad area, high resolution, spatially explicit, and balanced inventory of the state. Additionally, I demonstrate how estimates of forest volume can be further refined with the inclusion of traditional industrial cruise information. Chapter 4 describes the derivation of an enhanced high-resolution forest inventory using a multi-stage approach of data fusion based on:

1. derivation of a provisional high-resolution forest inventory based on the KNN methodology that uses the FIA data and TM images, and application of the mean-balancing methodology to remove the KNN produced bias as compared to the FIA unbiased inventory mean;

2. derivation of improved training data through a separate application of the mean balancing methodology to our collaborator's industry inventory data for further propagation with the KNN methodology;

3. use of data derived in step 2 to re-apply the KNN methodology to the entire state (scene by scene), and a double-weighted application of the mean-balancing methodology to the entire state (scene by scene) to remove the KNN produced bias as compared to the FIA inventory.

1.1 REFERENCES

[1] Blackard, J.A., M.V. Finco, E.H. Helmer, G.R. Holden, M.L. Hoppus, D.M. Jacobs, A.J. Lister, G.G. Moisen, M.D. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D.L. Weyermann, K.C. Winterberger, T.J. Brandeis, R.L. Czaplewski, R.E. McRoberts, P.L. Patterson, and R.P. Tymcio. 2008. Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sens Environ 112(4):1658-1677.

[2] Chen, X., S. Liu, Z. Zhu, J. Vogelmann, Z. Li, and D. Ohlen. 2011. Estimating aboveground forest biomass carbon and fire consumption in the U.S. Utah High Plateaus using data from the Forest Inventory and Analysis Program, Landsat, and LANDFIRE. Ecological Indicators 11(1):140-148.

[3] Chojnacky, DC. 1998. "Double Sampling for Stratification: a Forest Inventory Application in the Interior West." USDA Forest Service Rocky Mountain Rorest and Range Experiment Station Research Paper (RP-7): 1.

[4] Chopping, M., C.B. Schaaf, F. Zhao, Z. Wang, A.W. Nolin, G.G. Moisen, J.V. Martonchik, and M. Bull. 2011. Forest structure and aboveground biomass in the southwestern United States from MODIS and MISR. Remote Sens Environ 115(11):2943-2953.

[5] Cieszewski, C. J., K. Iles, R. C. Lowe, M. Zasada (2005). "Proof of Concept for an Approach to a Finer Resolution Inventory." In: McRoberts, R. E.; Reams, G. A.; Van Deusen, P. C.; McWilliams, W. H., eds. 2003 Proceedings of the Fifth Annual Forest Inventory and Analysis Symposium. 2003 November 18-20; New Orleans, LA. Gen. Tech. Rep. W)-69. Washington, DC: U.S. Department of Agriculture Forest Service. 222p.

[6] Cohen, W. B., T. A. Spies, et al. (1995). *Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A.* International Journal of Remote Sensing 16(4): 721 - 746.

[7] Coleman, T. L., L. Gudapati, et al. (1990). *Monitoring forest plantations using Landsat Thematic Mapper data.* Remote Sensing of Environment 33(3): 211-221.

[8] Czaplewski, R (1998). Use Of Advanced Remote Sensing Technologies for Annual State Inventories. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, USA.

[9] De Colstoun, E.C.B; M. Story; C. Thompson; K. Commisso; T.G. Smith, J.R. Irons (2003). National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. Remote Sensing of Environment. 85: 316-327.

[10] Dymond, C. C., D. J. Mladenoff, et al. (2002). *Phenological differences in Tasseled Cap indices improve deciduous forest classification.* Remote Sensing of Environment 80(3): 460-472.

[11] Evans, David (1994). Forest Cover from Landsat Thematic Mapper Data for use in the Catahoula Ranger District Geographic Information System. General Technical Report so-99. Southern Forest Experiment Station, USDA Forest Service. 14pp.

[12] FIA Business Report (http://www.fia.fs.fed.us/library/bus-org-documents/docs/FY2012FINALDRAFTBusinessReport_030113-opt.pdf)

[13] Finley, A.O., S. Banerjee, and R.E. McRoberts. 2008. A Bayesian approach to multi-source forest area estimation. Environmental and Ecological Statistics 15(2):241-258.

[14] Franco-Lopez, H., A. Ek., M.E. Bauer (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment. 77:251-274.

[15] Franklin, S. E., R. J. Hall, et al. (2003). *Discrimination of conifer height, age and crown closure classes using Landsat-5 TM imagery in the Canadian Northwest Territories.* International Journal of Remote Sensing 24(9): 1823 - 1834.

[16] Georgia Forestry Commission. 2008. Sustainable Forest Management in Georgia. Macon, GA: Georgia Forestry Commission.

[17] Gu, H. Y., L. M. Dai, et al. (2006). *Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data.* Science in China Series E-Technological Sciences 49: 54-62.

[18] Haapanen, R; A. Ek.; M. Bauer; A. Finley (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment. 89: 265-271.

[19] Hall, R. J., R. S. Skakun, et al. (2006). *Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume.* Forest Ecology and Management 225(1-3): 378-390.

[20] He, H.S., D.J. Mladenoff, V.C. Radeloff, and T.R. Crow. 1998. Integration of gis data and classified satellite imagery for regional forest assessment. Ecological Applications 8(4):1072-1083.

[21] He, L., J.M. Chen, S. Zhang, G. Gomez, Y. Pan, K. McCullough, R. Birdsey, and J.G. Masek. 2011. Normalized algorithm for mapping and dating forest disturbances and regrowth for the United States. International Journal of Applied Earth Observation and Geoinformation 13(2):236-245.

[22] Healey, S. P., Z. Q. Yang, et al. (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(1): 115-126.

[23] Healey, S.P., E. Lapoint, G.G. Moisen, and S.L. Powell. 2011. Maintaining the confidentiality of plot locations by exploiting the low sensitivity of forest structure models to different spectral extraction kernels. International Journal of Remote Sensing 32(1):287-297.

[24] Horler, D. N. H. and F. J. Ahern (1986). *Forestry information content of Thematic Mapper data.* International Journal of Remote Sensing 7(3): 405 - 428.

[25] Hyyppa, J., H. Hyyppa, et al. (2000). *Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes.* Forest Ecology and Management 128(1-2): 109-120.

[26] Jakubauskas, M. E. (1996). *Thematic Mapper characterization of lodgepole pine seral stages in Yellowstone National Park, USA*. Remote Sensing of Environment 56(2): 118-132.

[27] Johnson, Devin S, Michael S Williams, Raymond L Czaplewski, Johnson D.S., Williams M.S., and Czaplewski R.L. 2003. "Comparison of Estimators for Rolling Samples Using Forest Inventory and Analysis Data." Forest Science 49 (1) (February): 5063. http://www.ingentaconnect.com/content/saf/fs/2003/00000049/00000001/art00004.

[28] Katila, M.; E. Tomppo (2001). Selecting estimation parameters for the Finnish multisource National Forest Inventory. Remote Sensing of Environment. 76:16-32.

[29] Kellndorfer, J.M., W.S. Walker, E. LaPoint, K. Kirsch, J. Bishop, and G. Fiske. 2010. Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM plus , and ancillary data sets. Journal of Geophysical Research-Biogeosciences 115.

[30] Kimes, D. S., R. F. Nelson, et al. (1998). *Accuracies in Mapping Secondary Tropical Forest Age from Sequential Satellite Imagery.* Remote Sensing of Environment 65(1): 112-120.

[31] Lashbrook, J., R. Williams, J. Nichols, R. Heiligmann, B. Motsch, R. Romig, J. Vimmerstedt (2001). The Application of Remote Sensing to the Inventory of White Pine (Pinus Strobus L.) in Eastern Ohio. Ohio Agricultural Research

[32] Liu, Q. J., T. Takamura, et al. (2002). *Mapping of boreal vegetation of a temperate mountain in China by multitemporal Landsat TM imagery.* International Journal of Remote Sensing 23(17): 3385-3405.

[33] Lowe, R.C., C.J. Cieszewski, S. Liu, Q. Meng, J.P. Siry, M. Zasada, and J. Zawadzki. 2009. Assessment of Stream Management Zones and Road Beautifying Buffers in Georgia Based on Remote Sensing and Various Ground Inventory Data. Southern Journal of Applied Forestry 33(2):91-100.

[34] Lu, D., P. Mausel, et al. (2004). *Change detection techniques.* International Journal of Remote Sensing 25(12): 2365-2407.

[35] Luther, J.E.; R.A. Fournier; D.E. Piercey; L. Guindon; R.J Hall (2006). Biomass mapping using forest type and structure derived from Landsat TM imagery. International Journal of Applied Earth Observation and Geoinformatics. 8:173-187.

[36] Magnussen, S., R.E. McRoberts, and E.O. Tomppo. 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sens Environ 113(3):476-488.

[37] Makela, H. and A. Pekkarinen (2004). *Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data.* Forest Ecology and Management 196(2-3): 245-255.

[38] McRoberts, R. E., Bechtold, W. A., Patterson, P. L., Scott, C. T., & Reams, G. A. (2005). The Enhanced Forest Inventory and Analysis Program of the USDA Forest Service?: Historical Perspective and Announcement of Statistical Documentation. Health (San Francisco).

[39] McRoberts, R., D. Wendt, M. Nelson, M. Hansen (2002). Using a land cover

[40] classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of the Environment. 81: 3644.

[41] McRoberts, R.E. 2010a. The effects of rectification and Global Positioning System errors on satellite image-based estimates of forest area. Remote Sens Environ 114(8):1710-1717.

[42] McRoberts, R.E. 2010b. Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. Remote Sens Environ 114(5):1017-1025.

[43] McRoberts, R.E. 2011. Satellite image-based maps: Scientific inference or pretty pictures? Remote Sens Environ 115(2):715-724.

[44] McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management 272:3-12.

[45] McRoberts, R.E., E.O. Tomppo, and E. Naesset. 2010. Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research 25(4):368-381.

[46] McRoberts, R.E., S. Magnussen, E.O. Tomppo, and G. Chirici. 2011. Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. Remote Sens Environ 115(12):3165-3174.

[47] Meng, Qingmin, and Chris J Cieszewski. 2006. Spatial Clusters and Variability Analysis of Tree Mortality Physical Geography 27(6): 534553.

[48] Nelson, R. F., D. S. Kimes, et al. (2000). *Secondary Forest Age and Tropical Forest Biomass Estimation Using Thematic Mapper Imagery.* BioScience 50(5): 419-431.

[49] Nilson, T. and U. Peterson (1994). *Age dependence of forest reflectance: Analysis of main driving factors.* Remote Sensing of Environment 48(3): 319-331. NOAA. National Climatic Data Center. http://www.ncdc.noaa.gov/oa/ncdc.html

[50] Olsson, H. (1994). *Changes in satellite-measured reflectances caused by thinning cuttings in Boreal forest.* Remote Sensing of Environment 50(3): 221-230.

[51] Pax-Lenney, M., C. E. Woodcock, et al. (2001). *Forest mapping with a generalized classifier and Landsat TM data.* Remote Sensing of Environment 77: 241-250.

[52] Rack, J. (2000). Forest/nonforest classification of Landsat TM data for annual inventory phase one stratification. Paper presented at the Second Annual Forest Inventory and Anaylsis (FIA) Symposium, Salt Lake City, UT, October 1718, 2000.

[53] Reese, H., M. Nilsson, et al. (2002). *Applications using estimates of forest parameters derived from satellite and forest inventory data.* Computers and Electronics in Agriculture 37(1-3): 37-55.

[54] Reese, H.; M. Nilsson; T.G. Phalen; O. Hagner; S. Joyce; U. Tingelof; M. Egberth; H. Olsson (2003). Countrywide estimates of forest variables using satellite data and field data from the national forest inventory. AMBIO. 32(8): 542-548.

[55] Sader, S. A., D. J. Hayes, et al. (2001). *Forest change monitoring of a remote biosphere reserve.* International Journal of Remote Sensing 22: 1937-1950.

[56] Salajanu, D.; C.E. Olson Jr. (2001). The Significance of Spatial Resolution Identifying Forest Cover from Satellite Data. Journal of Forestry. 99(6):32-38.

[57] Salvador, R. (2000). *An assessment of the spatial variability of basal area in a terrain covered by Mediterranean woodlands.* Agriculture Ecosystems & Environment 81(1): 17-28.

[58] Sivanpillai, R., C. T. Smith, et al. (2006). *Estimation of managed loblolly pine stand age and density with Landsat ETM+ data.* Forest Ecology and Management 223(1-3): 247-254.

[59] Sivanpillai, R., C.T. Smith, R. Srinivasan, M.G. Messina, and X. Ben Wu. 2005. Estimating regional forest cover in East Texas using enhanced thematic mapper (ETM plus) data. Forest Ecology and Management 218(1-3):342-352.

[60] Song, C., C. E. Woodcock, et al. (2001). "Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?" Remote Sensing of Environment 75(2): 230-244.

[61] Song, C., C. E. Woodcock, et al. (2002). *The spectral/temporal manifestation of forest succession in optical imagery: The potential of multitemporal imagery.* Remote Sensing of Environment 82(2-3): 285-302.

[62] Song, C., T. A. Schroeder, et al. (2007). *Predicting temperate conifer forest successional stage distributions with multitemporal Landsat Thematic Mapper imagery.* Remote Sensing of Environment 106(2): 228-237.

[63] Tomppo, E.; M. Halme, (2004). Using coarse scale forest variables as ancillar information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote sensing of Environment. 92:1-20.

[64] Trotter, C. M., J. R. Dymond, C. J. Goulding (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. International Journal of Remote Sensing. 18[10]:2209-2223.

[65] Vieira, I. C. G., A. S. de Almeida, et al. (2003). *Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia.* Remote Sensing of Environment 87(4): 470-481.

[66] Van Deusen, P. 2010. "Carbon Sequestration Potential of Forest Land: Management for Products and Bioenergy Versus Preservation." Biomass and Bioenergy 34 (12) (December): 16871694. doi:10.1016/j.biombioe.2010.03.007.

[67] Wang, G., M. Zhang, G.Z. Gertner, T. Oyana, R.E. McRoberts, and H. Ge. 2011. Uncertainties of mapping aboveground forest carbon due to plot locations using national forest inventory plot and remotely sensed data. Scandinavian Journal of Forest Research 26(4):360-373.

[68] Wang, Cuizhen, Zhenqian Lu, and Timothy L. Haithcoat. 2007. "Using Landsat Images to Detect Oak Decline in the Mark Twain National Forest, Ozark Highlands." Forest Ecology and Management 240 (1-3) (March): 7078. doi:10.1016/j.foreco.2006.12.007.

[69] Walker, Wayne S., Josef M. Kellndorfer, Elizabeth LaPoint, Michael Hoppus, and James Westfall. 2007. "An Empirical InSAR-optical Fusion Approach to Mapping Vegetation Canopy Height." Remote Sensing of Environment 109 (4) (August): 482499. doi:10.1016/j.rse.2007.02.001.

[70] Wayman, Jared P, Randolph H Wynne, Richard G Oderwald, James B Campbell, and John A Scrivani. 2000. "Landsat TM-Based Forest Area Estimation Using Iterative Guided Spectral Class Rejection."

[71] Williams, Michael S, and HT Schreuder. 1995. *Documentation and evaluation of growth and other estimators for the fully mapped design used by FIA - a simulation study.* Forest Science 41 (3): 2645.

[72] Wilson, E. H. and S. A. Sader (2002). *Detection of forest harvest type using multiple dates of Landsat TM imagery.* Remote Sensing of Environment 80(3): 385-396.

[73] Woudenberg, Sharon W.; Conkling, Barbara L.; O'Connell, Barbara M.; LaPoint, Elizabeth B.; Turner, Jeffery A.; Waddell, Karen L. (2010). "The Forest Inventory and Analysis Database: Database description and users manual version 4.0 for Phase 2." U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Fort Collins, CO. RMRS-GTR- 245.

[74] Wulder, A. A., R. S. Skakun, et al. (2004). *Estimating time since forest harvest using segmented Landsat ETM+ imagery.* Remote Sensing of Environment 93(1-2): 179-187.

[75] Zhang, Q., G. Pavlic, et al. (2004). *Deriving stand age distribution in boreal forests using SPOT VEGETATION and NOAA AVHRR imagery.* Remote Sensing of Environment 91(3-4): 405-418.

[76] Zhang, Y., H.S. He, W.D. Dijak, J. Yang, S.R. Shifley, and B.J. Palik. 2009. Integration of Satellite Imagery and Forest Inventory in Mapping Dominant and Associated Species at a Regional Scale. Environ Manage 44(2):312-323.

[77] Zheng, D., L.S. Heath, and M.J. Ducey. 2007. Forest biomass estimated from MODIS and FIA data in the Lake States: MN, WI and MI, USA. Forestry 80(3):265-278.

[78] Zheng, D., L.S. Heath, and M.J. Ducey. 2008. Spatial distribution of forest aboveground biomass estimated from remote sensing and forest inventory data in New England, USA. Journal of Applied Remote Sensing 2.

CHAPTER 2

KNN BASED REPAIR OF ENHANCED THEMATIC MAPPER PLUS IMAGES AFFECTED BY THE SCAN LINE CORRECTOR FAILURE

2.1 INTRODUCTION

The Landsat series of satellites has been continuously recording and archiving images for nearly four decades, compiling the most complete set of image data about the Earth's surface. This vast archive enables us to evaluate the landscape not only in terms of its current surroundings, but in relation to its previous state as well. Landsat images play an important role in research and development of natural resources and related industries. The data has been used markedly for various productions of forest inventories and natural resource studies. Some national inventories (Gillis 2001; Gjertsen 2007; Reese et al. 2010; Tomppo et al. 2008) and much inventory work in the USA (McRoberts and Tomppo 2007; R. E. McRoberts et al. 2009; R. McRoberts, Tomppo, and Naesset 2010) are based on mapping the ground inventories onto Landsat imagery. Some have attempted to assess impacts of regulatory measures, such as streamside management zones and road buffers, on merchantable timber (Lowe et al. 2009). Studies such as these are especially important for a state like Georgia whose forests support a wide range of uses. The ability to address these types of issues in a quick and efficient manner is significant.

Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+), the satellites on which many of these studies are based, have operated well past their intended operational life spans. In May 2003, the ETM+ satellite suffered a failure of the Scan Line Corrector (SLC) that results in rows of missing pixels near the edges of an image. The data gaps (i.e., missing pixels) are approximately 450 meters (15 pixels) high (Fig. 2.1C) at the edges and diminish toward the center of the scene (Fig. 2.1D). A north/south swath in the center of the scene, approximately 20% (Fig. 2.1A, Fig. 2.1B) of the image's width, remains completely unaffected.

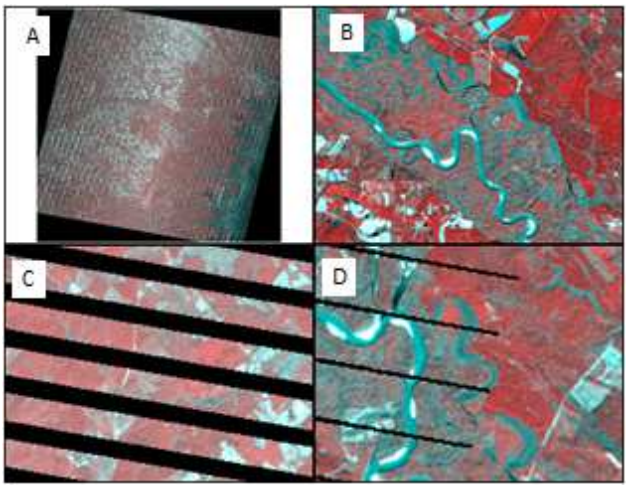

Figure 2.1: Landsat 7 Enhanced Thematic Mapper Plus satellite image affected by the SLC failure. A) Path 17, row 38 ETM+ base image; B) riparian area in the unaffected center portion of the image; C) missing pixels near the edge of the image (15 pixels high); D) missing pixels terminating near the center of the scene (2 pixels high).

Shortly after the malfunction, the US Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS) conducted a 'scientific usability and validity' study of the SLC-off imagery. They concluded that the overall radiometric and geometric quality of the recorded pixels remain unaffected (USGS 2003), though the missing pixels do degrade the overall image usability and without a correction only a small portion of the scene is suitable for detailed local land mapping.

In an attempt to fill in the missing pixels in SLC-off ETM+ imagery, the USGS EROS Data Center (EDC) developed a multi-date histogram matching process to fill the gaps based on a spatially or temporally adjacent image or images. The pixel-wise iterative process matches unaffected pixels within a 17-pixel window on the SLC-off image with spatially coincident pixels from an adjacent image(s). The adjacent image could be an ETM+ scene captured before the failure, a TM image, or even a SLC-off ETM+ image. The least squares solution is calculated and applied to the affected ETM+ image to yield a new gap-filled pixel. This method reportedly works well in and across large homogeneous areas, although artifacts may be introduced in regions with cloud and

cloud shadows, along narrow features and boundaries, and in areas that have undergone a cover type or phenological change (Howard, Lacasse, and Howard 2004). Caution is required when using an adjacent SLC-off image in the process since scan line gaps from spatially or temporally adjacent scenes do not overlap.

Maxwell et al. (2004, 2007) developed a segment-based process that utilizes landscape patch boundaries, which restrict pixel interpolation limiting it to remain within consistent cover types. In this approach, an adjacent TM or SLC-on ETM+ image is segmented into spectrally homogeneous groups. Values for the affected pixels are then interpolated using only pixels from within the segment in which it falls. This approach yields data well suited for landscape and regional scale studies, and crop-specific monitoring, because the crop boundaries typically do not change from season to season. However, caution is required these images for evaluations at the pixel-level and within areas with dynamic landscapes (Bedard et al. 2008, Maxwell et al. 2004, 2007).

The recent Neighborhood Similar Pixel Interpolator approach (J. Chen et al. 2011) and the Geostatistical Neighborhood Similar Pixel Interpolator approach (Zhu, Liu, and Chen 2012) to SLC-off ETM+ image repair, and by Jin et al. (2013) to recover clouded and cloud shadow areas build on the approach reported by Meng et al. (2009) and Lowe et al. (2009). The key to these approaches is the process used to select the nearest pixel(s), or in the case of Jin et al. (2013), the means by which similar pixels are selected for model development. The location of an affected image is identified on an unaffected, temporally adjacent TM or ETM+ image. The positions of the spectrally nearest pixels are determined on that companion image and then pinpointed on the original SLC-off image. The new value, then, is determined from spectral information from the original scene at those location(s) determined on the companion scene. Using this approach, they report no boundary shifts or striping in the repaired areas and explain that the spectral and spatial integrity of linear features are maintained since the new pixel values are determined from value combinations drawn from the ETM+ data itself. Unlike the most similar pixel approach reported by Meng et al. (2009), Chen et al.(2011) recommends using 20 nearest neighbors.

OBJECTIVES

The purpose of this study was to explore a gap-filling process that could be used to repair any suspect pixels, due to the SLC hardware failure, atmospheric conditions, or any other environmental disruptions, and it would allow for an arbitrary selection of reference scenes subject to availability of viable spectral definitions concurring with the suspect pixels. The method needed to retain the Landsat 7 Enhanced Thematic Mapper Plus spectral information across the landscape, and be consistent regardless of physical or phenological changes that appear in the imagery.

2.2 DATA

Our data consisted of ETM and TM satellite image scenes from path 17 and row 38 for the years of 2005 and 2006. Each complete ETM+ scene contains information on about 8.5 million acres with a resolution of 900 m^2 with the pixels defined by six spectral and one thermal band. Figure 2.2 shows small samples from path 17 row 38 that were analyzed in for the purpose of reconstructing complete image information from multiple ETM+ satellite passes. These images contain the black lines where pixels are affected due to the SLC failure.

While the image in Figure 2.2 may suggest that the missing pixels can be repaired from one adjacent image, typically it may take four to five ETM+ images to construct full-undamaged information for the entire area. The TM satellite has the same resolution and frame as the ETM+ satellite, and both TM and ETM+ produce spectrally similar scenes and encompass the same area at the same resolution. However, the TM satellite was still delivering reliable spectral information for all areas, with the usual exception of intermittent environmental interferences, such as those due to clouds, shadows, and smoke (Figs. 2.3A and 2.3B). Thus, we used the TM images for repairing the ETM+ images.

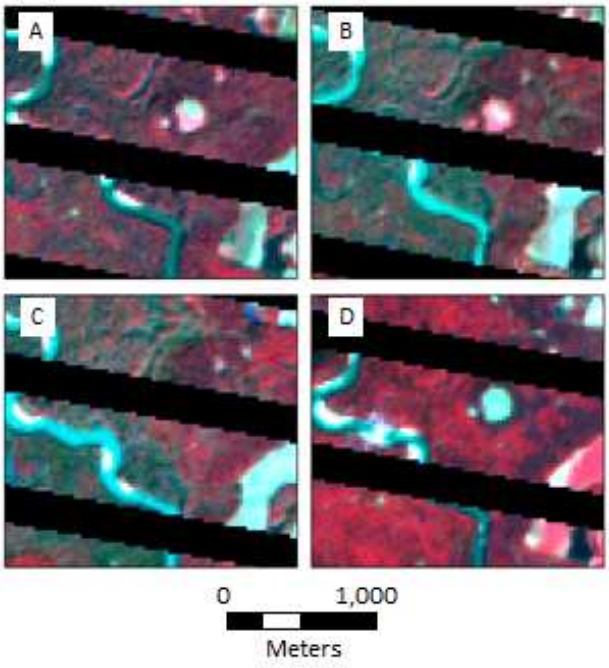

Figure 2.2: Non-overlapping dropped lines due to the SLC failure on the Landsat 7 Enhanced Thematic Mapper Plus satellite. A river flowing from the northwest to the southeast, a circular non-vegetated area near the center, and a pasture on the eastern portion of the image can be seen. Images were captured in A) November 2003, B) December 2004, C) December 2005, and D) September 2006. In this area of the scene, the dropped regions are approximately 270 meters high with approximately 700 meters between.

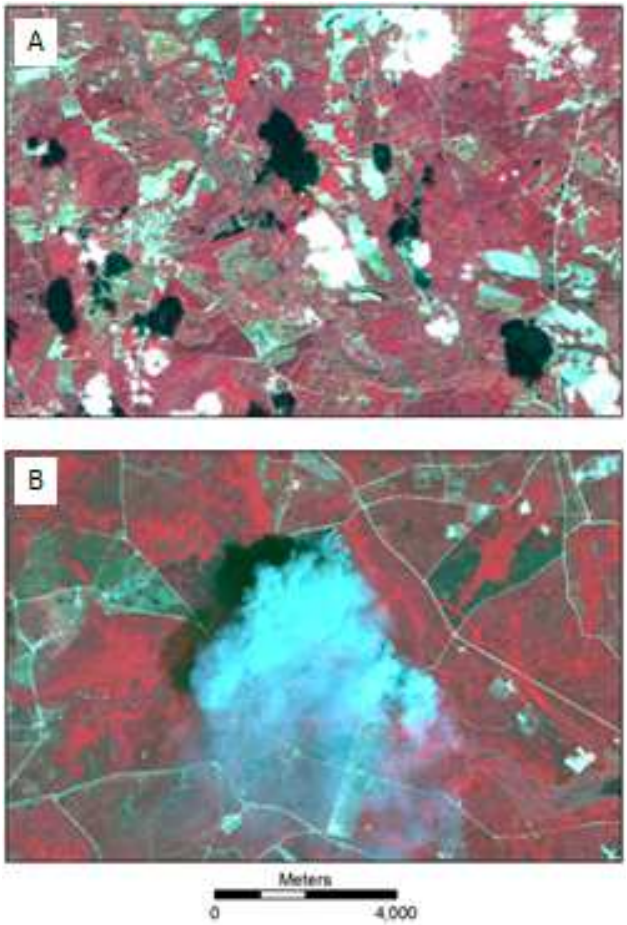

Figure 2.3: Areas affected by A) clouds and cloud shadows and B) smoke as seen on Landsat 5 Thematic Mapper images.

2.2.1 Study area

The study boundary for this project corresponds to the intersection of the path 17, row 38 ETM+ and TM images listed in Table 2.1 (ULx: -83.1834; ULy: 32.6456; LRx: -80.9306; LRy: 30.8399; rows: 6,741 px; cols: 7048 px). The site, encompassing 427,595 ha, is located in the coastal plain region of Georgia, approximately 320 kilometers southeast of Atlanta. The Altamaha River is located approximately in the middle of the scene and flows east to the Atlantic Ocean. Elevations range from sea level to 132 meters (Fig. 2.4). This area includes some of the most productive timberland sites in the region, and it is one of the most commercially important forest areas of the state. Seventy-two percent of it is forested, supporting 2.14 million hectares of forestland (Cieszewski 2007). Over 2.8 million of these acres are conifer species, a majority of which is loblolly pine with clear evidence of artificial regeneration. The remaining forestland is either of the mixed hardwood/pine type (521,000 acres; approximately 1/3 showing some evidence of artificial regeneration), or various hardwood or cypress species (1.9 million acres; 3% exhibiting some evidence of artificial regeneration.

Table 2.1: The path 17, row 38 ETM+ and TM images used in the gapfilling process.

			Days Since Capture	
EDC File Name	Sensor	Acq. Date	Summer ETM+	Winter ETM+
LE70170382005207EDC00	ETM+	26-Jul-05	0	0
LE70170382005335EDC00	ETM+	Dec. 01, 2005	0	0
LT50170382005167EDC00	TM	16-Jun-05	-40	-168
LT50170382005327EDC00	TM	Nov. 23, 2005	120	-8
LT50170382006122EDC00	TM	2-May-06	280	N/A
LT50170382006330EDC00	TM	Nov. 26, 2006	N/A	360

2.2.2 Satellite image data

I selected two virtually cloud-free SLC-off Landsat 7 Enhanced Thematic Mapper Plus (ETM+) images to test the gap filling process and downloaded from the USGS Global Visualization Viewer (http://glovis.usgs.gov). The summer scene was acquired on July 26, 2005 and the winter scene on December 01, 2005 (Tab. 2.1). Approximately 23%, 2.661 million ha, of each scene is affected

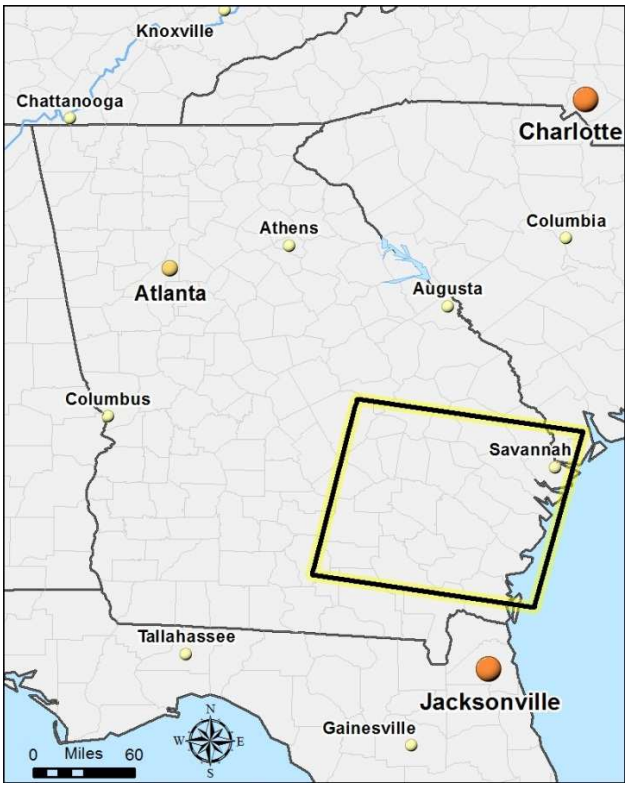

Figure 2.4: WRS2 path 17, row 38 study site located on the southeastern coast of Georgia.

by the SLC anomaly. The unaffected north to south center swath is approximately 16 km wide encompassing 280,245 ha.

The series of Landsat 5 Thematic Mapper images used in the gap filling process (Tab. 2.1) were paired and grouped into 3 categories:

1. near-date
2. opposite-season, and
3. near-anniversary.

The highest quality TM scene acquired as near to the date of each ETM+ scene were chosen as the near-date scenes. The summer near-date TM scene was acquired 40 days before and the winter near-date 8 days before their respective ETM+ counterparts. These TM scenes were also used to

form the summer and winter opposite-season image pairs. The summer opposite-season TM image was captured 120 days after the summer ETM+ image, and the TM winter opposite-season image was captured 168 days prior to its winter ETM+ counterpart. The near-anniversary TM scenes (i.e., the highest quality scenes acquired in the same season 1 year before or after the ETM+ capture date) were acquired 280 and 360 days after the Summer and winter ETM+ images, respectively. The winter scene was acquired in November of 2006, and the summer in May of 2006.

Sixteen well-distributed ground control points were located on each image for verification of image-to-image registration. We computed the root-mean square error (RMSE) using the summer ETM+ as the base image. No RMSE exceeded 12.5 meters. A visual comparison of the images revealed no egregious misalignment in the USGS terrain corrected L1T imagery, and we concluded that the co-registration was satisfactory. Enhanced Thematic Mapper Plus and TM bands 1 - 5 and 7 were layer stacked using the ERDAS Imagine software using the nearest neighbor resampling technique and a 30-meter pixel resolution. We did not use the thermal bands.

We masked out the pixels located in any image's collar (the non-data containing region around an image's edge) and those in an areas affected by the SLC failure. Additionally, the pixels within 3,000 meters of the image's bounding box and those within 1,500 meters of the actual data's edge were masked to minimize any edge effect. The total unmasked area is 1,842,279 ha (20,469,772 pixels).

2.2.3 Other GIS data

We used the Georgia Land Use Trends 2005 (NARSAL, 2001) and the National Land Cover Data 2001 (NLCD)(EPA, 2000) to identify homogeneous areas for testing the accuracy of the analysis. The GLUT dataset is a 13-class land cover classification based on satellite imagery from multiple seasons acquired primarily in 2005. It includes homogeneous land cover classes such as urban, crop/pasture, evergreen, deciduous, mixed, and clearcut/sparse forest, various wetlands, water, and other non-vegetated classes. Georgia overlays portions of the nationwide NLCD 2001 land cover dataset that was developed using ETM+ imagery from 1999 to 2003. It includes 16 cover types

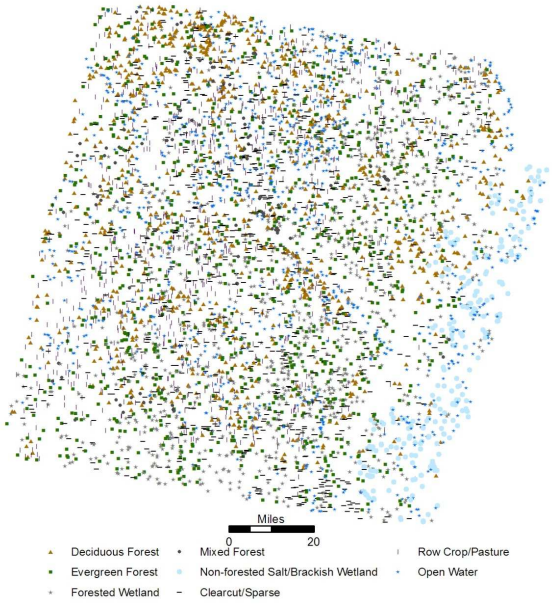

Figure 2.5: Small sample site locations, 0.81 to 1.98 ha., generated from the combination of the GLUT and NLCD land cover classification data sets that were used in the evaluation of the repair process.

that are similar to those of GLUT's, but due to the differing dates of the source imagery and other confounding factors, they are not identical. I overlaid the two land cover data sets and retained all cells that matched and eliminated any cells that did not. Clumps of adjacent pixels between 9 and 22 pixels were vectorized, yielding a polygon dataset containing areas that were classified as the same type by both agencies and are between 0.81 and 1.98 ha, 2 to 5 acres, in size.

The remaining 43,316 polygons represent the following land cover classes: 1) clearcut/sparse, 2) deciduous forest, 3) evergreen forest, 4) forested wetland, 5) mixed forest, 6) non-forest salt/brackish wetland, 7) open water, and 8) row crop/pasture. We reduced the data further by randomly selecting, when available, 1000 samples from each class to be set aside for image classification and validation. We retained all polygons in the mixed forest (112), and non-forested salt/brackish wetland (246), classes (Fig. 2.5) (Tab. 2.2).

Table 2.2: Combined GLUT/NLCD land cover samples used in classification and validation.

Cover Type	# Polygons	Mean Area (ha)	Standard Dev. Area (ha)	Total Area (ha)
Clearcut/sparse	1,000	1.14	0.31	1,137
Deciduous Forest	1,000	1.08	0.27	1,080
Evergreen Forest	1,000	1.25	0.35	1,249
Forested Wetland	1,000	1.15	0.32	1,148
Mixed Forest	112	1.01	0.2	113
Non-forested, Brackish Wetland	246	1.28	0.36	314
Open Water	1,000	1.19	0.33	1,193
Row Crop/Pasture	1,000	1.27	0.35	1,269

2.3 METHODS

2.3.1 K-nearest neighbor gap-filling

I applied a K-nearest neighbor (KNN) gap-filling method developed internally within the Fiber Supply Assessment group, WSFNR, UGA, (Cieszewski and Lowe 2008). The method was originally used to repopulate areas of an image affected by various disturbances, such the SLC failure, or clouds and smoke, and their shadows. Lowe and Cieszewski (2009) subsequently discussed the use of the method in SLC failure repair, while Meng et al. (2009a) discussed the use of the method in cloud removal. In this study, I applied this method to all of the pixels on the scene, as if they all needed repairing. I used the areas of the scene with the unaffected pixels, the areas that did not need the repair, to estimate the accuracy of the pixel repair.

The first step of the method is the pairing of spatially concurrent pixels from two different images – an affected pixel from the ETM+ image (pixels i) and its unaffected counterpart from the TM image (pixels j). The second step is to find the locations of the K nearest neighbors on the TM image based on pixel j. The nearest neighbor is defined using the minimum Euclidean Spectral Distance (ESD) (Eq. 2.1):

$$ESD = \sqrt{\sum_{i=1}^{n} (j_L - k_L)^2}, \tag{2.1}$$

where: *ESD* is the Euclidean Spectral Distance between TM pixel *j* and pixel *k*; *L* denotes the satellite band; and, *n* is the number of bands in the image. The average of the ETM+ pixel values at those K nearest locations, determined on the TM image, are stored as the new value for ETM+ pixel *i*.

I enforced three spatial constraints during the nearest neighbor selection. I masked the current pixels, pixel *i* and its counterpart on the TM image, pixel *j*, from the searches for the nearest neighbors. This ensured no pixel would be used in its own estimation. Second, I masked all affected pixels from the search for the near neighbors to ensure only valid data was used in the pixel repair process. Lastly, I limited the nearest neighbor search window to 750 meters (25 pixels) extending in the cardinal directions centered on pixel *i*. This distance corresponds to the maximum distance of spatial autocorrelation between Landsat-derived variables and pine basal area determined in previous studies (Meng et al. 2009).

2.3.2 Derived data layers and summary preparation

To test different scenarios of TM scene selections and numbers of K-neighbors, I ran the KNN gap-fill repair process separately for the summer and winter ETM+ images using K values 1 to 20. A *"near-date"* pairing refers to the matching of an ETM+ image in need of repair with a TM image captured as close to the same time as possible. An *"opposite-season"* pairing refers to the matching of an ETM+ image with a TM image that was captured one season before or season after. An *"anniversary-season"* pairing refers to matching an image for repair with an image acquired in the same season a year before or a year after.

I used only the unaffected pixels for the evaluation of the pixel repair, and summarized the residuals for each KNN-processed image at two scales, scene-wide and for small areas (polygons). The scene-wide summaries included all the image pixels that contain valid spectral information (do not need repair). I used a combination of ArcMap's Reclassify and SetNull tools to mask all the

pixels with faulty spectral information, such as the pixels affected by the SLC failure. I assigned null values to all the bad pixels and left all the good pixels unchanged. The null designation ensured that only cells containing valid spectral information would be included in subsequent processes and summaries. There are 27,007,535 pixels included in the assessment of the summer ETM+ repair and 26,999,061 pixels in the winter ETM+ process.

To simulate a small-area assessment, I performed a supervised classification on each ETM+ and all KNN-processed images based on a training set derived from the combined land cover described in Section 2.3. I randomly selected approximately half of each land cover type. One-half served for the development of spectral signatures and the image classification and the other half for the validation of the image classification (Tab. 2.3).

Table 2.3: Training and validation sites used for supervised classification and testing.

	Training		Validation	
Forest Type	# Groups	Total Acres	# Groups	Total Acres
Clearcut/Sparse	511	1,438	489	1,371
Deciduous Forest	508	1,354	492	1,313
Evergreen Forest	526	1,606	474	1,479
Forested Wetland	515	1,459	485	1,377
Mixed Forest	61	157	51	122
Non-forested, Brackish Wetland	134	417	112	359
Open Water	460	1,371	540	1,577
Row Crop/Pasture	484	1,519	516	1,614

I made the initial scene-wide examination of the repaired data in the 8-bit spectral radiance (DN) domain for bands 1-5 and 7. The DNs range from 0 to 255 where a value of 1 is recorded by the satellite over an area of minimum radiance. A value of 0 denotes NoData pixels. Conversely, pixels with a value of 255 were captured over an area with a maximum or near maximum radiance. Thus, the range of residuals for these data is -255 to 255. A low residual represents an area that is dark on the unrepaired ETM+ image such as a shadow or deep water that is estimated to be a bright feature, possibly a cloud or sand. Conversely, a positive residual will occur in a bright area that is estimated as a dark feature. A pixel whose residual equals zero was perfectly repaired. To

evaluate the pixel repair success in the DN domain, I summarized the residuals by determining the following:

- the percentage of pixels with a residual equal to 0, the
- residual representing the 95th percentile and its corresponding K, and the
- root-mean squared error.

The normalized difference vegetation index (NDVI) (Eq. 2.2) is a widely used index that has been related to the biophysical features on the ground like leaf area index (Curran, Dungan, and Gholz 1992; Turner et al. 1999; M. Chen and Cihlar 2000), land cover type ((Xiao et al. 2002; Vieira 2003), and others forest related activities ((Sader, Waide, and Lawrence 1989; Wilson and Sader 2002; Maselli and Chiesi 2006). To examine the effects of the repair process on imagery when used to map vegetation over large areas, I transformed both the winter and summer ETM+ and all KNN-processed images into NDVI. I assessed the differences by calculating the mean absolute difference (AD) (Eq. 2.3) for all season pairings and across K = 1, 2, , 20.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}, \tag{2.2}$$

where: *NDVI* is the normalized difference vegetation index, *NIR* is the spectral response from the near-infrared band (TM and ETM+ band 4), and *RED* is the spectral response from the red band (TM and ETM+ band 3).

$$AD = |NDVI_u - NDVI_k| \tag{2.3}$$

where: *AD* is the absolute difference, *NDVIu* is the NDVI value generated from the raw ETM+ imagery, and *NDVIk* is the NDVI value generated from the KNN-processed imagery. Image classification is a common method resource managers use to assess the landscape. To replicate that type of analysis, I applied a supervised Maximum Likelihood classification to the NDVI imagery derived from the:

- raw ETM+ image from the summer (NDVIs),

- raw ETM+ image from the winter (NDVIw),

and the K=1, 2, , 20, KNN-processed imagery from the ETM+/TM:

- summer/near pairing (NDVIsn),
- summer/opposite pairing (NDVIso),
- summer/anniversary pairing (NDVIsa),
- winter/near pairing (NDVIwn),
- winter/opposite pairing (NDVIwo), and the
- winter/anniversary pairing (NDVIwa).

I used ArcGIS' *Create Signatures* tool to generate signatures for the entries in the combined land cover training data set. Then I classified the images into discrete land cover categories using the *Maximum Likelihood Classification* tool. I used an error matrix to express the classification accuracies described in Story and Congalton (1986). The matrix consists of a series of columns that represent an entry's observed truth and rows that represent its assigned class (Tab. 2.4). Each cell contains the number of items observed as the column's truth yet classified as the row's class. A summation of the cells where the observed (column) truths and assigned classes (row) match discloses the number of correctly classified samples. A summation of all entries in a row reflects the number of samples classified as that class and the column summation reveals the number of samples observed as that class.

I assessed the differences in classifications using the overall classification accuracy (OA),

$$OA = \frac{\sum_{m=1}^{n} i_{mn}}{\sum_{m=1}^{n} i_{..}} * 100, \tag{2.4}$$

where: *OA* is the overall accuracy, *i* is the number of pixels assigned to class *m* and observed as class *n*; and Cohen's Kappa (CK) (Cohen 1960),

$$OA = \frac{\sum i_{..} - ef}{N - ef}, \tag{2.5}$$

Table 2.4: A traditional error matrix showing the calculation of the user's, producer's and overall classification accuracies.

		Reference Data				
		Class 1	Class 2	Class 3	Total	User's Accuracy
Classified As	Class 1	i11	i12	i13	i1.	(i11 / i1.)
	Class 2	i21	i22	i23	i2.	(i22 / i2.)
	Class 3	i31	i32	i33	i3.	(i33 / i3.)
	Total	i.1	i.2	i.3	i..	
Producer's Accuracy		(i11 / i.1)	(i22 / i.2)	(i33 / i.3)	Overall Accuracy	(i11+i11+i11) / i..

where: *CK* is Cohen's Kappa, $\Sigma i_{..}$ is the total number of correctly classified pixels, and *ef* is the expected frequency for the number of agreements that would have been expected by chance for each category. for this purpose, I assume NDVIs is the observed ground truth for the NDVIsn, NDVIso, and NDVIsa classified images. For the winter pairings, the NDVIw observations are considered to be the observed ground truths for the NDVIwn, NDVIwo, and NDVIwa classifications.

The small-area assessment measures accuracy on the entries in the combined land cover validation data set (Section 2.3). I used ArcGIS' *ZonalStatisticsasTable* to assign each polygon the majority land cover class from each classification. Similar to above, I used the NDVIs and NDVIw results as the observed truth and the results from KNN-repaired image classifications as the classified value to determine the overall classification rates (Eq. 2.4) and Cohen's Kappa (Eq. 2.5).

In addition to scene-wide and small-area assessments, I made visual evaluations of the repaired imagery. Though a subjective inquiry, I consider it an important aspect of any investigation of this sort to assure the outputs appear to be suitable for use. I examined the repaired imagery for the:

- presence of speckle or "salt and pepper" effects, the
- ability to reconstruct linear and irregular features, and the
- condition of areas with recent changes in land cover.

2.4 RESULTS

2.4.1 Percentage of exactly repaired pixels (residual = 0)

Near-season pairings for both winter and summer ETM+ scenes delivered the largest number of zero-error pixels in all bands with success rates ranging from 13% to 26% and 10% to 23%, respectively (Tab. 2.5, Fig. 2.6). The winter near-date pair for all bands had 2% to 3% more zero-error pixels than the anniversary-season pair and 3% to 5% more than the opposite-season pair. Summer near-season pairings had 0.4% to 2% more zero-error pixels than the anniversary-season and 2% to 6% more than the opposite-season grouping for the same range of bands. Anniversary date pairings used in the repair of both the winter and summer ETM+ yielded the next highest zero-error success

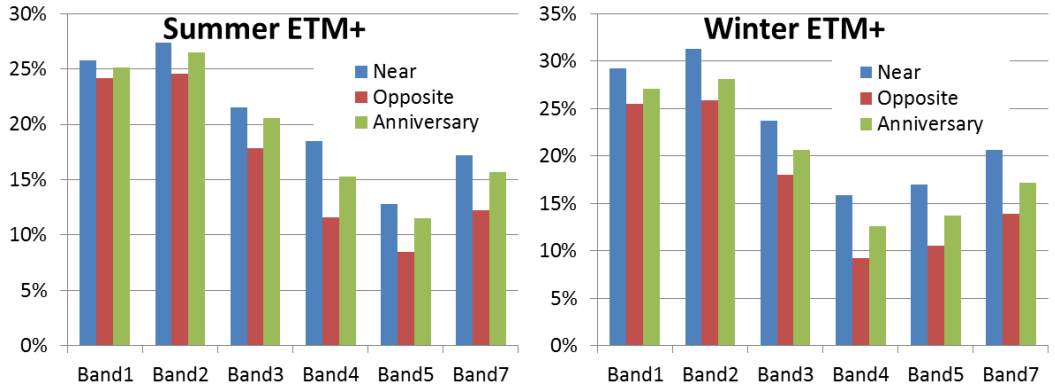

Figure 2.6: The percentage of zero-residual pixels for the summer and winter ETM+ and for each pairing.

rate with 10% to 24% and 9% to 23%, respectively. Opposite-season pairings that combined the summer ETM+ with a winter TM and the winter ETM+ with a summer TM yielded the fewest zero-errors (8% to 22% for the winter and 7% to 20% for the summer).

Table 2.5: Percentage of zero-residual pixels for each band and season pairing.

Pairing (ETM/TM)	Band1	Band2	Band3	Band4	Band5	Band7	Min.	Max.
Summer/Summer	21%	23%	17%	15%	10%	14%	10%	23%
Summer/Winter	19%	20%	14%	10%	7%	10%	7%	20%
Summer/Anniv.	20%	23%	17%	13%	9%	13%	9%	23%
Minimum	19%	20%	14%	10%	7%	10%	NA	NA
Maximum	21%	23%	17%	15%	10%	14%	NA	NA
Winter/Winter	24%	26%	19%	13%	14%	16%	13%	26%
Winter/Summer	21%	22%	15%	8%	9%	11%	8%	22%
Winter/Anniv.	22%	24%	17%	10%	11%	14%	10%	24%
Minimum	21%	22%	15%	8%	9%	11%	NA	NA
Maximum	24%	26%	19%	13%	14%	16%	NA	NA

2.4.2 Minimum DN representing the 95^{th} percentile residual

The near-season pairings and all bands 1-5 and 7 yielded the lowest DN threshold for the 95^{th} percentile residual. Residuals for the winter near-season pair ranged from 4 DN to 8 DN, while the

residuals for the summer near-season pair ranged from 5 DN to 17 DN (Tab. 2.6). Thresholds for winter pairings were 33% to 56% lower than the thresholds for the opposite-season and 20% to 43% lower than the anniversary pair. The pixel repairs of the summer ETM+ using the near-season pair yielded a 95^{th} percentile residual 18% to 31% lower than the summer opposite-season 95^{th} percentile residuals and 0% to 18% lower than the anniversary pairing 95^{th} percentile residuals. Summer near-season and anniversary-season pairing had equal band 1 residuals, yet the former required two nearest neighbors while the latter required three nearest neighbors to achieve that level. Anniversary pairings for both winter and summer produced the next lowest thresholds.

The band 4 is the only instance when the summer opposite-season pairing produced a smaller DN threshold than the winter opposite season pair. There were three cases where the summer pairings yielded equal thresholds to the winter pairings (opposite-season band 1 and anniversary season bands 1 and 2). The winter pairings produced lower thresholds for all the other combinations and bands. The number of nearest neighbors required to reach the minimum DN representing 95^{th} percentile residual threshold for all pairings initially decreased and then leveled off and remained relatively constant, and in a couple cases, it increased slightly (2.7). The optimal number of neighbors varied between different bands and neither K=1 nor K=20 yielded the optimal result for any of the bands.

Table 2.6: Minimum DN representing the 95^{th} percentile residual and the number of nearest neighbors (K) required to reach that point.

ETM+ Season	Pairing	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Summer	Near	5 (2)	5 (7)	9 (2)	9 (3)	17 (5)	16 (3)
Summer	Opposite	6 (2)	7 (2)	11 (3)	13 (4)	23 (4)	20 (5)
Summer	Anniversary	5 (3)	6 (2)	10 (2)	11 (3)	20 (5)	18 (4)
Winter	Near	4 (2)	4 (3)	6 (2)	8 (3)	8 (9)	7 (6)
Winter	Opposite	6 (2)	6 (6)	10 (3)	14 (5)	18 (4)	14 (5)
Winter	Anniversary	5 (3)	6 (2)	9 (2)	10 (6)	14 (6)	12 (5)

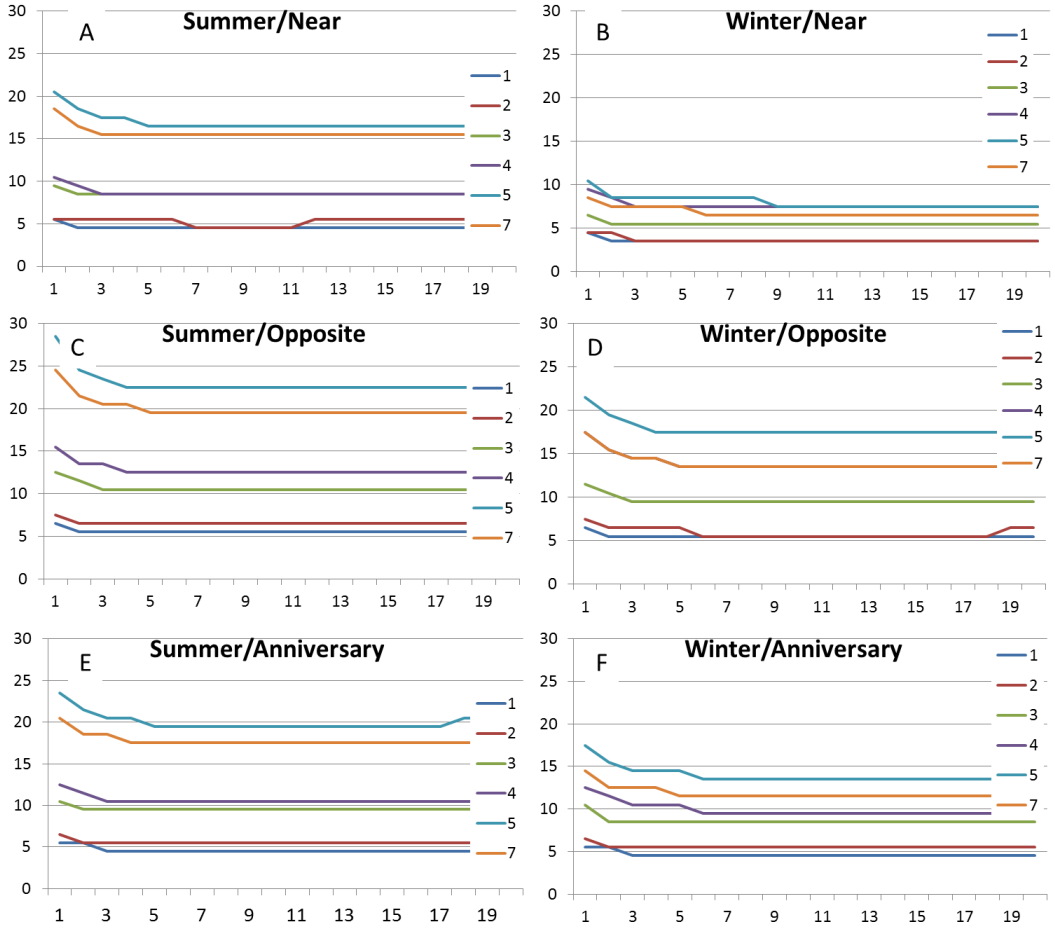

Figure 2.7: Minimum DN representing the 95^{th} percentile residual for bands 1 - 5 and 7 across K = 1 to 20.

2.4.3 RMSE based on bands 1 - 5 and 7

Near-season pairings produced the lowest RMSE across all values of K. The winter minimum occurred at K=7 with an RMSE=6.9 DN and at K=8 yielding an RMSE=11.6 DN for summer (Tab. 2.7, Figs. 2.8A and 2.8B). The winter ETM+ repair processed using the near-season pairing produced RMSEs 75% lower than the opposite-season and 44% lower than the anniversary-season pairings. The summer near-season pair yielded an RMSE 25% lower than the opposite-season and 11% lower than the anniversary-season summer pairs. Anniversary-groups for both winter and summer proved to be the next best by producing RMSEs 21% and 13% higher, respectively, than their opposite season counterparts. The best winter ETM+ repair process, for each pairing, resulted in a lower RMSE than their summer counterparts. The best near-season pairing was 69% lower than the best summer near season pair. The best winter anniversary pair was 21% lower and the opposite season, which was 29% lower than their summer equivalents.

Table 2.7: Scene-wide RMSE measures based on bands 1 - 5 and 7 for each season pairing.

	Winter			Summer		
Near Nbrs. (K)	Near	Opposite	Anniversary	Near	Opposite	Anniversary
1	8.387	14.964	12.266	14.119	18.065	15.589
2	7.416	13.182	10.823	12.509	15.905	13.82
3	7.103	12.582	10.346	11.984	15.176	13.252
4	6.965	12.3	10.125	11.755	14.833	13.002
5	6.901	12.147	10.01	11.643	14.654	12.887
6	6.87	12.061	9.949	11.588	14.553	12.83
7	**6.861**	12.013	9.918	11.567	14.498	**12.811**
8	6.863	**11.985**	9.905	**11.564**	14.466	12.813
9	6.874	11.973	**9.903**	11.574	**14.457**	12.827
10	6.89	11.969	9.91	11.591	14.457	12.85
11	6.911	11.972	9.922	11.615	14.466	12.877
12	6.933	11.98	9.938	11.643	14.48	12.908
13	6.958	11.992	9.956	11.673	14.499	12.942
14	6.985	12.007	9.976	11.705	14.522	12.977
15	7.012	12.023	9.998	11.738	14.547	13.014
16	7.04	12.041	10.02	11.772	14.574	13.05
17	7.069	12.06	10.044	11.806	14.601	13.088
18	7.099	12.079	10.068	11.842	14.63	13.125
19	7.128	12.1	10.093	11.877	14.659	13.162
20	7.158	12.12	10.118	11.913	14.689	13.199

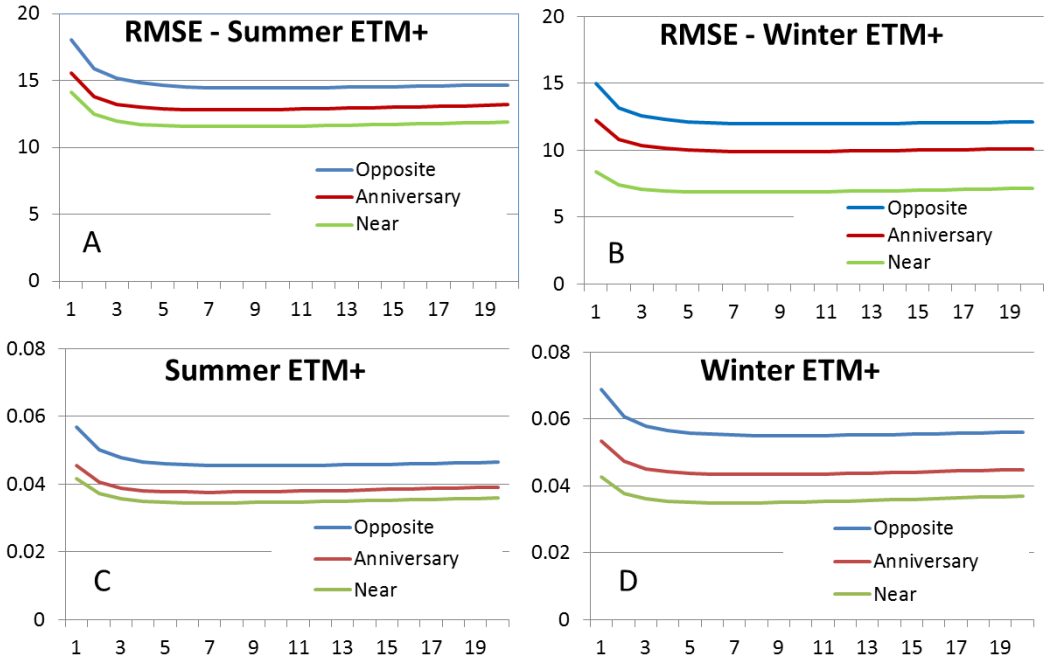

Figure 2.8: Root-mean squared error for each season pairing plotted across k = 1 to 20. Frames A) and B) display the RMSE based on the bands 1-5 DN and frames C) and D) the AD based on the NDVI transformation.

2.4.4 NDVI mean absolute difference

NDVI-based mean absolute difference (AD) behaved in a similar manner as the DN-based measures with an initial decrease and then a gradual leveling off or a slight increase (Fig. 2.8C and 2.8D). Absolute mean differences for winter near-season pairings were 20% lower than those from the winter-anniversary pairs and 37% lower than those generated with the opposite season group (Tab. 2.8). Summer near-season pairings resulted in ADs 8% and 24% lower than those generated from the anniversary and opposite-season groups. The summer near-season pair produced an AD slightly lower than the winter near-season pair, which produced 1% lower than the opposite season. Both of these minimums occurred at K=7. The summer anniversary AD was 15% lower than the winter anniversary-season pair and was processed at K=7, while the winter was processed

at K=9. Both opposite-season pairings produced their minimum using 9 nearest neighbors with the summer group 21% lower than the one from winter.

Table 2.8: NDVI-based mean absolute difference for K=1 to 20.

K	Sum/Near	Sum./Oppo.	Sum./Aniv.	Win./Near	Win./Oppo.	Win./Aniv.
1	0.0417	0.0569	0.0456	0.0426	0.0687	0.0535
2	0.0371	0.0501	0.0405	0.0377	0.0606	0.0473
3	0.0356	0.0477	0.0388	0.036	0.0578	0.0451
4	0.0349	0.0466	0.0381	0.0353	0.0565	0.0442
5	0.0346	0.046	0.0378	0.035	0.0558	0.0437
6	0.0345	0.0457	0.0376	0.0349	0.0554	0.0435
7	0.0345	0.0455	0.0376	0.0348	0.0552	0.0433
8	0.0345	0.0455	0.0376	0.0349	0.0551	0.0433
9	0.0345	0.0454	0.0377	0.035	0.055	0.0434
10	0.0346	0.0455	0.0378	0.0351	0.055	0.0434
11	0.0347	0.0455	0.0379	0.0353	0.0551	0.0435
12	0.0348	0.0456	0.038	0.0354	0.0551	0.0437
13	0.035	0.0457	0.0381	0.0356	0.0552	0.0438
14	0.0351	0.0458	0.0382	0.0358	0.0553	0.0439
15	0.0352	0.0459	0.0384	0.036	0.0554	0.0441
16	0.0354	0.046	0.0385	0.0362	0.0555	0.0442
17	0.0355	0.0461	0.0386	0.0364	0.0556	0.0444
18	0.0356	0.0462	0.0388	0.0366	0.0558	0.0445
19	0.0358	0.0463	0.0389	0.0368	0.0559	0.0447
20	0.0359	0.0464	0.0391	0.037	0.056	0.0448

2.4.5 Scene-wide classification accuracy and Cohen's Kappa.

Overall classification rates for all pairings increased from K=1 to around K=7 to K=9, and then leveled off or slightly decreased (Figs. 9A and 9B). The overall classification rates of the NDVI imagery ranged from 61% of success for the summer opposite-season pair to 73% of success for the winter near-season pair (Tab. 2.9). The maximum near-season classification (K=9) of the winter image was 4% higher than the best anniversary-season and 14% higher than the opposite-season pair. The best summer near-season classification rate, occurring at K=9, was 5% higher than the anniversary and 17% higher than the opposite season pair. The classifications based on the winter ETM+ were 0.1% to 3% higher than for the summer season pairings.

Table 2.9: Maximum overall classification rate based on the supervised classification of the NDVI imagery.

Season	Pairing	Max K	Max OA Rate (%)	Max CK (%)
Summer	Near	9	72.9%	66.3%
Summer	Opposite	7	61.1%	51.5%
Summer	Anniversary	7	69.8%	62.5%
Winter	Near	9	73.0%	66.4%
Winter	Opposite	14	63.0%	53.7%
Winter	Anniversary	9	70.1%	62.4%

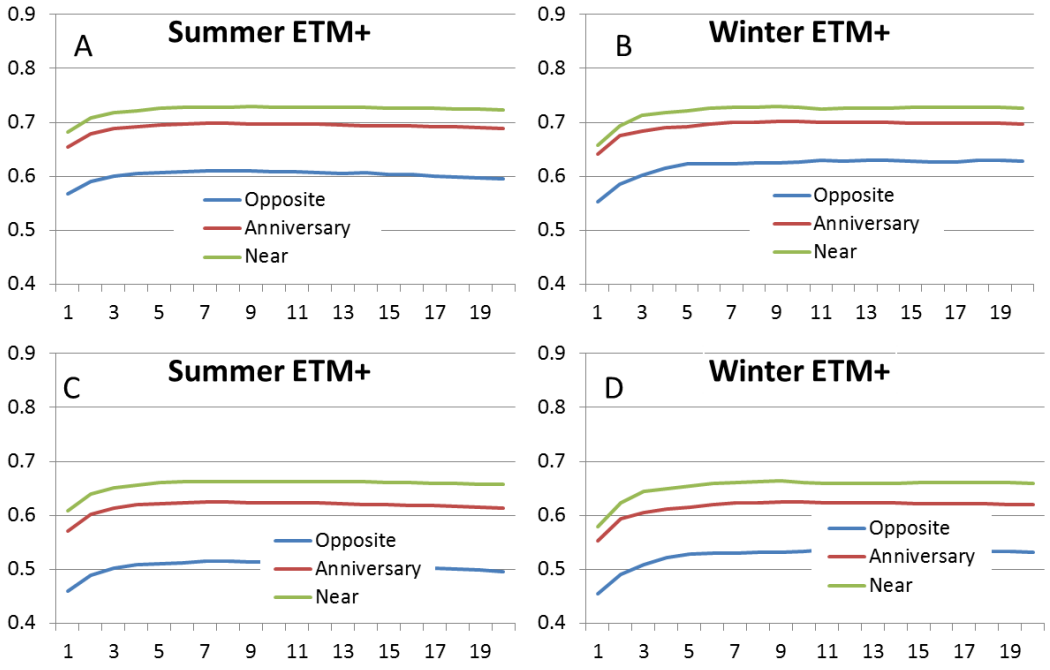

Figure 2.9: Overall classification accuracy and Cohen's Kappa for the classified NDVI images for each pairing and across k=1 to 20. Overall classification results for the summer ETM are shown in frame A and frame B for the winter classification. Cohen's Kappa for each pairing and across k=1 to 20 are shown in frames C and D, respectively.Overall classification accuracy and Cohen's Kappa for the classified NDVI images for each pairing and across k=1 to 20. Overall classification results for the summer ETM are shown in frame A and frame B for the winter classification. Cohen's Kappa for each pairing and across k=1 to 20 are shown in frames C and D, respectively.

The more conservative Cohen's Kappa rates for the near-season pairings were 3.9% and 3.8% higher than the winter and summer anniversary and 13% and 15% higher than the ones from the opposite-season pairing (Tab. 2.9). The winter near-season pairing produced a CK of 66% and the summer near-pairing a CK of 66%. The values increased, just as the overall classification accuracy (OA) did, from K=7 to K=9 and then leveled off or slightly decreased (Fig. 2.9C and 2.9D).

2.4.6 Assessment of combined land cover validation polygons

The winter near-season overall classification rate (81%, K=5) was higher than the winter opposite-season and anniversary-season rates (Tab. 2.10, Fig. 2.10B), and the winter anniversary-season pair outperformed the opposite-season results, 78% (K=4) success rate versus 74% (K=4) rate. Classification accuracy gradually increased for each winter pairing until K=4 or K=5 and then steadily decreased. Summer pairing success rates are ordered the same as the winter pairings (Fig. 2.10A). The summer near-date pair produced the largest OA rate (83%, K=1), then the anniversary-season pair (81%, K=1) and the opposite-season pair had the smallest (72%, K=1) (Tab. 10). Rankings according to Cohen's Kappa success rates are identical to the overall classification rates (Tab. 2.10). The near-season pairings for both winter (CK=77%, K=5) and summer (CK=79%, K=1) produce the largest rates. The anniversary-season pairings give the next highest rates for their respective seasons and the opposite-season pairings give the lowest. Depending on the season pairing, Cohen's Kappa increases to K=4 or K=5 (Fig. 2.5C and D) and then levels off. On the other hand, maximum Cohen's Kappa rates for the summer pairs occur at K=1 for each pair (Tab. 2.10, Figs. 2.10C and 2.10D).

Table 2.10: Maximum overall classification rate, Cohen's Kappa and the K at which it occurred.

Season	Pairing	Max OA	K @ Max OA	Max CK	K @ Max CK
Summer	Near	83%	1	79%	2
Summer	Opposite	72%	1	65%	2
Summer	Anniversary	81%	1	76%	1
Winter	Near	81%	5	77%	5
Winter	Opposite	74%	4	67%	4
Winter	Anniversary	78%	4	73%	4

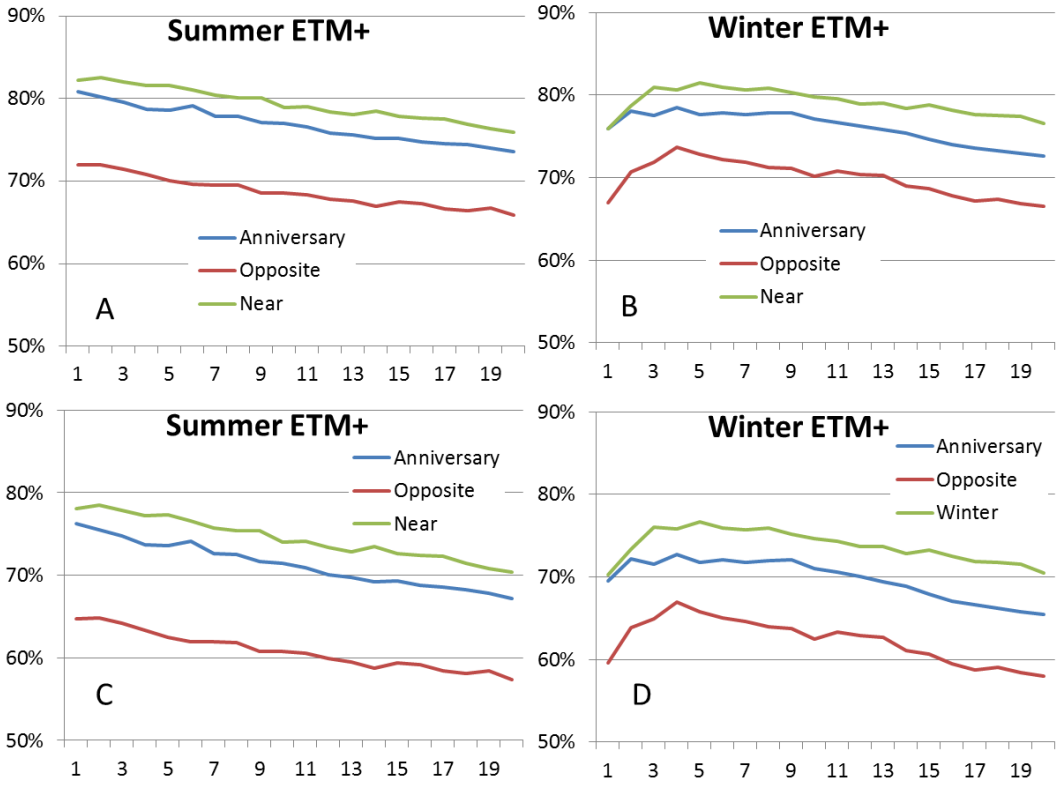

Figure 2.10: Assessment at combined GLUT/NLCD land cover polygons. Frames A) and B) show the summer and winter overall classification rate for each pairing across k=1 to 20. Frames C) and D) show the Cohen's Kappa.

2.4.7 Visual assessment

Speckle was visible in areas of homogeneous land cover and near road intersections when using few near neighbors (Fig. 2.11B and Fig. 2.12B). However, it gradually disappeared as K increased (Figs. 2.11C 2.11F and 2.12C - 2.12F). I observed artifacts in areas with clouds (Fig. 2.13) that did not dissipate with K (Figs. 2.13B 2.13F). Narrow features similar to the river visible in Figure 2.11, the road in Figure 2.12, transitions among crops and clearcut areas lose detail at low values of K. Often, though, the boundaries tended to come back as K increased.

Land cover change did affect the visual appeal of the repaired image when it occurred after the capture of the TM image employed in the repair process. Figure 2.14 highlights such a case. A harvest occurred on this site between November 23, 2005 and December 1, 2005 the near-date ETM+ (Fig. 2.14B) and TM (Fig. 2.14C) capture dates. Thus, the near-date image repair was not optimal (Figs. 2.14D, 2.14E, and 2.14F). Visual appeal degenerated further for the opposite-season image repair (Fig. 2.14G). In contrast, the anniversary-season repair process captured most of the detail of the site since the TM's acquisition date was post-harvest (Fig. 2.14F).

2.5 Discussion and Conclusions

The discontinuity of spectral information in the SLC-off ETM+ satellite data renders much of its imagery unusable for analysis across the continuous landscape. The nearest neighbor method presented in this research and elsewhere (Cieszewski and Lowe 2008; Lowe and Cieszewski 2009; Qingmin Meng et al. 2009; J. Chen et al. 2011; Jin et al. 2013) is a simple approach for repair of defective images. The process relies on the availability of companion images with suitable pairs of repair pixels. This research demonstrates the importance of selecting the companion images used in the repair process. It is understandable that the best choice is to select a companion image captured at a date the closest possible to the ETM+ acquisition date. In this study, the near-date pairings yielded the highest success rates in terms of the number of zero-residual pixels (Tab. 2.5) and with the lowest DN threshold representing the 95^{th} percentile residual (Tab. 2.6) for all bands 1-5 and 7. Moreover, these pairings produced the lowest DN-based RMSEs (Tab. 2.7), resulted in

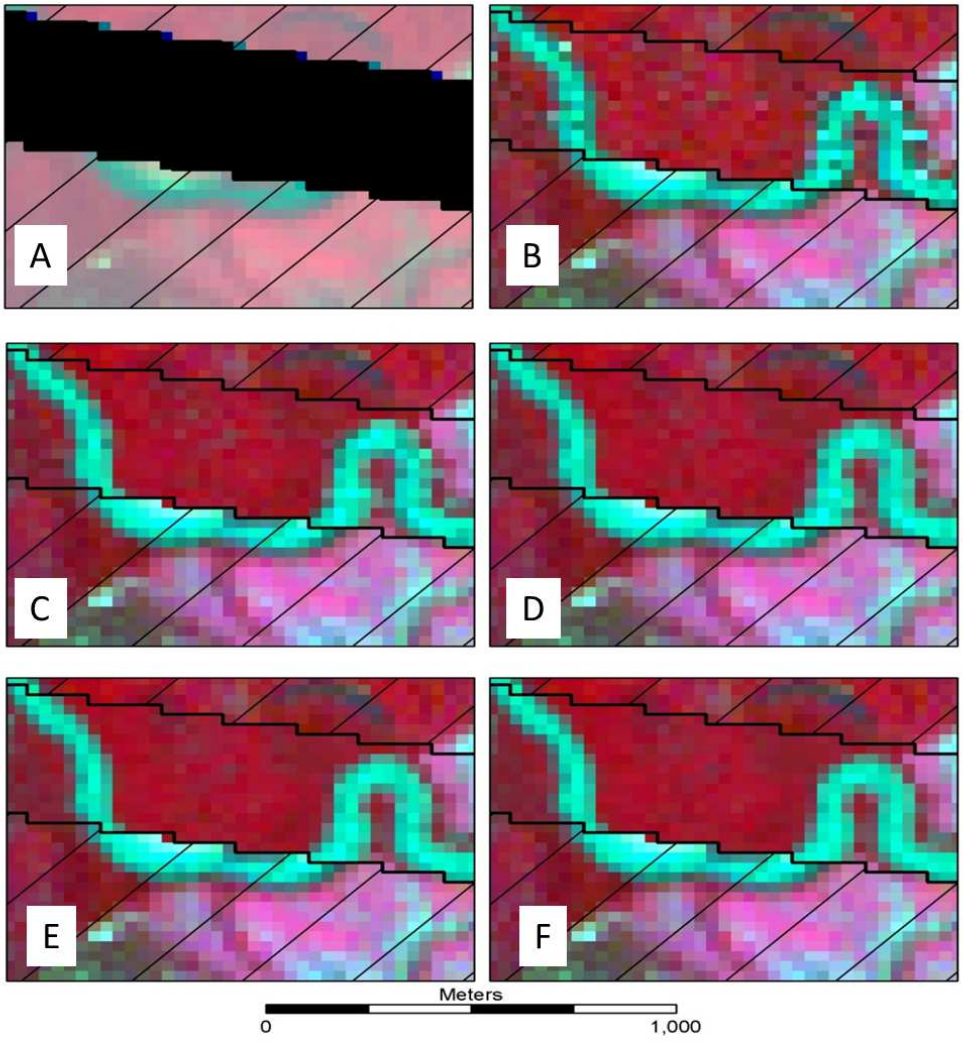

Figure 2.11: Riparian area in the summer ETM+ image restored using the KNN-Gap filling process. Frame A) shows the original unprocessed image, frame B) the scene processed using K=1, C) using K=5, D) using K=10, E) using K=15, and F) using K=20.

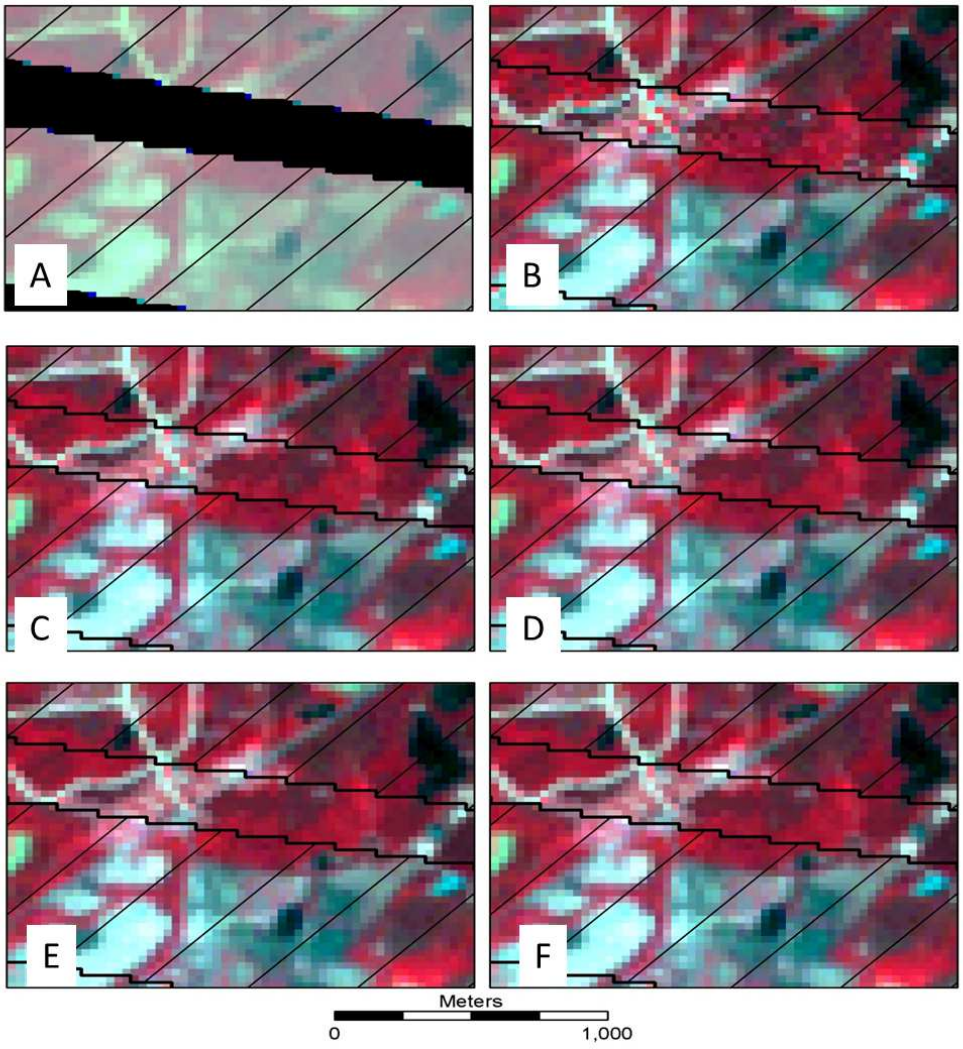

Figure 2.12: Urban area in the winter ETM+ image restored using the KNN-gap filling process. Frame (A) shows the original unprocessed image, frame B the scene processed using K=1, (C) using K=5, (D) using K=10, (E) using K=15, and (F) using K=20.

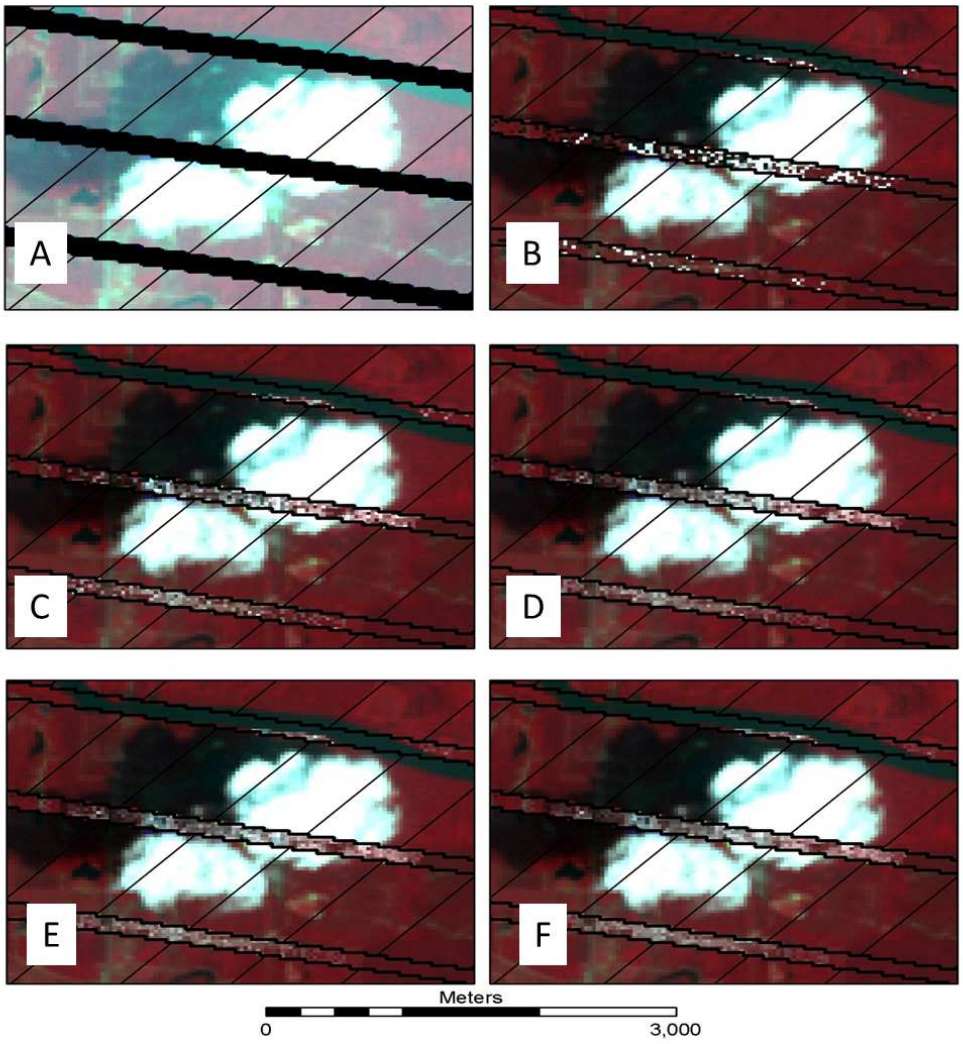

Figure 2.13: Cloud and cloud shadow shown in the summer ETM+ image repaired using the KNN-gapfill process. A) An EMT+ unprocessed image and the same image processed with B) K=1,C) K=5,D) K=10, E) K=15, and F) K=20.

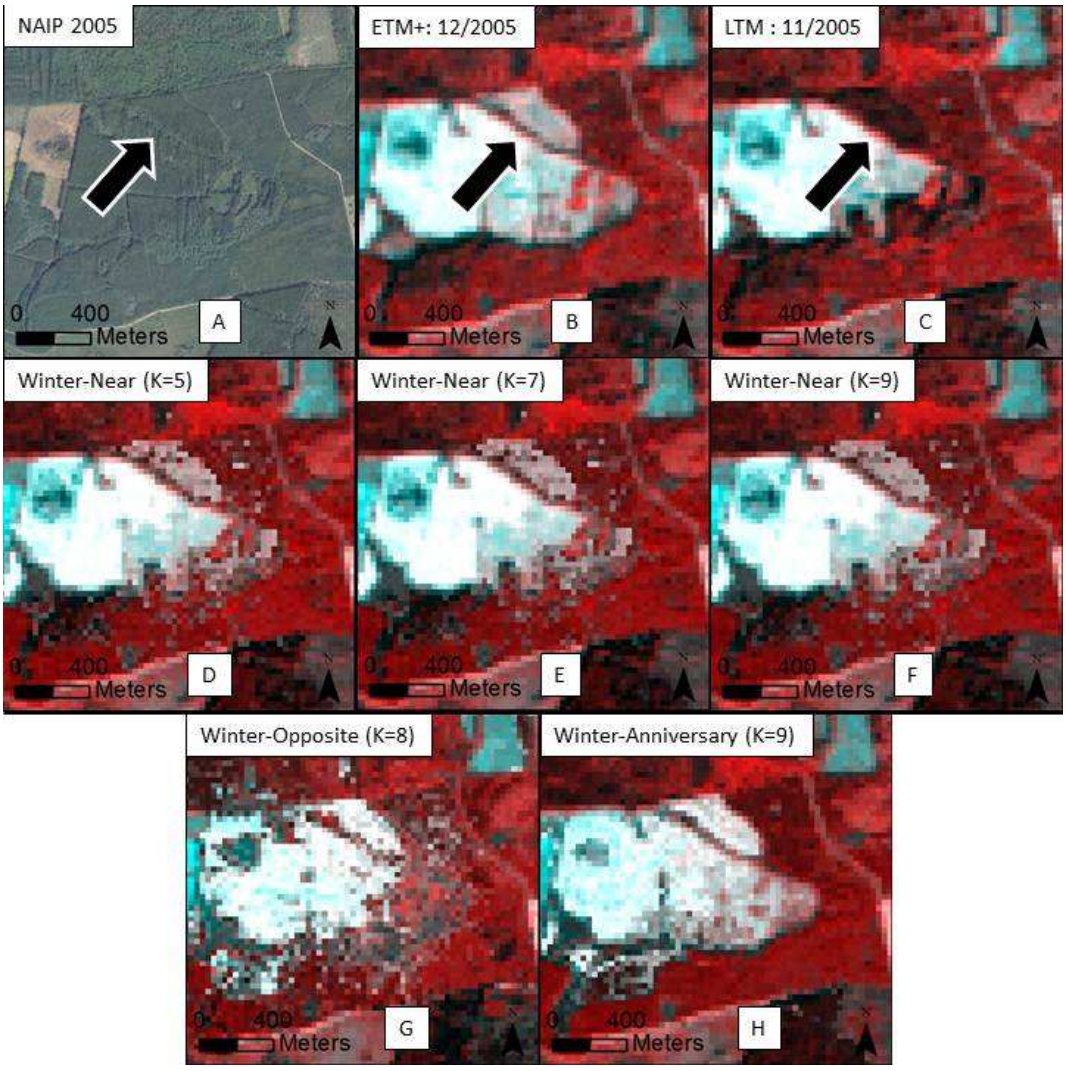

Figure 2.14: The visual effect of a harvested area present on the repaired ETM+ yet not on the near-season or the opposite-season TM. A) The NAIP 2005 aerial photograph showing eth unharvested area; B) the ETM+ image before repair; C) the TM image used in the near-date repair which does not show the harvest; D) the repaired ETM+ image using K=5; E) the repaired ETM+ image using K=7; F) the repaired ETM+ image using K=9; G) the repaired ETM+ using the opposite-season TM and K=8, and H) the repaired ETM+ using the anniversary-season TM and K=9.

the lowest absolute mean differences (Tab 2.8), and returned the largest classification accuracies for the entire scene (Tab. 2.9) and for a polygon-based validation dataset (Tab. 2.10). However, it is unclear if the next best choice for a companion image is one captured closer in time yet in a different season or further in time but in the same season as the damaged image. It is expected that in case of some dynamically changing landscape conditions, such as fast growth of young plantations or presence of disturbances the proximity in time might have the highest value. On the other hand in the presence of fairly static conditions minimizing the distance of time might be less beneficial than maximizing similarities in seasonal conditions.

In this study, anniversary-season imagery, a TM image captured in the same season yet a year later, yielded the next highest success rates. Despite a shorter time between the dates of the repair scene and the companion scene, the opposite-season pairs yielded the lowest success rates in all measures. Thus, when selecting companion imagery for the gap-filling process, it seems the best to consider the imagery captured in the same season as close in time to the repaired image as possible. If the same season near-date image is not available, the next best choice is an image captured in the same season in the preceding or following year. Imagery captured in a different season are the least desirable choices.

One needs to consider carefully the dates of the companion imagery in relation to the capture date of the repair image in areas where a disturbance may have occurred. It is clear that any changes in land cover that occur between the acquisition time of the companion image and the repaired image will result in erroneous repairs (Figs. 2.14D 2.14F). However, in the example shown in Figure 2.14, there was no issue repairing the harvest site using the alternate-season ETM+/TM pair since the TM was captured after the forest activities. In areas experiencing drastic changes one should exercise caution in selecting companion images whose acquisition times corresponded with the greatest differences between the companion and the repair images. For examples if a stand was harvested around the time of the acquisition of the repair image there could be the following scenarios. If the preceding image captured mature stands and the repair image captured a cleared land, then it is clear that the better choice for a companion image is at the latter time. However,

if the repair image captured a still mature stand then the clear choice is to work with the earlier capture time of the companion image. Since the land changes captured by the satellite imagery may be considered random in space and time, this leads to the conclusion that a single repair image may call for both selections of the companion images from before and after its capture time. Furthermore, since there could be various other obstacles in matching the best pixel pairs for images repair, the process would conceptually work the best with a multi-scene and multi-date companion image analysis whereby repair of each individual pixel could be optimized based on different date selections of the companion imagery choices.

The optimal number of nearest neighbors that yielded the best results varied depending on the season pairings. The DN-based RMSE (Tab. 2.7), NDVI-based AD (Tab. 2.8), and the scene-wide classification accuracy results (Tab. 2.9) showed the best K ranging from 7 to 9 when using the near-date and anniversary-season companion images. The range was slightly higher for the opposite-season companion images. Notwithstanding the above, the ideal number of nearest neighbors determined at other sites is likely to vary; and therefore, it seems desirable to evaluate each individual project for a range of K before deciding what value of K is the best for that project. For example, Chen et al. (2011) determined that 20 nearest neighbors were the best in their study while the same value of K in this study yielded inferior results.

Overall, the visible bands 1 through 3 were repaired more accurately than the near-infrared and mid-infrared bands 4, 5, and 7. The number of zero-residual estimates from the near-season visible bands ranged from 17% to 23% for the summer and 19% to 26% for the winter while the infrared bands ranged from 10% to 15% for the summer and 13% to 16% for the winter. Furthermore, the DN representing 95th percentile residual ranged from 5 to 9 for the summer and 4 to 6 for the winter while the ranges for the infrared bands were 9 to 17 and 7 to 8, which might be of interest to users wishing to use the repaired data in transformed space like NDVI.

One issue in the repair process associated with areas that contain cloud or other environmental disturbances like smoke from forest fires (Fig. 2.13) appeared to be that the repair process across all K for all season pairing produced speckled results. This was likely due to the fact that the

constrained nearest neighbor search area was too restrictive, which implies that the search area should be identified and processed with a special consideration for the fact that a too restrictive range will hinder the repair process (Qingmin Meng et al. 2009; Zhu et al. 2012).

2.6 REFERENCES

[1] Bdard, F., G. Reichert, R. Dobbins, and I. Trpanier. 2008. *Evaluation of Segment-based Gap-filled Landsat ETM+ SLC-off Satellite Data for Land Cover Classification in Southern Saskatchewan, Canada.* International Journal of Remote Sensing 29 (7) (April): 20412054. doi:10.1080/01431160701281064.

[2] Chen, Jin, Xiaolin Zhu, James E. Vogelmann, Feng Gao, and Suming Jin. 2011. *A Simple and Effective Method for Filling Gaps in Landsat ETM+ SLC-off Images.* Remote Sensing of Environment 115 (4) (April 15): 10531064. doi:10.1016/j.rse.2010.12.010.

[3] Chen, M, and Josef Cihlar. 2000. *Retrieving Leaf Area Index of Boreal Conifer Forests Using Landsat TM Images* 162 (August 1995): 153162.

[4] Cieszewski, CJ., RC Lowe. *Biomass InFORM (Interactive Fast Online Reports and Maps).* GrowthAndYield. 2007. Available At: http://www.growthandyield.com/maps/InFORMB/GA/state_tmplate.html

[5] Cieszewski, CJ and RC Lowe. 2008. Generic gapfilling method for reconstructions of missing data using KNN approach on multitemporal scene pairings. Fiber Supply Assessment Technical Report 2008-1. WSFNR, University of Georgia. Athens GA 30602. 2p.

[6] Curran, Paul J, Jennifer L Dungan, and Henry L Gholz. 1992. *Seasonal LAI in Slash Pine Estimated with Landsat TM.* Remote Sensing of Environment 39 (1) (January): 313. doi:10.10160034-4257(92)90136-8.

[7] EPA (United States Environmental Protection Agency), 2000. Multi-Resolution Land Characteristics Consortium (MRLC) database. Available At: http:www.epa.govmrlcpage.

[8] Cohen, J. 1960. *A coefficient of agreement for nominal scales.* Educ. Psychol. Meas. 20, 37-46.

[9] Gillis, Mark D. 2001. *Canada's National Forest Inventory (Responding to Current Information Needs).* Environmental Monitoring and Assessment 67 (1) (February): 121129. doi:10.1023/A:1006405820244.

[10] Gjertsen, Amt Kristian. 2007. *Accuracy of Forest Mapping Based on Landsat TM Data and a kNN-based Method.* REMOTE SENSING OF ENVIRONMENT 110 (4) (October): 420430. doi:10.1016/j.rse.2006.08.018.

[11] Howard, By Stephen M, James M Lacasse, and S M Howard. 2004. *An Evaluation of Gap-Filled Landsat SLC-Off Imagery for Wildland Fire Burn Severity Mapping.* Photogrammetric Engineering Remote Sensing 70 (August): 877880.

[12] Jin, Suming, Collin Homer, Limin Yang, George Xian, Joyce Fry, Patrick Danielson, and Philip A. Townsend. 2013. *Automated Cloud and Shadow Detection and Filling Using Two-date Landsat Imagery in the USA*. International Journal of Remote Sensing 34 (March) (March 10): 3741. doi:10.1080/01431161.2012.720045.

[13] Lowe, RC, CJ Cieszewski. 2009. KNN Gapfill method to repair SLC-off Landsat ETM+ satellite imagery. Presentation given at the 7th Southern Forestry and Natural Resources GIS Conference, held on Dec. 7-9, 2009, Athens, Georgia.

[14] Maselli, F., and M. Chiesi. 2006. *Integration of Multi-source NDVI Data for the Estimation of Mediterranean Forest Productivity.* International Journal of Remote Sensing 27 (1) (January 10): 5572. doi:10.1080/01431160500329486.

[15] Maxwell, and S. 2004. *Filling Landsat ETM Plus SLC-off Gaps Using a Segmentation Model Approach* 70 (10): 11091111.

[16] Maxwell, S. K., G. L. Schmidt, J. C. Storey, Maxwell, K. S., Schmidt, L. G., Storey, and C. J. 2007. *A Multi-scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-off Images.* International Journal of Remote Sensing 28 (23) (January): 53395356. doi:10.1080/01431160601034902.

[17] McRoberts, R., and E. Tomppo. 2007. *Remote Sensing Support for National Forest Inventories.* Remote Sensing of Environment 110 (4) (October): 412419. doi:10.1016/j.rse.2006.09.034.

[18] McRoberts, Ronald E., Erkki Tomppo, Klemens Schadauer, Claude Vidal, Goran Stahl, Gherardo Chirici, Adrian Lanz, Emil Cienciala, Susanne Winter, and W. Brad Smith. 2009. *Harmonizing National Forest Inventories.* Journal of Forestry (June): 179187.

[19] McRoberts, Ronald, Erkki Tomppo, and Erik Naesset. 2010. *Advances and Emerging Issues in National Forest Inventories.* Scandinavian Journal of Forest Research 25 (4) (August): 368381. doi:10.1080/02827581.2010.496739.

[20] Meng, Q, C Cieszewski, and M Madden. 2009. *Large Area Forest Inventory Using Landsat ETM+: A Geostatistical Approach.* ISPRS Journal of Photogrammetry and Remote Sensing 64 (1) (January): 2736. doi:10.1016/j.isprsjprs.2008.06.006.

[21] Meng, Qingmin, Bruce E Borders, Chris J Cieszewski, and Marguerite Madden. 2009. *Closest Spectral Fit for Removing Clouds and Cloud Shadows.* Photogrammetric Engineering & Remote Sensing 75 (5): 569576.

[22] NARSAL (Natural Resources Spatial Analysis Laboratory), 2001. Georgia Land Use Trends (GLUT) project data. Athens: University of Georgia Institute of Ecology.

[23] Reese, Heather, Mats Nilsson, Tina Granqvist Pahln, Olle Hagner, Steve Joyce, Ulf Tingelf, Mikael Egberth, and Hkan Olsson. 2010. *Countrywide Estimates and Data Satellite Using Inventory Forest of Forest Data Field Variables the From.* Environment.

[24] Sader, Steven A, Robert B Waide, and William T Lawrence. 1989. *Tropical Forest Biomass and Successional Age Class Relationships to a Vegetation Index Derived from Landsat TM Data.* Biomass 156: 143156.

[25] Story, M., R. Congalton. 1986. *Accuracy assessment: a user's perspective.* Photogrammetric Engineering and Remote Sensing 52 (3): 397-399.

[26] Tomppo, E, H Olsson, G Stahl, M Nilsson, O Hagner, and M Katila. 2008. *Combining National Forest Inventory Field Plots and Remote Sensing Data for Forest Databases.* Remote Sensing of Environment 112 (5) (May): 19821999. doi:10.1016/j.rse.2007.03.032.

[27] Turner, David P, Warren B Cohen, Robert E Kennedy, Karin S Fassnacht, and John M Briggs. 1999. *Relationships Between Leaf Area Index and Landsat TM Spectral Vegetation Indices Across Three Temperate Zone Sites* 68 (April 1998): 5268.

[28] USGS (United States Geological Survey), 2003. Preliminary Assessment of Landsat 7 ETM+ Data Following Scan Line Corrector Malfunction.

[29] Vieira, I. 2003. *Classifying Successional Forests Using Landsat Spectral Properties and Ecological Characteristics in Eastern Amaznia.* Remote Sensing of Environment 87 (4) (November 15): 470481. doi:10.1016/j.rse.2002.09.002.

[30] Wilson, Emily Hoffhine, and Steven a Sader. 2002. *Detection of Forest Harvest Type Using Multiple Dates of Landsat TM Imagery.* Remote Sensing of Environment 80 (3) (June): 385396. doi:10.1016/S0034-4257(01)00318-2.

[31] Xiao, Xiangming, Stephen Boles, Jiyuan Liu, Dafang Zhuang, and Mingliang Liu. 2002. *Characterization of Forest Types in Northeastern China, Using Multi-temporal SPOT-4*

VEGETATION Sensor Data. Remote Sensing of Environment 82 (2-3) (October): 335348. doi:10.1016/S0034-4257(02)00051-2.

[32] Zhu, Xiaolin, Desheng Liu, and Jin Chen. 2012. *A New Geostatistical Approach for Filling Gaps in Landsat ETM+ SLC-off Images.* Remote Sensing of Environment 124 (September): 4960. doi:10.1016/j.rse.2012.04.019.

CHAPTER 3

ASSESSMENT OF STREAM MANAGEMENT ZONES AND ROAD BEAUTIFYING BUFFERS IN GEORGIA BASED ON REMOTE SENSING AND VARIOUS GROUND INVENTORY DATA[1]

[1]Lowe, R.C., C.J. Cieszewski, S. Liu, Q. Meng, J.P. Siry, M. Zasada, and J. Zawadzki. 2008. *South. J. Appl. For.* SF-08-010. Reprinted here with permission of the publisher.

ABSTRACT

Stream management zones (SMZs) and road beautifying buffers (RBBs) are voluntary in Georgia and have an unknown extent and impact on the state's forest production. We describe analyses of these buffers, including an estimation of their potential areas and volumes, and their distributions in different forest cover types under an assumption of their full implementation. We base this analysis on Landsat 7 Enhanced Thematic Mapper Plus imagery and various sources of ancillary data, such as those from the Georgia Gap Analysis Program, the Forest Inventory and Analysis large-scale forest survey, and various industrial forest ground inventories. We considered stream data classified into trout, perennial, and intermittent streams, which we combined with elevation and slope information to assess buffer widths consistent with Georgia's Best Management Practices rules. Our results indicate that minimum width 12.2-m SMZ buffers would occupy about 4.01% of the total forested area in Georgia and would cover about 4.32% of the state's volume. The area of the wider, 30.5-m SMZ buffers would cover about 8.65% of the total forested area in Georgia and would cover about 9.27% of the state's total volume. The minimum-width 12.2-m RBBs would occupy about 3.64% of the total forested area in Georgia and would cover about 3.52% of the state's volume. The area of the wider, 30.5-m RBBs would occupy almost 8.68% of the total forested area in Georgia and would cover about 8.40% of the state's total volume.

3.1 INTRODUCTION

Our ever-growing population and urban expansion increase demand for clean water and various forest products/uses and decrease areas of forests available for commercial production (Wear and Greis 2002). To compensate for the shrinking land base of commercial forests and to meet the growing demand for forest products, forest managers must improve the efficiency in their management and planning practices. Such improvements are conditioned on the availability of effective tools for evaluation and analysis of potential impacts of various forest management and policy

decisions. In turn, outcomes of such analysis are critical to the knowledge of, for example, how various regulatory constraints affect such business decisions as locations of pulp or saw mills.

One of the main natural resource-related concerns of modern societies is maintaining clean water supplies, which may be affected by nonpoint source (NPS) pollution such as runoff from agricultural fields and forest clearcuts. To protect water resources from NPS pollution, forest managers and landowners in Georgia apply protective buffers, or stream management zones (SMZs), which separate rivers from adjacent pollution sources while filtering sediments, absorbing nutrients, and stabilizing stream banks. SMZs also provide habitat for moist-zone animals and plants and stream organisms, and they shade and cool the streams (Welsch 1991). They consist mostly of riparian habitat area, that is, the area directly adjacent to a waterway that includes the bank vegetation and likely a strip of forest.

One major difference in the riparian area is the vegetation it contains. The vegetation within these areas is strongly influenced by the surface and subsurface water associated with the stream itself. Riparian areas contain distinctly different forest cover types, such as bottomland hardwoods. They also may contain species similar to the surrounding uplands but with relatively higher productivity.

Stream management zones were introduced as a result of federal water quality legislation (the Clean Water Act) created in 1972 when the U.S. Environmental Protection Agency recommended using Best Management Practices (BMP) as a primary method for controlling NPS pollution [1]. The Georgia Forestry Commission issued the current *Georgia's Best Management Practices for Forestry* manual in 1999 (GFC 1999). Currently the state of Georgia chooses a nonregulatory system of voluntary compliance, which it may eventually change to a mandatory system. In the future, the BMP may become mandatory with increased SMZ widths and with further management restrictions (Wenger 1999). The potentially mandatory system could include road beautifying buffers, which are protective strips established along the roads. The road buffers are similar to the SMZs in that they have some positive ecological value by stabilizing the ground, reducing ero-

sion from road runoff, and reducing wind erosion by providing natural windbreaks, but they are established primarily to ensure an aesthetically pleasing landscape for road travelers.

Currently, for perennial streams, BMPs recommend leaving an evenly distributed 11.5 m^2 of basal area per hectare or at least 50% of the canopy cover after a harvest. If the stream is classified as a trout stream, it is necessary to also create an additional no-harvest zone around the stream's bank. For intermittent streams, requirements include leaving 5.7 m^2 of basal area per hectare or at least 25% of the canopy cover after a harvest (GFC 1999). There are no stringent parameters regarding the road buffers, which are likely to have similar widths for all roads since their main aesthetic and wind-breaking functions are independent of road widths.

The extent of potential harvesting limitations on all of Georgia's protective buffers is unknown in terms of both the absolute and relative values of involved areas and volumes in the state. Although a large number of studies on riparian/stream management zones in the southern Unites States have been conducted (see e.g., Wenger 1999), the literature on their extent, assessment, and statistics in the region is rather scarce.

According to rough estimates by the US Forest Service, which to date has not conducted specific inventories of these areas, wetlands and riparian zones occupy less than 10% of total national forest area in the United States (Sedell et al. 2000). Furthermore, because practically nobody has performed such inventories on private lands, it is not known how much of the area is occupied by protective buffers in private forest.

The Southern Appalachian Riparian Team (Van Sickle 2001), working on uniform standards for large-scale riparian areas and watershed assessments for the southern U.S. national forests, calculates the area of riparian zones as the product of stream length and the riparian zone/buffer widths. They identify the lengths of perennial and intermittent stream channels on topographic maps in the form of graphic information systems (GIS) stream layers while assuming minimum, 30.5-m riparian area widths on either side of all perennial and intermittent streams, wetlands, lakes, and ponds. They also propose to estimate ephemeral streams from a sample of topographic

quadrangle maps or from GIS stream simulation, and to address any riparian areas wider than 30.5 m using stream survey data.

Cubbage and Woodman (1993) present an approach to estimating effective harvest area losses and costs by management classes in hypothetical forest management units based on the Georgia Forest Inventory and Analysis (FIA) inventory data. Their estimates of the forested areas in stream management zones varied. They were 4.8% when the calculations were based on the minimum 10.7-m recommendations from BMPs of 1989, 5.3% for 10.7-30.5-m buffers (depending on slope), and 7.09% for when the primary buffer width was assumed 91.5 m.

Limited information regarding riparian zones has been provided by different small-scale studies, such as ecological, biochemical, physiological, socioeconomic, and hydrologic assessments and experiments (e.g., Coweeta Long Term Ecological Research). Rough estimates of riparian areas are also included in Land and Resources Management Plans (LRMPs) prepared for national forests on the basis of the National Forest Management Act of 1976 [2]. However, this information is quite limited and does not provide any detailed statistics, although some preliminary estimates of the areas and volumes contained in the riparian buffers in Georgia have been reported by Zasada et al. (2003).

OBJECTIVES

In this study we set out to make a detailed and complete assessment of all the potential stream and road buffers in Georgia using FIA and other available ground inventory, GIS, and remote sensing data. Our goal was to estimate the areas and volumes of possible stream and road buffers and to assess their potential impact on the wood supply in Georgia using refined methods from the study initiated by Zasada et al. (2003). This article describes our assessment methodology and forest-type-dependent statewide estimates of those buffers.

3.2 DATA

3.2.1 Ground inventory data

We used two types of ground-measurement forest inventory data. To stratify Landsat Enhanced Thematic Mapper Plus (ETM+) images, we used models calibrated on our industrial partners' ground inventory data. These data contained GIS polygons with stand-level inventory statistics based on ground measurements from 1998. The second type of ground inventory data came from the US Forest Service FIA program. The FIA program provides the data for all states through the North Central Research Station website (FIA 2006). This hierarchically organized database comprises files for various inventory levels, such as state survey, county, plot, and within-plot condition, allowing for analysis at various levels of resolution. The plot-level data includes forest type, stand age, size class, stand origin, productivity class, site index, land use class, slope, aspect, elevation, physiographic class or soil group, treatment opportunity class, percentage of unstocked area, stocking, remeasurement period, area expansion factors, volume, growth, mortality, and removals. Condition classes further stratify the plots according to specific combinations of plot attributes. FIA records more than 120 variables on the tree level, some of which are measured (e.g., species, DBH, height, quality, crown ratio and class, damage and its cause) and others of which are computed (e.g., volume, removals, growth and biomass, mortality expansion factors) from formulas (Miles et al. 2001).

3.2.2 Landsat ETM+ imagery data

We used summer and winter Landsat 7 ETM+ images for evaluation of land cover throughout the state. The imagery encompasses the entire state except the extreme northwest and southern parts of Georgia (Fig. 3.1). The following ETM+ scenes were used: (path/row; summer capture date; winter capture date):

- 17/37; July 26, 1999; Nov. 15, 1999
- 17/38; July 26, 1999; Nov. 15, 1999

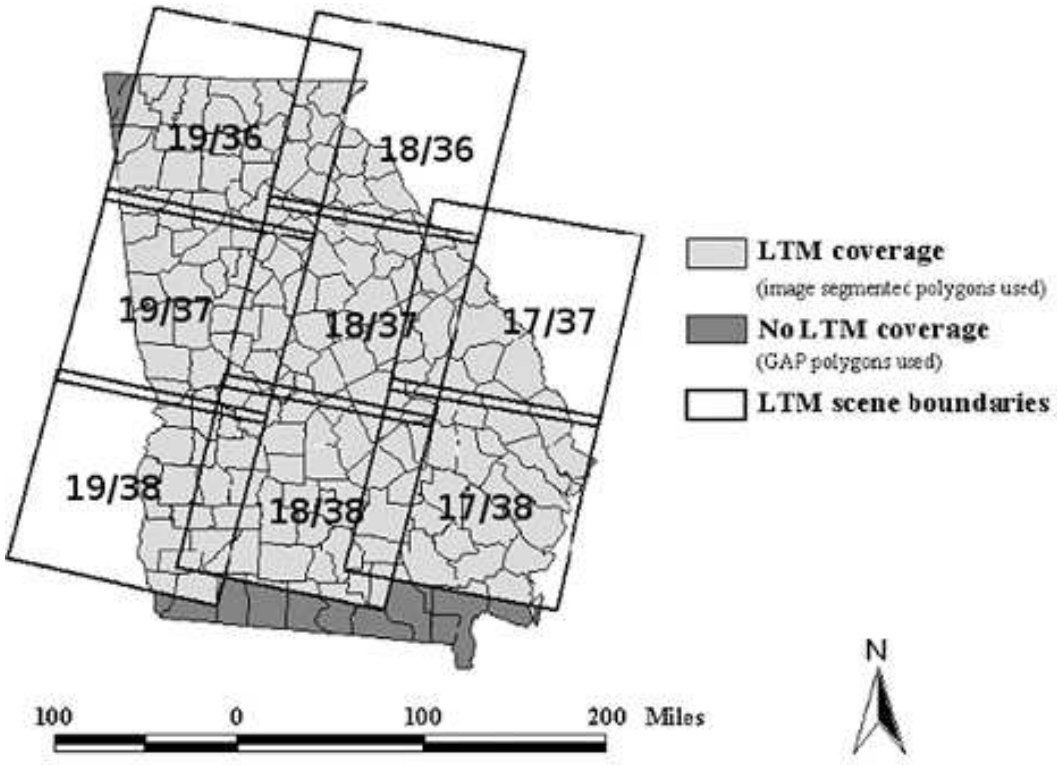

Figure 3.1: Landsat Enhanced Thematic Mapper Plus scene boundaries. TM, Landsat Thematic Mapper.

- 18/36; Apr. 30, 2000; Dec. 8, 1999
- 18/37; Apr. 30, 2000; Dec. 24, 1999
- 18/38; Apr. 30, 2000; Dec. 24, 1999
- 19/36; Sept. 10, 1999; Nov. 29, 1999
- 19/37; Sept. 10, 1999; Nov. 29, 1999
- 19/38; Sept. 10, 1999; Nov. 29, 1999

In this analysis we used only the three visible, one infrared, and two mid-infrared bands captured by the ETM+ sensor. We did not use the thermal band because of its 60-m ground resolution,

which is inconsistent with the other information and gives information with too low of a resolution for analysis of the buffer widths.

3.2.3 Georgia GAP land cover data

For classification of natural land cover types, we used the raster land cover data of the Georgia Gap Analysis Program (GAP) obtained from the Natural Resources Spatial Analysis Laboratory (NARSAL) located at the Institute of Ecology, University of Georgia (Natural Resource Spatial Analysis Laboratory 1998). Using these data we evaluated areas classified as clearcut/sparse, deciduous forest, evergreen forest, mixed forest, and forested wetland, masking out all other cover type classes.

The GAP data encapsulate assessments of the conservation status of native vertebrate species and natural land cover types throughout the United States. Their land cover classification, which was derived from Landsat Thematic Mapper imagery from 1997 and 1998, contains descriptions of 18 general land cover types, including three forest cover types (deciduous, evergreen, and mixed), clearcut/sparse areas, and forested wetlands.

3.2.4 Other GIS data

We used various GIS data obtained from the Georgia GIS Clearinghouse (2009), which serves vector data, including administrative boundaries, roads, and rivers, and also provides raster elevation data (Digital Elevation Model [DEM]/National Elevation Data Sets) for Georgia. The hydrology data set includes rivers, streams, and other linear water bodies captured from United States Geological Survey 7.5-minute topographic maps. It classifies the streams into perennial and intermittent types. Perennial streams flow in a well-defined channel throughout most of the year under normal climatic conditions. Intermittent streams flow in a well-defined channel only during wet seasons. We complemented the vector stream data set with additional water bodies extracted from the GAP data. We also identified trout streams using a Trout Streams of Georgia vector data set from NARSAL.

3.3 METHODS

Following the methodology initially described in Lowe et al. (2003), Zasada et al. (2003), Cieszewski et al. (2005), and Iles (2009), we conducted the analyses in the following six phases:

1. The Image Segmentation phase converted the images to homogeneous polygons and classified them into coniferous, deciduous, and mixed timber stands.

2. The Image Classification phase classified the Landsat ETM+ images relative to pine basal area using industrial ground data.

3. The FIA Information Distribution phase distributed the FIA information to the created polygons using two alternative methods.

4. The Buffer Creation phase defined the buffers using two different buffer widths and three slope classes.

5. The Data Intersection phase intersected the riparian and transportation buffers with the ETM+ -generated polygon data sets.

6. The Buffer Area and Volume Calculation phase summarized the buffer areas and volumes by cover type classes and summarized the results obtained.

3.3.1 Image segmentation

We used the GAP land cover data set to locate the timbered and nontimbered areas. The timbered areas were further stratified into clearcut/sparse, deciduous, evergreen, and mixed forests and forested wetlands. Then, we generated separate evergreen, deciduous, and mixed forest data sets by masking out all areas in the ETM+ scenes that did not match the following criteria:

1. GAP-classified as evergreen forest for the evergreen ETM+ data to derive ETM+ signatures for average, low, and high basal area set,

2. stands based on the ground inventory data. The industry data contained a limited number of large stands with low basal areas, so we

3. GAP-classified as deciduous forest or forested wetland for the deciduous ETM+ data set, or

4. GAP-classified as mixed forest or clearcut/sparse for the mixed ETM+ data set in the absence of any other information about the subsequent regeneration on these areas.

Next, we converted the evergreen, deciduous, and mixed ETM+ data sets to homogeneous polygons using an image segmentation module for ERDAS Imagine (ERDAS 2002) that was developed by the US Forest Service Remote Sensing Applications Center (Ruefenacht et al. 2002). This module iteratively calculates the Euclidean spectral distance between the first pixel in the image (a seed pixel) and other cells in the considered vicinity, that is,

$$\sqrt{\sum_{i}^{6} (S_i - A_i)^2}, \tag{3.1}$$

where *S* the seed pixel's ETM+ value; *A* the compared pixel's ETM+ value; and *i* the ETM+ band (excluding the thermal band).

If the Euclidean spectral distance between two cells is less than or equal to a threshold value specified by the user, the pixel is assigned as a member of the same group as the seed. The assignment of pixels to this region is finished when there are no more pixels adjacent to this region that satisfy the threshold criteria. When the region is closed, the next seed pixel is selected and the process is repeated.

When vectorized, each image-segmented scene produced a data set with more than one million polygons. Since there were too many polygons for our hardware and software to process, we clipped the image-segmented ETM+ scenes to the county boundaries. In-GIS processing, buffering, slope calculation, intersection, and so forth, described from this point on were applied on the county-level data subsets (image classification was carried out and applied at the scene level). We vectorized the image-segmented ETM+ data subsets and combined the evergreen, deciduous,

and mixed ETM+ -generated polygon data. At the end of this phase we refined the final segmentation by merging all polygons smaller than 4.05 ha with larger, adjacent polygons.

3.3.2 Image classification

As a first-stage sample land cover type, we used a GAP data subset containing all coniferous lands. We then sub classified these lands with respect to their basal areas using the ETM+ images and the ground-measurement forest inventory data from our industrial partners. To accomplish the basal area sub classification, we first needed to derive ETM+ signatures for average, low, and high basal area to derive ETM+ signatures for average, low, and high basal area stands based on the ground inventory data. The industry data contained a limited number of large stands with low basal areas, so we needed to first derive the high basal area signature, which could be subsequently used in generating the low basal area signature. To obtain the high basal area signature, we sampled ETM+ pixels within the high basal area stands and averaged their pixel values for bands 1-5 and 7. Next, to define a measure of difference between different basal areas in terms of their associated spectral signatures, we used the Euclidean distances defined by Equation 1.

We calculated the Euclidean spectral distances from the high basal area signature to all the other considered cells. Assuming that the pixels spectrally furthest (as measured by Equation 1) from the high basal area represent the low basal areas, we calculated the low basal area signature as the average signature of the cells with the largest spectral distances from the high basal area signature. To rank all the stands according to their various basal areas (not just low or high), we fitted separate regression models of basal areas as functions of the Euclidean distances from the high basal areas and as functions of Euclidean distances from low basal areas. We used these models to of the Euclidean distances from the high basal areas and as functions of Euclidean distances from low basal areas. We used these models to assign the classified basal areas to both ETM+ scenes for which we had the industrial ground inventory data.

To apply a similar approach to a different unclassified ETM+ scene for which we did not have the ground inventory data, we used the scene overlap areas, treating the classified basal area

values as the ground truth data; we used the previously classified basal areas as if they were based on ground measurements to calibrate new models for the entire unclassified scene. We repeated these steps for all the unprocessed scenes adjacent to those ETM+ scenes that had been previously processed/classified.

3.3.3 FIA information distribution

We distributed the FIA information throughout the ETM+ generated polygon data sets using two separate approaches. In the first approach we iteratively assigned FIA information to the ETM+ -generated polygons using the actual FIA plot forest type, ETM+ -generated polygon forest type, and proximity (distance) to each of the approximate FIA plot locations (the US Forest Service truncated the plot location coordinates before making them available to the public). The FIA data were distributed to the polygons until the sum of all the polygon area was within 99% of the FIA plot's per area unit expansion factor or until there were no more polygons to fill. We applied the following data distribution criteria to each FIA point in sequence.

1. Assign FIA information to polygons that are within 1,000 m of the plot location and of the same general type: evergreen, deciduous, or mixed.

2. Assign FIA information to polygons that are within 12,800 m of the plot location and of the same general type: evergreen, deciduous, or mixed.

3. Assign information from evergreen FIA plots to evergreen or mixed polygons within 6,040 m.

4. Assign information from deciduous FIA plots to deciduous or mixed polygons within 6,040 m.

5. Assign information from mixed FIA plots to mixed or deciduous polygons within 6,040 m.

6. Assign information from any FIA plot to polygons of any type within 6,040 m.

7. Assign information from any FIA plot that has not been fully distributed to polygons of any type that have not been fully filled.

Our second method iteratively assigned FIA information to the ETM+ -generated polygons using ETM+ -estimated pine basal area, polygon size, FIA volume per area unit, and the FIA per area unit expansion factor. First we assigned the average ETM+ -estimated basal area (for the areas that fell within the stand) to the evergreen polygons and ranked them according to basal area from the highest to the lowest. We then ranked deciduous and mixed polygons, and then the FIA plots. We distributed the FIA forest inventory information to the polygons in proportions relative to polygon area and the FIA plot per area unit expansion factor according to the following criteria.

1. Assign information from evergreen FIA plots to evergreen

2. ETM+ -generated polygons and then to mixed and hardwood polygons if more area is needed.

3. Assign information from deciduous FIA plots to deciduous ETM+ -generated polygons and then to any unassigned mixed and hardwood polygons if more area is needed.

4. Assign information from mixed FIA plots to mixed ETM+ -generated polygons and then to any unassigned deciduous or evergreen polygons.

3.3.4 Buffer creation

We converted DEM elevation values to slopes using the ArcInfo *slope*(*grid*) command and reclassified them according to BMP recommendations into slight (0 -20%), moderate (2140%), and steep (40%) (GFC 1999). We generated a mask for all water by first buffering the hydrology data set. We buffered streams and rivers classified as perennial stream types and those streams classified as trout streams in NARSAL's trout data set by 4.6 m on each side. The process created a 9.2-m primary stream/river mask. Then we buffered the intermittent streams by 2.3 m on each side, creating a 4.6-m secondary stream mask. We vectorized the bodies of water contained in the GAP data set (all

cells classified as ”open water” or ”coastal marsh”) and merged them with the stream/river water masks to create the total water-mask polygon dataset. We considered all areas within this mask to be water.

We vectorized the slope data and incorporated it into the water masks by intersection. Then we buffered the intersected water mask data set, excluding the area existing in the mask, following the criteria in Table 3.1. Buffers were stored in the GIS software (ArcView and ArcInfo) as a separate information layer (Fig. 3.2A-C).

Table 3.1: Water buffer widths according to Georgia's current Best Management Practices.

Stream type	Slope	Prim. Buffer (m)	Sec. Buffer (m)
Trout	All	30.5	30.5
Perennial	Slight (<=20%)	12.2	30.5
	Moderate (21%-40%	21.3	30.5
	Steep (>40%)	30.5	30.5
Intermittent	Slight (<=20%)	6.1	15.2
	Moderate (21%-40%	10.7	15.2
	Steep (>40%)	15.2	15.2

In the beginning, we wanted to use the GA transportation development data available from the Georgia GIS Clearinghouse and assume a certain width of each road according to road classification. However, to maintain consistency with the land cover stratifications, we used GAP as the main (basic) source of information about roads. We extracted a transportation class from the GAP data set and stored it as a separate information layer. To make sure that all major roads were included in our analyses, we compared the resulting road class with the road development data for the state. This revealed that the transportation class in the GAP data closely corresponded with the publicly available data, so we chose to use the GAP transportation class as our sole source of road information. We defined primary and secondary road buffer widths of 12.2 and 30.5 m, respectively.

3.3.5 Data intersection

The GAP land cover classes and ETM+ -generated polygon attributes were incorporated into the hydrology and transportation buffers by intersection (Fig. 3.2D-F). This process produced data

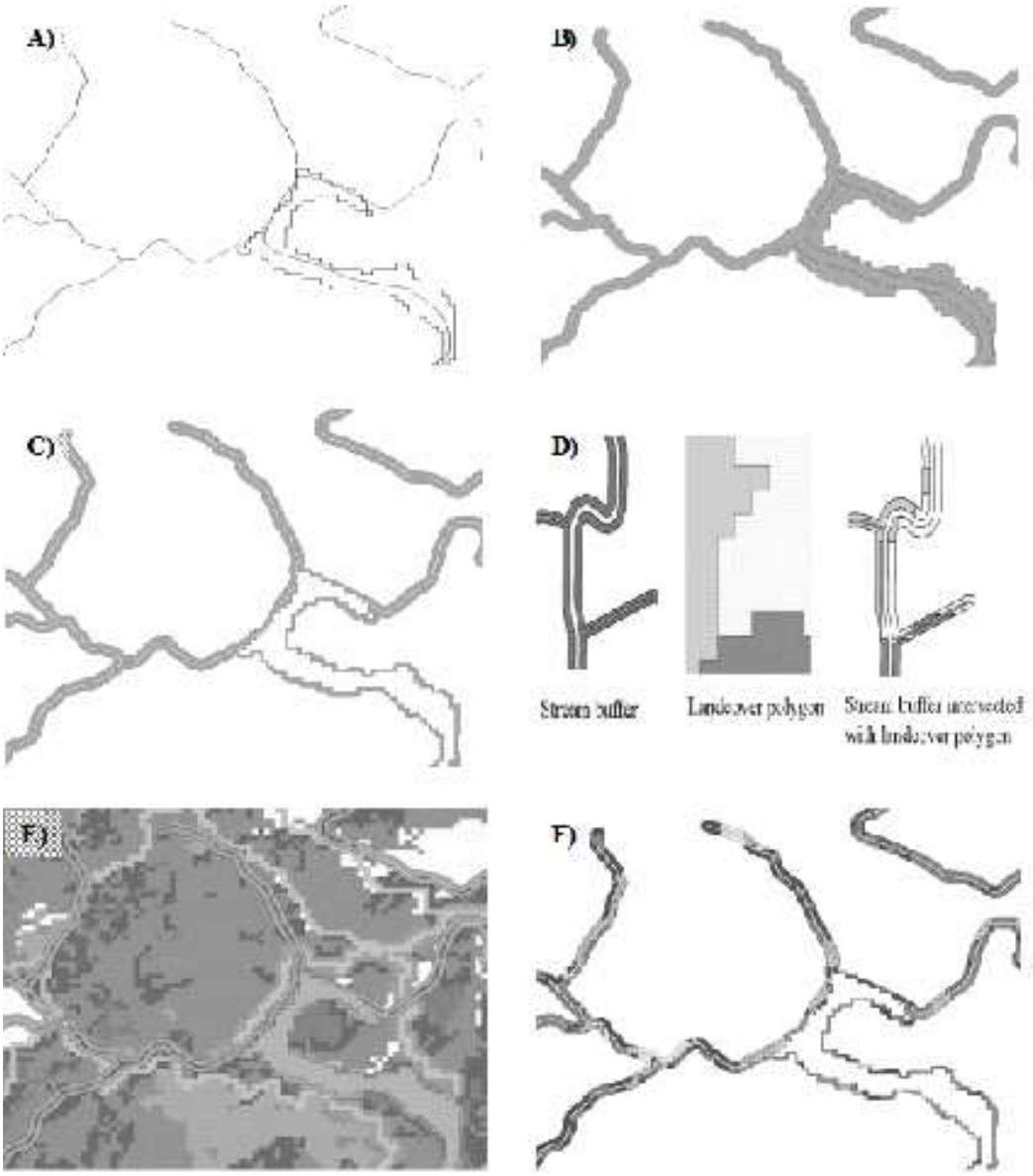

Figure 3.2: Buffer creation and intersection: (A) merged vector hydrology and raster GAP data sets; (B) buffered with respect to water body class and elevation data; (C) water masked out. Resulting buffers were intersected with the polygon data set containing distributed FIA information (D,E), giving final buffer data set with incorporated forest inventory data (F).

sets similar to those created when the hydrology masks were intersected with the reclassified slope data. During the intersection process, all attributes contained in both data sets were maintained. Table 3.2 contains the complete list of the 12 analyzed intersections.

3.3.6 Buffer area and volume calculation

We calculated the total sum of forested area inside the buffers for all regimes. We also calculated the area of various forest classes for each scenario. Division into classes depends on the regime. In the scenario based on the GAP data, we were able to calculate detailed area distribution by GAP classes; we have area of deciduous, evergreen, and mixed forests; forested wetlands; and clearcut/sparse. Because in the ETM+ -based approach we used the forest inventory data, we were able to calculate detailed area statistics by forest types, which were defined based on the FIA data. Using this approach, we can also calculate detailed statistics for any combination of attributes included in standard FIA inventory reports (e.g., Thompson 1998).

We combined the area summaries with the volume data. For each broad forest cover type in the GAP-based scenarios, we calculated average volume per area unit at the state level using the FIA database (Hansen et al. 1992, Miles et al. 2001) and multiplied it by the number of hectares in a given class. Areas and volumes were compared with the latest results of the FIA program (Thompson 1998). For the Landsat-based regimes, we calculated average volume per hectare of distinguished forest types at the county level. The total volume within buffers was calculated as a sum of volumes of various forest types in all counties.

3.4 Results and Discussion

The derived results showed consistently higher estimates of the areas and volumes in the considered buffers than the estimates reported by Zasada et al. (2003). The models based on the Euclidean spectral distances from the high and low basal area signatures to all the other considered cells had a high $R2$, varying between the scenes from 80% to 82% for the ETM+ scenes, rows 37 and 38 in path 19 (Tab. 3.3; Fig. 3.3). The $R2$ values associated with the fittings of the models for

Table 3.2: Intersection data set description.

Dataset name	Description
Primary hydro/GAP	Primary (BMP-based) hydrology buffer intersected with land cover containing GAP cover-class information
Primary hydro/proximity	Primary (BMP-based) hydrology buffer intersected with land cover containing FIA information assigned using the proximity criteria
Primary hydro/rank	Primary (BMP-based) hydrology buffer intersected with land cover containing FIA information assigned using the ranking criteria
Secondary hydro/GAP	Secondary (15.2 meters for intermittent / 30.5 meters for perennial streams) hydrology buffer intersected with land cover containing GAP cover-class information
Secondary hydro/proximity	Secondary (15.2 meters for intermittent / 30.5 meters for perennial streams) hydrology hydroloy buffer intersected with land cover containing FIA information assigned using the proximity criteria
Secondary hydro/rank	Secondary (15.2 meters for intermittent / 30.5 meters for perennial streams) hydrology buffer intersected with land cover containing FIA information assigned using the ranking criteria
Primary road/GAP	12.2-meter road buffer intersected with land cover containing GAP cover class information
Primary road/proximity	12.2-meter road buffers intersected with land cover containing FIA information assigned using the proximity criteria
Primary road/rank	12.2-meter road buffers intersected with land cover containing FIA information assigned using the ranking criteria
Secondary road/GAP	30.5-meter road buffer intersected with land cover containing GAP cover class information
Secondary road/proximity	30.5-meter road buffers intersected with land cover containing FIA information assigned using the proximity criteria
Secondary road/rank	30.5-meter road buffers intersected with land cover containing FIA information assigned using the ranking criteria

basal areas on the other adjacent scenes varied from 80% to 93%. Table 3.4 summarizes the results of the regression models applied to all the adjacent scenes. The high $R2$ values initially seemed surprisingly high, but it soon became apparent that they were mainly a result of pixel averaging, which in turn made the final models much less accurate for estimating basal area for the remaining pixels. The results obtained from all three described methods of water and road assessment are very similar (Tabs. 5 and 6). This was not a surprise since we used the GAP data set to stratify Landsat images and eliminate nonforest areas. However, the territories occupied by stream and road buffers determined by different scenarios were not identical because the GAP forested area was larger than the FIA forested area, and therefore, the different scenarios associated different areas with unassigned forest cover types. Volumes of the forests in the buffers also show very close values, even though they were calculated using completely different approaches. The differences in the results from different methods may likely be attributed to additional or more detailed information. The scenario based on the exclusive use of GAP data resulted in the maximum SMZ areas and their volumes by broad GAP cover types. In this scenario the stream buffers established according to Georgia's BMPs occupy almost 396,460 ha, which makes up 4.01% of the total forested area of the state. Assuming all buffer widths of 15.2 m for intermittent and 30.5 m for perennial streams, we estimated that 8.65% of forested land would be occupied by the buffers. Forests in the determined buffers maintain 4.32% and 9.27% of Georgia's total inventoried volume (Tabs. 3.5 and 3.6).

Table 3.3: Basal area-spectral distance regression results, model parameters (a, b, c), standard errors of regression (SE) and adjusted R2, for the fitted Hossfeld-based (Y a/ 1 b/(X d)c) and Richards-based (Y a 1 exp(bX) c) models.

ETM+ scene	Base Model	a	b	c	d	N	SE	R2
Path 19, Row 37	Hossfeld	378.327	10.394	0.722	13.516	14	19.2	0.82
Path 19, Row 38	Richards	162.808	0.273	71.849	N/A	15	21.11	0.8

Area estimates differ for the various methods used in this study because the forested area defined by GAP is greater than the forested area estimated by FIA. In addition, buffer area differs due to the method by which the FIA information was distributed. In many instances at the county level, GAP area estimates for evergreen, deciduous, and mixed cover types exceed those

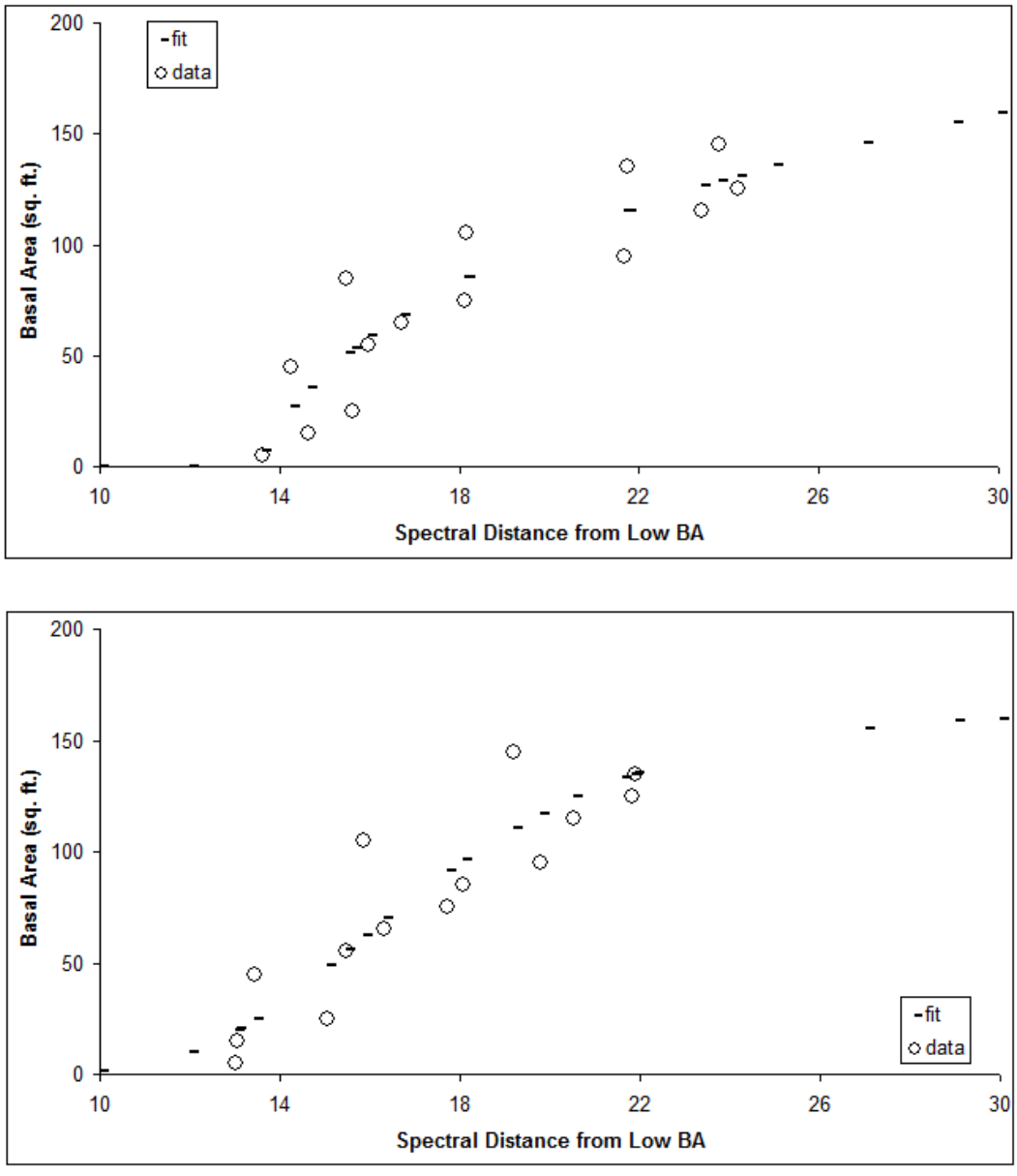

Figure 3.3: Predicted basal area as a function of euclidean spectral distance from a "low basal area" signature for Enhanced Thematic Mapper Plus scene. (A): Hossfeld model (Basal_Area = a/(1 + b/(E_Distance –c)^d) fitted to data from scene in path 18, row 37; (B): Richards model (Basal_Area = a(1 – exp(-b[E_Distance^c])^d) fitted to data from scene of path 19, row 38. BA, basal area.

Table 3.4: Adjacent bsal area spectral distance regresion results.

ETM+ scene	Parent scene	Adjusted R^2	Standard error
Path 19, Row 36	Path 19, Row 37	0.82	19.2
Path 18, Row 36	Path 19, Row 36	0.8	21.11
Path 18, Row 37	Path 19, Row 37	0.88	22.49
Path 18, Row 38	Path 19, Row 38	0.89	16.23
Path 17, Row 37	Path 18, Row 37	0.93	10.67
Path 17, Row 38	Path 18, Row 38	0.86	13.01

from FIA. In these cases, not all GAP-stratified ETM+ -generated polygons were populated with FIA information. Those polygons not populated with FIA information are not represented in the buffer summary data sets (Fig. 3.4).

Results obtained for the scenario based on current BMP are smaller than the numbers coming from studies by Cubbage and Woodman (1993) (4.8% and 5.3%), which might be a result of the authors of the cited study using different assumptions about buffer widths and a different (a spatial) approach. The area of the buffers in primary SMZ zones (7.09%) seems to correspond with our results from the scenario assuming all buffers having 15.2/30.5-m width.

The area of potential road beautifying strips is very similar to the area of stream buffers. The average volume of stands inside the road buffers is lower than those in the stream buffers, which makes sense considering that the riparian forests grow on better sites and contain, in general, older stands.

This study demonstrates that the potential stream management zones and the road beautifying buffers could occupy a significant portion of Georgia forests. The area in the case of the enhanced widths of these buffers may exceed 15% of total forested area in the state, which is a larger impact than generally expected. Average volume per hectare in these areas is higher for water buffers than for road buffers, which is expected since site quality typically increases with proximity to water and reduced elevation. The dominating presence of deciduous forests in water buffers is also

Table 3.5: Comparison of results for primary (BMP-based) and secondary (15.2/30.5-m-wide) water buffer assessment methods.

		GAP		Proximity		Rank	
	Buffer	Area	Volume	Area	Volume	Area	Volume
Forest	Regime	[ha]	[x10$\hat{3}$m3[ha]	[ha]	[x10$\hat{3}$m3[ha]	[ha]	[x10$\hat{3}$m3[ha]
Evergreen	Primary	91,506(0.9%)	7,704(0.8%)	115,458(1.2%)	11,150(1.2%)	105,496(1.1%)	9,424(1.0%)
	Secondary	212,166(2.1%)	17,863(1.9%)	257,975(2.6%)	23,772(2.5%)	238,729(2.4%)	19,980(2.1%)
Mixed	Primary	56,962(0.6%)	4,704(0.5%)	51,174(0.5%)	4,811(0.5%)	59,594(0.6%)	5,434(0.6%)
	Secondary	117,817(1.2%)	9,730(1.0%)	106,260(1.1%)	9,424(1.0%)	126,080(1.3%)	11,094(1.2%)
Deciduous	Primary	247,992(2.5%)	28,718(3.0%)	186,364(1.9%)	21,451(2.3%)	203,229(2.1%)	23,206(2.4%)
	Secondary	524,609(5.3%)	60,750(6.4%)	3963,13(4.0%)	44,997(4.7%)	432,338(4.4%)	48,478(5.1%)
Total	Primary	396,460(4.0%)	41,126(4.3%)	352,996(3.6%)	37,413(3.9%)	368,319(3.7%)	38,064(4.0%)
Timber	Secondary	854,592(8.6%)	88,343(9.3%)	760,549(7.7%)	78,193(8.2%)	797,074(8.1%)	79,551(8.4%)
Unallocated	Primary	NA	61,628(11.0%)	44,763(7.1%)	NA	NA	NA
Area	Secondary	NA	128,296(11.0%)	92,272(6.7%)	NA	NA	NA

Table 3.6: Comparison of results for primary (12.2 m) and secondary (30.5 m) road buffer assessment methods.

		GAP		Proximity		Rank	
	Buffer	Area	Volume	Area	Volume	Area	Volume
Forest	Regime	[ha]	[x10$\hat{3}$m3[ha]	[ha]	[x10$\hat{3}$m3[ha]	[ha]	[x10$\hat{3}$m3[ha]
Evergreen	Primary	162,189(1.6%)	13,650(1.4%)	161,124(1.6%)	14,688(1.5%)	173,698(1.8%)	13,046(1.4%)
	Secondary	390,104(3.9%)	32,832(3.4%)	382,904(3.9%)	34,639(3.6%)	414,614(4.2%)	31,470(3.3%)
Mixed	Primary	90,866(0.9%)	7,507(0.8%)	53,334(0.5%)	4,613(0.5%)	56,121(0.6%)	5,462(0.6%)
	Secondary	210,708(2.1%)	17,407(1.8%)	125,898(1.3%)	10,924(1.1%)	131,352(1.3%)	12,848(1.3%)
Deciduous	Primary	106,696(1.1%)	12,356(1.3%)	105,653(1.1%)	12,395(1.3%)	111,273(1.1%)	12,197(1.3%)
	Secondary	256,995(2.6%)	29,761(3.1%)	253,900(2.6%)	29,913(3.1%)	267,169(2.7%)	29,432(3.1%)
Total	Primary	359,752(3.6%)	33,513(3.5%)	320,112(3.2%)	31,696(3.3%)	341,092(3.5%)	30,706(3.2%)
Timber	Secondary	857,807(8.7%)	80,000(8.4%)	762,702(7.7%)	75,476(7.9%)	813,135(8.2%)	73,750(7.7%)
Unallocated	Primary	NA	39,640(11.0%)	18,660(5.2%)	NA	NA	NA
Area	Secondary	NA	95,105(11.1%)	44,672(5.2%)	NA	NA	NA

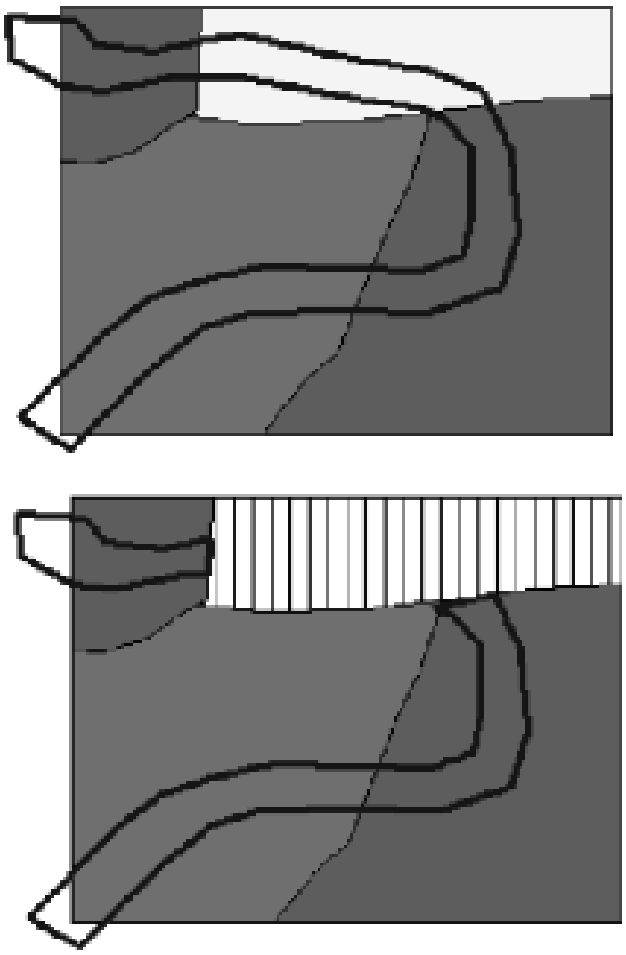

Figure 3.4: Differences in the buffer area estimates by the two methods applied in this study, which use different models for distributing the FIA data on the GAP-based area that is greater than the FIA defined area. Scenario 1 (top image), all four polygons were distributed FIA information. Scenario 2 (bottom image), the upper-right polygon was not assigned any FIA information, and therefore, its area was not incorporated into the final acreage tally.

expected due to the high productivity of the bottomland hardwoods and forested wetlands. On the other hand, it is not clear why evergreen forests dominate in road buffers.

3.4.1 Policy and economic implications

SMZs and road beautifying buffers (RBBs) will likely have some impacts on timber production returns. Their magnitude will depend on the scale at which SMZs and RBBs are implemented and on local site conditions. Setting aside protection zones will at least delay harvest, which may involve some immediate costs such as lost income from timber sales. However, as long as partial timber harvest is permitted, delaying harvest may also provide some market benefits as landowners diversify their wood production and let portions of their trees grow to larger sizes and more valuable product classes. Furthermore, at least in principle, setting aside protection zones will increase management costs. However, areas containing ?owing streams in gullies and ravines are often of marginal value for timber production, and timber management there would make little sense under any conditions. In those cases, increases in management costs would be small, especially after the protection zones are initially laid out. In cases where protection is extended to wider areas, including intermediate streams, management costs could increase quite substantially.

A recent survey of the implementation of forestry BMP in the southern region (Southern Group of State Foresters 2008) indicates relatively high implementation levels. For all southern states, SMZ implementation rates ranged from 76% to 99% over the period from 1997 through 2007. The most recent round of surveys indicates an average implementation rate of 88%. In cases where multiple surveys were available, positive trends were noted in SMZ and stream crossing implementation rates. The relatively high rates of voluntary compliance indicate that current SMZ requirements are reasonable and offer a cost-effective way to protect our water resources. The case is much less established for RBBs, in terms of both management costs and purported environmental benefits.

3.4.2 Using incompatible data to improve high-resolution inventory information

It is a known statistical fact that data from different sources with differing statistical and sampling characteristics cannot be pooled together for the purpose of deriving a new, unbiased estimator. This fact is the primary factor contributing to the scarcity and high cost of natural resource information. Although there is abundant information from various sources, such as atmospheric, soil, geographic, remote sensing, and a multitude of various independent special studies sampling data for different variables, the costs of forest inventories remain nearly as high as they would be in the complete absence of any auxiliary information. Most of the research on improving natural resource assessment concentrates on finding new measurement methods and technologies. A typical tactic for increasing inventory accuracy is to increase sample and plot size. The study described represents an unorthodox approach in this context in that it concentrates its efforts on increasing the use of various available data and addressing the problems caused by their incompatibilities according to the methodology described in Cieszewski et al. (2005) and Iles (2009).

Using the proposed methodology of combining information from various sources of satellite-based images with FIA data, we are able to provide more detailed high-resolution inventory statistics, which the large-area FIA inventory is not designed or intended to provide. In this study we demonstrated that the use of a variety of data sources with differing statistical and sampling characteristics can be combined to derive improved and more extensive results as compared with the alternative of using only similarly designed but much more limited data sources. The approach presented is not an alternative to the FIA inventory but rather an extension or supplement, which can offer much added value to the large area survey. The approach presented uses information pooled from various incompatible data sources, which is then conditioned to have a mean equal to another, unbiased mean. This approach to forest inventory was proposed by Kim Iles and first implemented in Iles' design of the British Columbia forest inventory (Iles 2009). Subsequently, this idea was implemented by a number of forest industry members and by the Fiber Supply Assessment program (Cieszewski et al. 2004a) at the Warnell School of Forestry and Natural Resources,

University of Georgia and used in this and various other research projects (Cieszewski et al. 2005, Meng and Cieszewski 2006, Meng et al. 2009) in the Fiber Supply Assessment program.

3.4.3 Applicability and potential use of the results

The results presented here provide a reasonable measure of the area and timber volume contained within Georgia's riparian zones. Such information will be helpful to forest managers in improving the effectiveness of their management planning. The fact that all the three approaches applied in this study produced similar results, which were also quite similar to the results reported by Zasada et al. (2003), even though the latter seemed to have a slight negative bias, suggests that the problem is robust and results should be dependable. This does not mean that there is no difference between using different methodologies. The methods used in this study were the most detailed and produced slightly but consistently higher estimates than the cruder methods used by Zasada et al. (2003), which suggests that the latter methods may have a consistent bias. Ultimately, a recommendation for follow-up research is that when suitable data or funding becomes available, efforts should be directed toward assessing the accuracy of the described methods and their potential biases.

3.5 REFERENCES

[1] Natural Resource Spatial Analysis Laboratory. 1998. *1998 land cover map of Georgia.* Natural Resource Spatial Analysis Laboratory, Institute of Ecology, University of Georgia. Available online at gis1.state.ga.us/data/landcov/1998landcover.html; last accessed April 2, 2009.

[2] Cieszewski, C.J., M. Zasada, R.C.B.E. Borders, R.C. Lowe, J.M.L. Clutter, and R.F. Daniels. 2003a. *Various aspects of sustainability analysis in Georgia.* P. 180 187 in Proc. of the Third annual Forest Inventory and Analysis symposium, October 1719, 2001, Traverse City, MI. McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and J.W. Moser (Eds.). US For. Serv. Gen. Tech. Rep. NC-230. US For. Serv. North Central Res. Stn., St. Paul, MN. 208 p.

[3] Cieszewski, C.J., M. Zasada, R.C. Lowe, B. Borders, M.L. Clutter, R.F. Daniels, R.I. ELLE, R. IZLAR, and J. Zawadzki. 2003b. *Analysis of pooled FIA and remote sensing data for fiber supply assessment at the Warnell School of Forest Resources at the University of GeorgiaOther studies and effective information dissemination.* P. 7179 in Proc. of the Fourth Annual Forest Inventory and Analysis Symposium, November 1921, 2002, New Orleans, LA, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and C.J. Cieszewski (eds.). US For. Serv. Gen. Tech. Rep. NC-252. US For. Serv. North Central Res. Stn., St. Paul, MN.

[4] Cieszewski, C.J., K. Iles, R.C. Lowe, and M. Zasada. 2004a. *Proof of concept for an approach to a finer resolution inventory.* P. 69 74 inProc. of the Fifth Annual Forest Inventory and Analysis Symposium. November 1820, 2003, New Orleans, LA,McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and C.J. Cieszewski (eds.). US For. Serv. Gen. Tech. Rep. WO-69. US For. Serv., Washington, DC.

[5] Cieszewski, C.J., M. Zasada, B.E. Borders, R.C. Lowe, J. Zawadzki, M.L. Clutter, and R.F. Daniels.2004b. *Spatially explicit sustainability analysis of long-term fber supply in Georgia, USA*. For. Ecol. Manag. 187(23):345359.

[6] Cieszewski, C.J., K. Iles, R.C. Lowe, and M. Zasada. 2005. *Proof of concept for an approach to a finer resolution inventory.* In 2003 Proc. of the Fifth Annual Forest Inventory and Analysis Symposium, McRoberts, R.E., Reams, G.A., Van Deusen, P.C., and McWilliams, W.H. (eds.). November 18 20, 2003, New Orleans, LA. Gen. Tech. Rep. WO-69. US For. Serv., Washington, DC. 222 p.

[7] Cubbage, F.W., and J.N. Woodman. 1993. *Impacts of streamside management zones on timber supply in Georgia.* P. 130 137 in Policy and forestry: Design, evaluation, and spillovers. Southern Forest Economics Workshop. Wear, D.N. (ed.). Duke University, Durham, NC.

[8] ERDAS. 2002. Imagine, version 8.4. Erdas, Inc., Athens, GA.

[9] Forest Inventory and Analysis (FIA). 2006. *Tools and data*. Available online at www.ncrs.fs.fed.us/4801/tools-data/data; last accessed April 2, 2009.

[10] Gadow, K., C. ZHANG, and X. Zhao. 2009. *Science-based forest design*. Math. Comput. For. Nat. Resour. Sci. 1(1):14 25. Available online at mcfns.com/index.php/Journal/article/view/MCFNS-1:14/46; last accessed Apr. 13, 2009.

[11] Georgia Forestry Commission (GFC). *1999.Georgia's Best Management Practices for forestry*. Georgia Forestry Commission, Macon, GA.

[12] Georgia GIS Clearinghouse. 2009. Georgia GIS Data Clearinghouse. Available online at gis1.state.ga.us; last accessed April 2, 2009.

[13] Hansen, M.H., T. Frieswyk, J.F. Glover, and J.F. Kelly. 1992. *The Eastwide forest inventory database: Users manual.* US For. Serv. Gen. Tech. Rep. NC-151. US For. Serv. North Central For. Exp. Stn., St. Paul, MN. 48 p.

[14] Iles, K. 2009. *Total-Balancing an inventory: A method for unbiased inventories using highly biased non-sample data at variable scales.* Math. Comput. For. Nat. Resour. Sci. 1(1):10 13. Available online at mcfns.com/index.php/Journal/ article/view/MCFNS-1:10/18; last accessed April 13, 2009.

[15] LIU, S., C.J. Cieszewski, R.C. Lowe, and M. Zasada. 2009. *Sensitivity analysis on long-term fber supply simulations in Georgia.* South. J. Appl. For. 33(2):8190.

[16] Lowe, T., C.J. Cieszewski, M. Zasada, and J. Zawadzki. 2003. *Distributing FIA inventory information onto segmented Landsat Thematic Mapper images stratifed with industrial ground data.* P. 111116 in Proc. of the Fourth Annual Forest Inventory and Analysis Symposium, November 1921, 2002, New Orleans, LA, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and C.J. Cieszewski (eds.). US For. Serv. Gen. Tech. Rep. NC-252. US For. Serv. North Central Res. Stn., St. Paul, MN.

[17] Meng, Q., and C.J. Cieszewski. 2006. *Using geo-spatial methods for derivation of one resolution forest inventory in Georgia from ground inventory data and Landsat imagery.* InProc. of the Fifth Southern Forestry and Natural Resources GIS Conference, June 1214, 2006, Asheville, NC. Prisley, S., P. Bettinger, I.-K. Hung, and J. Kushla (eds.). Athens, GA, Warnell School of Forestry and Natural Resources, University of Georgia.

[18] Meng, Q., C.J. Cieszewski, and M. Madden. 2009. *Large area forest inventory using Landsat ETM+: A geostatistical approach.* ISPRS J. Photogram. Remote Sensing 64:2736. Available online at dx.doi.org/10.1016/j.isprsjprs.2008. 06.006; last accessed April 2, 2009.

[19] Miles, P.D., G.J. Brand, C.L. Alerich, L.F. Bednar, S.W. Woundenberg, J.F. Glover, and E.N. Ezzell. 2001. *The Forest Inventory and Analysis database: Database description and user's manual version 1.0, Revision 8 (June 1, 2001).* US For. Serv. Gen. Tech. Rep. NC-218. US For. Serv. North Central For. Exp. Stn., St. Paul, MN.

[20] Ruefenacht, B., D. Vanderzanden, M. Morrison, and M. Golden. 2002. *New technique for segmenting images. RSAC program manual.* US For. Serv., Remote Sensing Application Center, Salt Lake City, UT. 14 p.

[21] Sedell, J., M. Sharpe, D. Dravnieks Apple, M. Copenhagen, and M. Furniss. 2000. *Water & the Forest Service.* FS-660. US For. Serv., Washington, DC. 26 p.

[22] Southern Group of State Foresters. 2008. *Implementation of forestry best management practices: A southern region report.* Prepared by Water Resource Committee for Southern Group of State Foresters, Winder, GA. 17 p.append.

[23] Thompson, M.T. 1998. *Forest statistics for Georgia, 1997.* Resour. Bull. SRS-36. US For. Serv. Southern Res. Stn., Asheville, NC. 92 p.

[24] Van Sickle, C. 2001. *Southern Appalachian case study.* P. 472 488 in A guidebook for integrated ecological assessments, Jensen, M.E. and P.S. Bourgeron, (eds.). Springer-Verlag, New York.

[25] Wear, D.N., and J.G. Greis. 2002. *Southern forest resource assessment.* US For. Serv.Gen. Tech. Rep. SRS-53. US For. Serv. Southern Res. Stn., Asheville, NC.635 p.

[26] Welsch, D.J. 1991. *Riparian forest buffers; function and design for protection andenhancement of water resources.* US For. Serv. Northeastern Area, state, andprivate forestry, NA-PR-07-91, Radnor, PA.

[27] Wenger, S.J. 1999. *A review of the scientifc literature on riparian buffer width, extent and vegetation.* Public Service & Outreach, Institute of Ecology, University of Georgia, Athens, GA. 59 p.

[28] Zasada, M., C.J. Cieszewski, R.C. Lowe, J. Zawadzki, M. Clutter, and J.P. SIRY. 2003. *Using FIA and GIS data to estimate areas and volumes of potential stream management zones and road beautifying buffers.* P. 253257 in Proc. of the Fourth Annual Forest Inventory and Analysis Symposium, November 1921, 2002, New Orleans, LA, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and C.J. Cieszewski (eds.). US For. Serv. Gen. Tech. Rep. NC-252. US For. Serv. North Central Res. Stn., St. Paul, MN.

[29] Zasada, M., C.J. Cieszewski,and R. C. Lowe. 2004. *Impact of stream management zones and road beautifying buffers on long-term fber supply in Georgia.* P. 41 47 in Proc. of the Fifth Annual Forest Inventory and Analysis Symposium, November 1820, 2003, New Orleans, LA, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and W.H. McWilliams (eds.). US For. Serv. Gen. Tech. Rep. WO-69. US For. Serv., Washington, DC.

CHAPTER 4

MEAN BALANCED, HIGH RESOLUTION FOREST INVENTORY OF GEORGIA

4.1 INTRODUCTION

Under a multi-use sustainable management regime, the provision of timely, reliable, and accurate information about forests, their forested ecosystems, and adjacent areas is essential for maintaining their ecological balance and sustained productivity. This is especially important where forests tend to be fast growing and changing, highly fragmented in area and ownership, and the demand for their wood products is high, such as those in Georgia and other southeastern states. However great the need though, there is a lack of detailed stand-level information for large portions of this region.

The United States Forest Service Forest Inventory and Analysis (FIA) program collects forest information and produces regular reports on the condition of forests throughout the country. In Georgia, the FIA data is used in various large area inventory based analysis ranging from carbon studies to tree mortality analysis ((Van Deusen 2010; Meng and Cieszewski 2006). The inventory provides reliable, unbiased estimates suitable for reporting across large areas (Blackard et al. 2008; Walker et al. 2007; Wang, Lu, and Haithcoat 2007; Sivanpillai et al. 2005; Wayman et al. 2000; Chojnacky 1998). However, since the large-area FIA inventories are not suitable for applications to smaller areas, there is still a compelling need for higher-resolution forest information. A more suitable source for this information is compiled by the local agencies familiar with those areas whose intimate knowledge is needed for their management. Forest product industry and other large area forest owners typically maintain their own private inventories that are more detail oriented and suitable for small area, stand level, forest management.

Nearest neighbor methods are an established means to generate estimates of forest volume (Trotter, Dymond, and Goulding 1997; Kajisa et al. 2008; Franco-Lopez, Ek, and Bauer 2001; McRoberts 2012), basal area (McRoberts et al. 2007; Holmgren et al. 2000; McRoberts 2008;

Meng, Cieszewski, and Madden 2009; Sivanpillai et al. 2006), biomass (Gjertsen 2007; Tomppo et al. 2008; Reese et al. 2010), and carbon (R. McRoberts, Tomppo, and Nsset 2010; Labrecque et al. 2006; Blackard et al. 2008; Fournier et al. 2003), to only name a few. This method's popularity, for some part, stems from its intuitive implementation, the ability to simultaneously generate estimates for multiple variables using the same parameters usually the number of nearest neighbors K, and the ability to make use of noisy data for prediction. However, as Iles (2010) demonstrates and others have reported in their results (R. E. McRoberts 2008; Trotter, Dymond, and Goulding 1997; McRoberts and Tomppo 2007; Hilker et al. 2009; Magnussen, Tomppo, and McRoberts 2010), the nearest neighbor methods are inherently biased, thus the risk cannot be ignored.

The total-balancing concept proposed by Iles (2009) is the foundation of our approach to addressing the issue of bias in our forest inventory for the state of Georgia. In this approach, the large-area FIA information and local forest inventories are used in tandem to develop a spatially explicit inventory that maintains the large-area unbiased properties of the FIA and the local precision of the forest industry inventories, even though they are traditionally viewed as having incompatible variances.

Objectives

The purpose of this research is twofold. First, I generate a broad area, high resolution, spatially explicit inventory for Georgia that is balanced to an unbiased mean volume per acre derived from the FIA. Second, I demonstrate the potential gains in local precision we can obtain by fusing local inventory information with the explicit inventory while maintaining overall balancing.

4.2 DATA

4.2.1 Study area

The study area for this research is the state of Georgia (Fig. 4.1). As a whole, the state of Georgia is a typical southern state with vast natural resources. It has more than 24 million acres of forestland, of which approximately 45% are conifer, 42% are deciduous, 12% are a mixed forest type and the

Figure 4.1: The 12 Landsat WRS2 scenes covering the state of Georgia.

remaining classified as non-stocked. Adding to the complexity of the landscape, an approximate 650,000 non-industrial landowners hold seventy-five percent of these acres whose average parcel size is decreasing (Georgia Forestry Commission 2008).

There are distinct differences in the forest composition moving from north to south in the state. The northern parts are dominated by hardwood species (Fig. 4.2A) (Tab. 4.1) . The forests transitions to a conifer-dominated ecosystem to as you proceed southward and to the east (Fig. 4.2B) (Tab. 4.1).

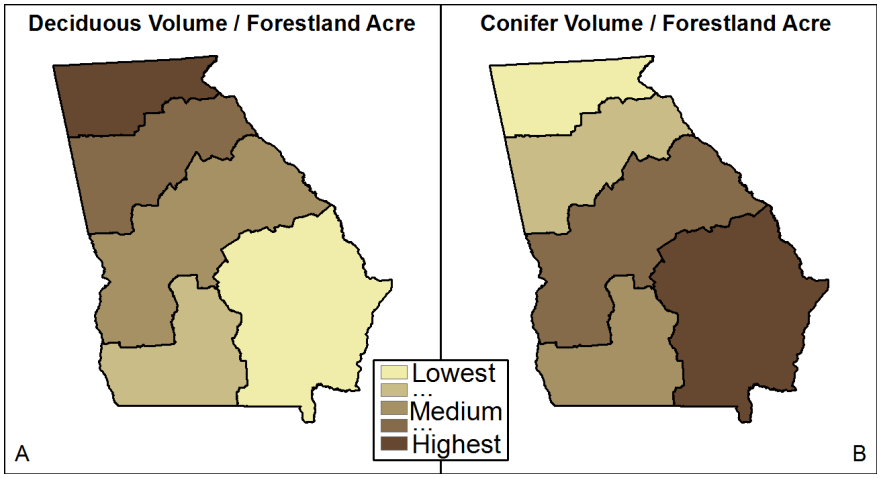

Figure 4.2: Total volume summarized by FIA regions for the A) coniferous forestland, and B) deciduous forestland for the state of Georgia.

Table 4.1: Forestland area and total volume summarized by FIA regions reported by the FIA.

	Conifer Forest		Mixed Forest		Deciduous Forest	
FIA Region	Forestland (1000 Ac)	Volume (Mil. ft^3)	Forestland (1000 Ac)	Volume (Mil. ft^3)	Forestland (1000 Ac)	(Mil. ft^3)
Northern	564	1,202	467	1,075	1,933	4,401
North Central	1,118	2,252	435	778	1,679	3,717
Central	3,675	5,877	842	1,144	3,067	5,092
Southeastern	4,361	6,327	770	970	2,742	4,147
Southwestern	1,404	2,068	360	452	1,103	1,719

4.2.2 Satellite imagery

The initial cubic-foot per acre estimates are based on Landsat 5 Thematic Mapper satellite imagery. I attempted to attain imagery from the leaf-off season early in the year, leaf-on from the summer months, and another leaf-off image from late in the year, however, due to cloudy conditions, it was not possible for all scenes (Tab. 4.2). I acquired two to four cloud-free images for each of the 12 scenes that wholly or partially cover the state. Two UTM zones, zone 16 and zone 17, overlap the state. To facilitate processing, I created a custom coordinate system definition that shifted UTM zone 17 west 500,000 meters and projected each image to that custom UTM zone.

Table 4.2: Acquisition dates of the Landsat 5 satellite imagery used in the volume estimation processes.

TM	TM 1	TM 2	TM 3	TM 4
Path 17, Row 37	4/11/2010	9/2/2010	12/7/2010	NA
Path 17, Row 38	3/13/2010	6/14/2010	12/7/2010	NA
Path 17, Row 39	6/14/2010	10/4/2010	11/21/2010	NA
Path 18, Row 36	4/18/2010	6/21/2010	11/12/2010	NA
Path 18, Row 37	4/2/2010	10/11/2010	12/14/2010	NA
Path 18, Row 38	1/12/2010	5/20/2010	10/11/2010	12/14/2010
Path 18, Row 39	2/13/2010	10/11/2010	12/14/2010	NA
Path 19, Row 36	11/16/2009	3/24/2010	10/2/2010	11/19/2010
Path 19, Row 37	2/20/2010	7/30/2010	11/19/2010	NA
Path 19, Row 38	2/20/2010	7/30/2010	11/19/2010	NA
Path 19, Row 39	2/20/2010	10/18/2010	11/19/2010	NA
Path 20, Row 36	1/29/2010	10/9/2010	NA	NA

4.2.3 Inventory data

The Forest Service, in an attempt to maintain the integrity of each plot, does not release cruise plot coordinates to the public. However, they do allow their use at one of their secure data centers. I processed the satellite imagery using their field measured GPS locations at the Southern Research Station in Knoxville, TN in December of 2010 I used a series of arcpy (ESRI 2010) scripts to extract the TM band 1 – 5 and 7 pixel values for each FIA cruise locations for all images used in this study (Tab. 4.2).

A fundamental aspect of the FIA's measurement protocol is the fact that measured plots shall not be given preferential treatment by the inventory crew or the public. The landowner is permitted to manage the forest as they see fit. Thus, there is the possibility that the database may contain outdated information about a plot since any changes to the land that occurs after the inventory are not recorded until the next measurement cycle.

Absent of the plot locations outside the FIA's data center, I was unable to perform a visual inspection of the TM data at each plot center. However, I did evaluate the spectral information stored in the training sample list using a series of pseudo-image composites. I used the following steps to generate the pseudo-images for each scene:

1. sort the training samples according to their NDVI (Eq. 4.1) values,
2. reorganize the data into a square grid that stores the information from one TM band,
3. repeat step 2 for each spectral band in the training sample list,
4. import each band into ArcGIS, and generate the pseudo-image using the *Composite Bands* command, and then
5. repeat steps 2 and 4 for the NDVI values.

$$NDVI = \frac{NIR - RED}{NIR + RED}, \tag{4.1}$$

where: NIR is the near-infrared layer (TM band 4), and RED is the red layer (TM band 3) The pseudo-image shown in figure 4.3 displays representative pixels from the actual TM data, yet they are sorted (left to right) by their NDVI values. By visual inspection of the actual TM and NDVI images, and with the aid of the 2010 NAIP aerial photography, I was able to loosely define the cover types represented in sections A – G as seen in figure 4.4 . The pixels in sections A) and B) were captured in areas void of green vegetation such as a cultivated field (Fig. 4.4) or a place inundated with water. Near the other end of the NDVI spectrum, the samples in frame F) (Fig. 4.3) are sites captured in mature forested areas with full canopy closure (Fig. 4.4). The sites in frames C through

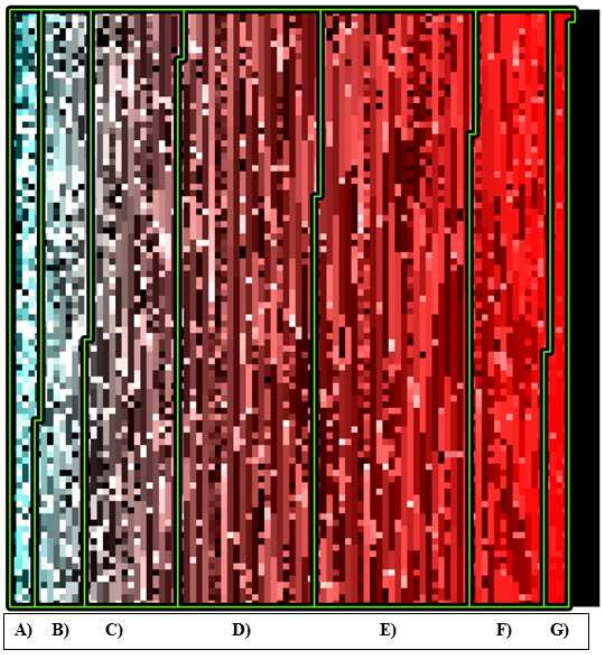

Figure 4.3: Pseudo-Landsat image generated from FIA sample sites representing A) bare ground sites, B-C) the transition to forest, D-F) the transition to a closed canopy forest, and G) cropland.

E contain sample sites that include old fields, young pine plantations, and thinned forests. Frame G contained the samples with the highest NDVI values. These are cropland sites with abundant, low-lying, fast green vegetation. I used the forested/non-forested thresholds determined by this process for each scene to assess which, if any, FIA plots had been harvested between the time a plot was measured and the capture of the late-winter TM image. I assigned those plots a volume per acre (ft^3) of zero.

I used the Forest Vegetation Simulator (Wykoff et al 1992, Dixon 2002) to project the FIA field measurements to a common end-of-year 2010. I implemented the SN variant and accepted the data processing defaults. The projected dataset contains 6,367 total plots that I have classified as either deciduous, mixed, or evergreen according to their predominant species (Tab. 4.3). There were 150 non-stocked and 2,122 non-forested plots within the state that were not used in the analysis.

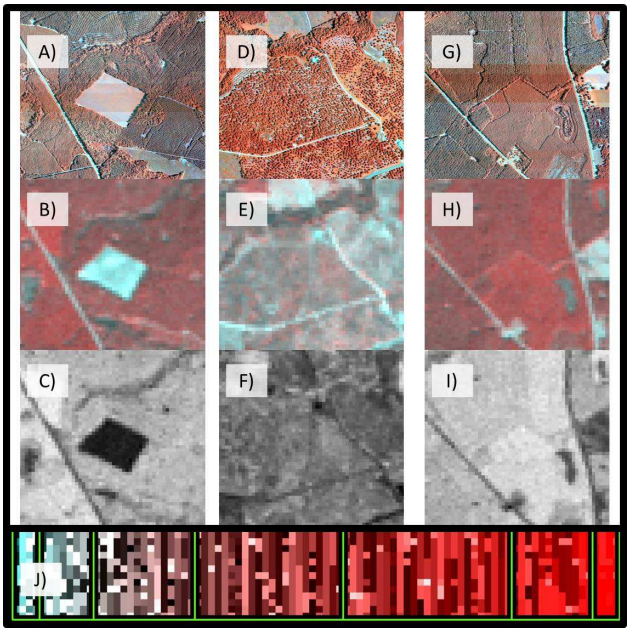

Figure 4.4: Visually assessed A-C) non-forest, D-F) sparse forest, and G-I) closed canopy sites as they relate to samples in the J) pseudo-Landsat image (Fig. 4.3 and how they appear in the 2010 color-infrared NAIP (A, D, G), the winter TM (B, E, H), and the NDVI (C, F, I) images.

Table 4.3: Summary of age, bsal area, and cubic-foot volume per acre for all FIA ground measurements

		Hardwood	Mix	Pine	Non-stocked	No Forest
	# of Stands	1,628	634	1,833	150	2,122
Age	Mean	48	38	27	3	0
	St. Dev.	30	24	17	2	0
	Min.	0	1	0	0	0
	Max.	149	162	115	5	0
Basal Area	Mean	98	95	95	0	0
	St. Dev.	53	45	46	0	0
	Min.	0	0	0	0	0
	Max.	412	263	430	0	0
Vol / Acre (cu.ft.)	Mean	2,060	1,789	1,668	0	0
	St. Dev.	1,569	1,284	1,189	0	0
	Min.	0	0	0	0	0
	Max.	11,585	6,232	8,882	0	0

I stratified the plots further according to which WRS2 scene it was located. Since there is overlap among the scenes in both a north-south and east-west direction some plots were used multiple times in different scene-level mean calculations (Tab. 4.4). These are the mean cubic feet volume per acre values that I used in the scene-level scaling process to generate the estimates.

These compiled data are the *training* samples used in the volume estimation process. The data files, one for each TM scene, includes FIA plot age, basal area per acre (BA), cubic-foot volume per acre (CF), county FIPS code, the TM scene identifier, and the TM spectral summaries that were recorded at each plot center.

We obtained 918 conifer forest polygons and associated stand summary information from our various industrial partners with holdings in WRS2 path 18, row 37. I visually inspected each area on the early and late in the year leaf-off TM and on the 2010 USDA Farm Service Agency National Agriculture Imagery Program aerial photography to ensure the data did not include any partially harvested stands. I manually recoded the stand summaries to zero for any stand that reflected a total harvest. Individual stand age, volume and basal area measures were projected forward to a common 2010 end-of-year timeline. The final industrial data set contained 47,469 acres.. Their ages ranged from zero to 61 and average volume per acre was 2,261 ft^3.

4.2.4 Land cover

I used a composited 2008 Land Use Trends Land Cover of Georgia (GLUT) (NARSAL 2006) and NLCD 2006 Land Cover (NLCD) (USDOI 2006) to initially stratify the land base into generic conifer, mixed forest, and deciduous forest types. The composite was created using a raster intersection where

$$GLUT/NLCD = GLUT * 1000 + NLCD. \tag{4.2}$$

This procedure outputs a single raster layer raster whose values represent the input from both data sets. For example, a cell whose output is "31042" represents an area classified by GLUT as *clearcut*, class 31, and classified by NLCD as *evergreen,* class 42. The overlay resulted in almost

Table 4.4: Summary of FIA plot grond measurements for the 12 TM scenes encompassing the state of Georgia.

			Volume / Acre (ft^3)			
Path / Row	Cover Type	# Plots	Mean	St. Dev.	Min.	Max.
17 / 37	Pine	358	1,794	115	0	8,134
	Mix	88	1,937	245	48	5,022
	Hwd	217	1,916	174	0	11,118
17 / 38	Pine	682	1,590	77	0	8,134
	Mix	138	1,486	168	0	6,232
	Hwd	366	1,853	132	0	9,345
17 / 39	Pine	264	1,390	109	0	4,978
	Mix	49	1,385	265	0	4,333
	Hwd	117	1,746	240	0	11,585
18 / 36	Pine	107	2,153	252	0	8,882
	Mix	79	2,427	319	48	5,976
	Hwd	267	2,488	176	0	9,737
18 / 37	Pine	613	1,777	86	0	8,134
	Mix	230	1,798	146	0	6,225
	Hwd	481	1,955	115	0	11,118
18 / 38	Pine	689	1,508	71	0	8,134
	Mix	170	1,449	144	0	6,225
	Hwd	478	1,812	114	0	11,053
18 / 39	Pine	183	1,453	129	0	4,214
	Mix	50	1,436	258	0	3,953
	Hwd	106	1,605	202	0	5,003
19 / 36	Pine	143	1,982	210	0	8,882
	Mix	143	2,334	224	37	5,976
	Hwd	461	2,452	130	0	9,737
19 / 37	Pine	345	1,791	116	0	7,324
	Mix	166	1,808	170	0	5,946
	Hwd	358	2,070	132	0	7,826
19 / 38	Pine	253	1,503	113	0	5,373
	Mix	75	1,294	188	0	4,836
	Hwd	218	1,811	164	0	9,203
19 / 39	Pine	37	1,546	310	0	3,939
	Mix	7	1,634	769	590	2,692
	Hwd	25	1,795	447	0	3,946
20 / 36	Pine	84	1,641	225	0	7,324
	Mix	49	1,993	330	37	3,839
	Hwd	148	2,172	207	0	5,754

200 unique combinations. I reduced the number of classes by reclassifying the cells using the cross-matrix shown in Table 4.5. I assumed the cells classified by GLUT as a clearcut would ultimately result in an evergreen forest. I assigned all cells classified as evergreen by one agency and deciduous by the other to the mixed forest type. The classes not listed in the table (i.e. urban, GLUT cropland/pasture, and water) were used as a non-forest mask.

Table 4.5: Reclassification matrix used to combine the GLUT and NLCD land cover products.

	2008 GLUT					
NLCD 2006	Clear-cut(31)	Decid-uous(41)	Ever-green(42)	Mixed(43)	Forested Wetland(91)	Nonforest Wet.(93)
Deciduous(41)	Ever.	Decid.	Mixed	Mixed	Decid.	NA
Evergreen (42)	Ever.	Mixed	Ever.	Mixed	Ever.	NA
Mixed (43)	Ever.	Mixed	Mixed	Mixed	Mixed	NA
Evergreen (52)	Ever.	Mixed	Ever.	Mixed	Ever.	NA
Clearcut (71)	Ever.	Decid.	Ever.	Mixed	Ever.	NA
Crop (81, 82)	Ever.	Decid.	Ever.	Mixed	NA	NA
Wetland (90)	Ever.	Decid.	Ever.	Mixed	Decid.	Decid.

Statewide, there were a total of 24,079,775 acres of forested land represented in this dataset. Adhering to the above re-classification scheme, 44% were labeled as conifer, 18% as mixed, and 37% assigned to deciduous class (Tab. 4.6).

Table 4.6: Area comparison with the FIA and summary of the various types contained in the combined GLUT/NLCD.

	FIA	GLUT/NLCD	Pct.	GLUT/NLCD Area(ac)		
NAME	Area (ac)	Area (ac)	Diff. (%)	Conifer	Mixed	Decid.
Central	7,583,550	7,307,525	3.6%	3,083,015	1,708,000	2,516,510
North Central	3,231,760	3,428,687	-6.1%	1,230,685	588,812	1,609,190
Northern	2,964,410	3,079,595	-3.9%	435,484	753,301	1,890,810
Southeastern	7,873,220	7,485,480	4.9%	4,553,737	597,463	2,334,280
Southwestern	2,867,720	2,777,584	3.1%	1,357,839	597,729	822,016

4.2.5 Software

I used a variety of commercial software and in-house programs to process the data. Image co-registration, data projection, land cover re-classification, and data cataloging tasks were performed

in ESRI's ArcGIS (ESRI 2010) and ERDAS' Imagine (ERDAS Inc., Norcross, GA). I converted the data layers among common GIS image formats and generic binary formats using the GDAL interfaced with Python (Van Rossum 2008) and Perl (The Perl Foundation). I developed custom programs written with Lahey/Fujitsu LF95 v. 8.1b Fortran compiler to implement the nearest-neighbor processing, data summarization, and image generation.

4.3 METHODS

4.3.1 Initial KNN estimation based on the FIA data

In this study, the volume prediction for a pixel was determined using:

$$\hat{y}_i = \frac{1}{k} \sum_{j=1}^{k} y_j^i, \tag{4.3}$$

where $\hat{y}_i$ is the predicted value for pixel *i* and $\left\{ y_j^i,\ j = 1,\ 2,\ \ldots,\ k \right\}$ are the *k*spectrally nearest response values stored in the training list. This process can be modified using a weighting factor which is commonly based on the distance between pixel *i* and the location of the K neighbor(s). In this study, however, I did not implement a weighting factor since I did not have access to the plot locations outside of the FIA data center.

I determined the number of near neighbors for each scene using leave-one-out cross-validation analyses. Using the FIA training list data as input, I generated volume estimates for K=1 to 20. In this process, I limited the nearest neighbor selection to entries with the same GLUT/NLCD cover type. Following the recommendation of McRoberts (2002), the optimal K was selected as the value of k that produces an RMSE (Eq. 4.4) no larger than 2.5% of the minimum (RMSE value across the same range of K).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{b} (y_i - \hat{y}_i)^2}{n}}, \tag{4.4}$$

where: y_i is the ground-observed, assumed to be true, measurement for sample *i*, $\hat{y}_i$ is the predicted value for sample *i*, and *n* is the total number of samples.

The process of generating volume per acre (ft^3) estimates for pixel *i* initiated with the selection of the K-nearest entries in the training list. Nearness in this study refers to the Euclidean spectral distance and is calculated using equation 4.3. The process was executed with the following steps:

1. calculate ESD from each forested pixel *i* to each entry in the training list having the same GLUT/NLCD cover type,
2. use Fortran's intrinsic *minval* and *minloc* to find the first closest neighbor in the list,
3. store the volume per acre value associated with the spectrally nearest entry in the training list and mask it from the list of spectral distances,
4. repeat 2 & 3 K times, and
5. generate the KNN-based volume per acre (ft^3) using equation 4.3.

An advantage of the KNN method is the ability make many estimates for a single location as long as the information is available in the training list. The additional information I stored for each pixel included the mean age, mean basal area, mean spectral distance and a blended land cover. The land cover was created by storing the majority GLUT/NLCD forest type. A mixed type was assigned in cases where there was no majority.

In step 1 of the initial estimation process, I use GLUTNLCD land cover data is used to influence the nearest neighbor selection by limiting the available entries in the training list to only those with similar cover types (conifer, mixed, or deciduous). While the GLUT and NLCD were the most current state and/or national scale data sets publically available, they were not current to the dates of the TM used in this study and required fine-tuning to bring them up to the current timeline. I transformed the late-season TM to NDVI (Eq. 4.1) and generated two derivatives. The first ($NDVI_F$) contained the NDVI information for all forested pixels represented in the GLUTNLCD. The remaining non-forested pixels were masked out. The second derivative ($NDVI_{NF}$) had only NDVI information for the non-forested pixels represented in the GLUTNLCD and the forested pixels were masked out.

I used a series of thresholding and visual inspections of both NDVI derivatives to create a current timeline 1) forested mask, and 2) land cover layer (LCOV). I removed the areas from the forested mask that were originally classified as a forest, yet through visual inspection of the $NDVI_F$ data, were determined to be non-forested. On the other hand, I added to the forested mask the areas I determined to be wrongly classified as non-forest in the $NDVI_{NF}$ data. I assigned the forested pixels the blended land cover label to create the LCOV layer.

4.3.2 Balancing to the FIA mean volume per-acre

The objective of the mean balancing process was to adjust individual pixel estimates up or down so the TM-based mean equals the mean derived from the FIA plot measurements. Throughout the iterative process, pixels with the largest Euclidean spectral distance are adjusted first. In each subsequent pass, the ESD threshold for pixel selection and adjustment is lowered to include a larger number of pixels. Some pixels, especially those with a large ESD, may be adjusted multiple times while it is possible others won't be adjusted at all. Conifer, mixed, and deciduous cover types as denoted in LCOV were processed separately using the following procedures:

1. calculate the TM-based mean volume per acre (VAC_L) , include only cells attributed with the current LCOV type (conifer, deciduous, or mixed);

2. calculate the mean volume per acre of the FIA plots (VAC_F) attributed with the current LCOV type (conifer, deciduous, or mixed);

3. select the pixels equal to or larger than the ESD threshold, and either

   (a) adjust the selected pixels by the ratio of the maximum FIA plot volume per acre to the maximum estimated volume per acre represented in this set of pixels if VAC_L is less than VAC_F, or

   (b) decrease the selected pixel values by 2.5% if VAC_L is greater than VAC_F;

4. recalculate VAC_L,

5. repeat steps 3 and 4 until VAC_L is within 2% of VAC_F, and a

6. final rescale all pixels by the ratio of VAC_F to VAC_L to ensure the mean volume per acre pixel estimates for the scene equals the FIA's estimate for the scene.

I scaled the secondary age and basal area estimates to maintain their original slope relationships with volume per acre (ft^3). I used equation 4.5 to determine the slopes based on the observed measurements in the FIA data.

$$yi = B + B1X1i \tag{4.5}$$

where y is the scaled estimate for age or basal area, b is the slope parameter, b1 is the slope parm for hardwood

The slope parameters B1 and B2, were then applied based on the amount each volume per acre estimate changed (Eq. 4.6)

$$y - hat = (deltacuft) * B + B2, \tag{4.6}$$

where: y-hat is the estimated age or basal area measure, delta cuft is the amount a pixel's estimate changed from the initial KNN and mean balancing process, and B1 and B2 are slope parameters.

4.3.3 Fusion of industry and initial KNN estimates

I demonstrate the fusion of industry and the FIA data on the central Georgia, path 18, row 37 scene. The goal of this process was to equalize the local estimates from the initial KNN process with the industry cruise reports. Equalization was carried out on a stand-by-stand basis, where only the pixels within one of the industry stand boundaries were processed. All other areas received their initial KNN estimate and ESD measure. For each industrial stand individually, I:

1. determined the mean of the initial KNN estimate, then

2. raised or lowered the individual KNN estimates within that boundary by the ratio of the industry mean to the mean from step 1, and

3. reset the ESD measure for each of the pixels and then I

4. re-ran the mean balancing routine.

By resetting the ESD measures within each stand to zero, I reduce the chance, but do not eliminate the possibility, an individual estimate will be adjusted during the mean balancing process.

4.3.4 Assessment

All records in the training list were used in the volume estimation processes, thus no samples were withheld for accuracy assessment. I used Leave-one-out cross-validation to determine the optimal K for each scene. I present the RMSE associated with each optimal K as a measure of accuracy of the initial KNN estimation process. Since I only had access to the FIA plot locations inside the Forest Service's data warehouse, I was unable to assess the predictions in the GIS software. I developed secondary routines to track the would-be changes to the training list entries assuming they were actual pixels which enabled me to calculate the root-mean squared error (RMSE) (Eq. 4.4). Additionally, I calculated mean absolute errors (MAE) (Eq. 4.7)for the mean balancing results. Also, I present summaries of the mean balancing output for each scene stratified by the conifer, deciduous, and mixed types contained in LCOV for the initial KNN estimates and the mean balanced estimates.

$$MAE = \frac{(\sum_{i=1}^{n} |\hat{y}_i - y_i|)}{n}, \tag{4.7}$$

where y_i is the ground-observed, assumed to be true, measurement for sample *i*, $\hat{y}_i$ is the predicted value for sample *i*, and *n* is the total number of samples.

I also present an assessment of the estimates generated by the 1) initial KNN, 2) the mean balancing, and the 3) industry-infused and mean balanced processes for the central-Georgia scene, path 18, row 37. The field measurements and GIS data obtained from our industrial partners were not used in the first two estimations. Therefore, I present the RMSE and MAE calculated across each industrial stand as an assessment of their accuracy based on an independent data source.

However, that data is an integral part of the industry-infused and mean balanced routines, so I do not consider them suitable samples for independent validation. However, I present their summaries to confirm the improvement in prediction accuracy achieved through this process.

Finally, to demonstrate the varying results one would attain by querying the 1) unprojected FIA database, the 2) initial KNN, the 3) mean balanced, and the 4) industry-infused data, I present the results of a series of queries at varying scales. I first present sumamries for Hancock County, Georgia for the conifer, deciduous, and mixed types. There are 97 industrial forest stands, approximately 5,000 acres, located in the county. The final two are centered at 33.3141 degrees north and -82.9368 west with a radius of miles (502.7 acres) and 3 miles (24,630.1 acres). There are 5, 205 acres, forested stands within the 3.5-mile radius and one stand, less than 20 acres, within the 0.5-mile radius.

4.4 RESULTS

The path 19, row 39 scene is located in the extreme southwestern part of the state. The image covers approximately 185,000 acres of forested land and contains 69 FIA plots. The scene with the next smallest coverage of the state, path 18, row 39, encompasses 1.2 million acres of forested land and 339 forested FIA plots. Due to the small number of plots and a relatively large percentage of overlap by adjacent scenes, nearly 88%, I processed the path 19, row 39 data using K=1. While I scaled the volume per acre estimates to its FIA scene mean using the same approach as the others, where possible, I substituted the scaled estimates from scenes path 19, row 38 and path 18, row 39. Unless specified, the following sections focus on the remaining 11 scenes used in this study.

4.4.1 Selection of the optimal K

The leave-one-out KNN assessment of cubic-foot volume per acre based on the training data revealed an initial decrease in RMSE as the number of neighbors was increased. The gain in precision continued to K=3 to K=10 and then the leveled (Fig?. 4.5(Tab. 4.7). Root-mean squared error

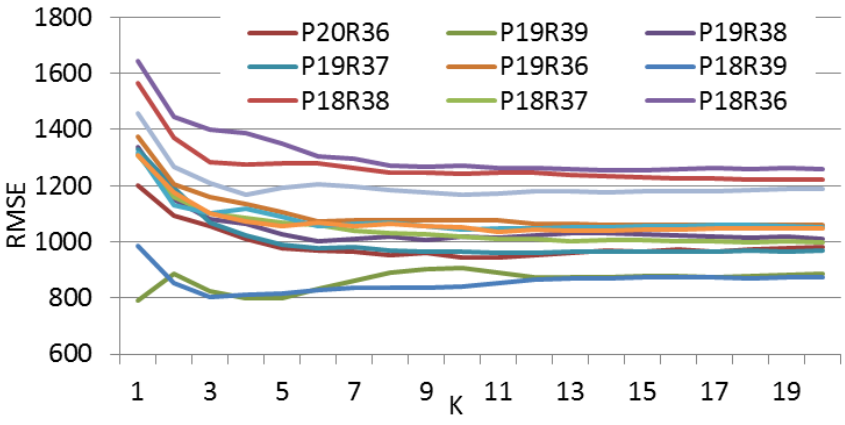

Figure 4.5: Root-mean squared error measures for K=1 to K=20 for the 12 TM scenes that were generated during the determination of the optimal K.

values for the optimal K ranged from 790 ft^3, 55% of the FIA mean for path 19, row 39, to 1.246, or 71% of the FIA mean for path 18, row 38 (Tab. 4.7)

Table 4.7: Optimum K and resulting combined-type RMSE for each TM scene used in the study.

Path / Row	Optimal K	RMSE	Relative RMSE
17 / 37	4	1169	67%
17 / 38	5	1055	60%
17 / 39	6	1056	60%
18 / 36	8	1271	73%
18 / 37	10	1019	58%
18 / 38	8	1246	71%
18 / 39	3	802	46%
19 / 36	6	1071	61%
19 / 37	6	976	56%
19 / 38	5	1027	59%
19 / 39	1	790	45%
20 / 36	6	969	55%

The compression of the range of initial volume per acre estimates is apparent in this study. Initial volume per acre estimates assessed on the entries of the training list data ranged from 0 to 5,553 ft^3, less than half of the range of the FIA measurements (Fig. 4.6). The KNN-derived mean

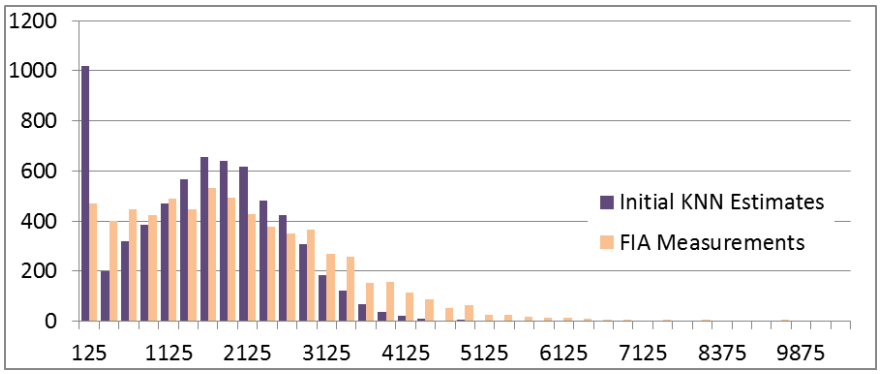

Figure 4.6: Histogram comparing the distributions of the FIA and the remotely sensed estimates made during the initial KNN process.

for the training list entries was 21% below the mean calculated from the FIA data, 1,454 ft^3 and 1,851, respectively.

The northern Georgia path 18, row 36 scene yielded the largest RMSE (1,271 ft^3) and the southern Georgia scene, path 18, row 39, produced the smallest (802 ft^3). The range of the simultaneous estimates, basal area and age, were also compressed. The range of basal area was 41% of the FIA's and the range of age was 22%. Both their means were approximately 22% below the means derived from the FIA data.

There were approximately 9% more forested acres, a total of 26.2 million acres, in LCOV than reported by the initial GLUTNLCD land cover data. There are 13 million acres of conifer represented in LCOV, 11.3 million acres of deciduous, and 7.2 million acres of mixed forest type.

4.4.2 Assessment of results on training samples

The summaries shown below are products of an assessment made on the training list samples compiled during both estimation processes where the training list entries were treated as a separate list of pixels in need of an estimate. The cover type designations used in these summaries are the same ones assigned in the LCOV data layer.

The initial KNN estimates of conifer volume per acre (ft^3) were on average 26% below the estimates derived from the FIA data (Tab. 4.7) and the mean balanced estimates 15% below. Root-mean squared error ranged from 974 ft^3 in the northern 19/36 scene to 1,541 ft^3 in the central 18/37 scene. Mean absolute errors ranged from 734 ft^3 in scene 18/39 which is located on the southern border of the state to 1,123 ft^3 (Tab. 4.7) in scene 18/38 which is located immediately to the north. As a percentage of the predicted mean, RMSE ranged from 81% in 18/37 to 148% in the northern Georgia 19/36 scene, and MAE from 64% in 17/37 to 99% in the north Georgia 20/36 scene.

There were two instances, scenes 17/37 in northeast Georgia and 18/39, where the scene-wide mean from the initial KNN process was larger than those from the mean balanced routine. In each case, the difference between the KNN estimates was closer to the FIA mean than the mean balanced results. Additionally, in every case, the initial KNN process yielded a lower RMSE and MAE. The largest difference between the KNN and mean balancing efforts occurred in scene 19/36 where the KNN process produced a scene-wide mean estimate of 974 ft^3 and the mean balanced process an estimate of 1,814 ft^3 (Tab. 4.7).

Table 4.8: Comparison of conifer FIA, KNN, and Mean balanced cubic-foot volume per acre estimates for entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	1,794	1,460	1,301	934	1,431	1,555	1,136
17 / 38	1,590	1,239	1,176	852	1,348	1,271	939
17 / 39	1,390	1,032	1,065	817	1,162	1,254	977
18 / 38	2,153	1,520	1,448	1,123	1,789	1,857	1,529
18 / 37	1,777	1,541	1,253	924	1,557	1,367	1,002
18 / 38	1,508	1,094	1,269	918	1,295	1,560	1,191
18 / 39	1,453	1,177	999	734	1,010	1,029	779
19 / 36	1,982	974	1,438	1,059	1,814	1,838	1,479
19 / 37	1,791	1,358	1,297	942	1,602	1,412	1,058
19 / 38	1,503	1,156	1,083	834	1,193	1,164	926
19 / 39	1,546	1,893	256	73	1,732	505	425
20 / 36	1,641	1,092	1,194	931	1,568	1,879	1,548

Mean estimates from both the initial KNN and mean balancing routines were below the FIA's estimates, 16% and 20%, respectively. The range of RMS errors for the KNN estimation process

ranged from a RMSE of 1,231 ft^3 in scene 18/39 to a RMSE of 1,980 ft^3 in the southeastern Georgia 17/39 scene, and MAE from 1,181 ft^3 in 18/39 to 2,155 ft^3 in 19/36 (Tab. 4.9). As a percentage of the mean, estimates on the the north Georgia 19/36 scene were the most accurate, yielding a RMSE relative to the mean of 68% and a MAE of 74%. On the other hand, scene 17/39 yielded the least accurate with a relative RMSE of 149% and relative MAE of 154%.

The initial KNN process produced deciduous means nearer to the FIA's than mean balancing in eight scenes. In all but one case, the southeastern 17/38 scene, the initial KNN process yielded lower RMSE and MAE measure. The differences in the two estimates were on average 4% with the largest occurring in 17/37, 1,548 ft^3 and 1,393 ft^3, respectively, and 18/38, 1,418 ft^3 and 1,253 ft^3, respectively (Tab. 4.9).

Table 4.9: Comparison of deciduous FIA, KNN, and Mean balanced cubic-foot volume per acre estimates for entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	1,916	1,584	1,500	1,053	1,393	1,538	1,082
17 / 38	1,853	1,701	1,634	1,163	1,606	1,633	1,188
17 / 39	1,746	1,325	1,980	1,303	1,349	2,081	1,316
18 / 38	2,488	1,869	1,666	1,224	1,817	1,817	1,352
18 / 37	1,955	1,577	1,519	1,067	1,473	1,633	1,151
18 / 38	1,812	1,418	1,717	1,269	1,253	1,826	1,317
18 / 39	1,605	1,284	1,231	913	1,181	1,276	915
19 / 36	2,452	2,109	1,439	1,040	2,155	1,590	1,162
19 / 37	2,070	1,653	1,462	1,076	1,599	1,580	1,172
19 / 38	1,811	1,675	1,429	968	1,538	1,511	997
19 / 39	1,795	1,697	382	135	1,539	705	461
20 / 36	2,172	1,624	1,540	1,119	1,657	1,543	1,130

Mixed type means estimated by the initial KNN and mean balancing approaches underestimated the FIA mean by 31% and 24%, respectively. The largest RMSE, 46% of the mean, and MAE, 73% of the mean, were in the south Georgia scene 17/38 (Tab. —reftab:tab49). The smallest RMSE was found in 19/38, south Georgia, and the smallest MAE in 18/38. There were three instances, 17/38, 18/39, and 19/38, where the KNN estimation process produced a larger mean and

a smaller FIA residual than the mean balancing process. Each of these scenes are located in the southern part of the state.

Table 4.10: Comparison of mixed type FIA, KNN, and Mean balanced cubic-foot volume per acre estimates for entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	1,937	1,310	1,284	870	1,626	1,538	1,216
17 / 38	1,486	796	1,253	902	397	1,462	1,134
17 / 39	1,385	1,035	814	615	1,282	771	600
18 / 38	2,427	1,569	2,236	1,398	1,948	2,173	1,287
18 / 37	1,798	1,371	1,425	1,016	1,396	1,465	1,083
18 / 38	1,449	1,077	1,120	835	1,352	1,189	861
18 / 39	1,436	963	1,372	938	668	1,496	1,091
19 / 36	2,334	1,366	1,512	1,104	1,929	1,743	1,318
19 / 37	1,808	1,402	1,162	861	1,778	1,372	1,070
19 / 38	1,294	1,094	1,020	785	928	1,013	783
19 / 39	1,634	3,010	26	19	2,109	883	883
20 / 36	1,993	990	1,490	1,243	1,306	1,511	1,266

scene-wide mean estimates of conifer basal area ranged from 60 ft^2 in 18/38 and 18/39 to 87 ft^2 in 18/37 for the initial KNN process and 49 ft^2 in 18/39 to 84 ft^2 in 18/37 for mean balancing (Tab.. 4.11). The KNN estimates differed from the FIA measures by 23% and the mean balanced by 24%. The largest RMSE relative to the mean is the northern 19/36 scene (76%) and the smallest in the central Georgia 18/37 scene (45%) The western Georgia scene 19/37 produced the smallest MAE in conifer basal area (51%) and south-central scene 18/38 the largest (84%). There were 5 instances where the initial KNN process produced a larger mean. In all cases, the KNN process resulted in a smaller RMSE and MAE (Tab. 4.11).

Both methods produced deciduous basal area means for the scene lower than the the means derived from the FIA. Means differed from the FIA on average by 21% for the initial KNN method and 28% for mean balancing. The southeastern Georgia scene 17/39 yielded means most different from their FIA measures for both methods. In terms of relative error, 17/39 generated the largest relative RMSE for the KNN (114%) and mean balancing (116%). The same scene, 17/39, yielded the largest relative MAE (82%) for the KNN process, however scene 18/38 yielded the largest for

Table 4.11: Comparison of conifer basal area estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	100	78	55	38	72	67	47
17 / 38	89	71	52	38	72	56	41
17 / 39	83	64	51	39	65	60	46
18 / 38	113	83	62	50	80	80	62
18 / 37	104	87	54	39	84	60	44
18 / 38	84	60	59	43	63	67	53
18 / 39	74	60	43	31	49	49	36
19 / 36	98	63	63	48	80	74	60
19 / 37	105	81	55	42	83	58	43
19 / 38	85	68	48	38	64	52	42
19 / 39	77	76	8	3	74	20	14
20 / 36	96	77	60	46	79	79	65

the mean balancing process. Scene 17/39 was the only scene where the mean balancing process produced lower errors (Tab. 4.12)

Estimates for mixed forest type basal area for the scenes were on average 28% to 30% below the FIA estimates. The south Georgia scene 18/39 produced the least accurate initial KNN estimate when compared to the FIA, yielding a relative RMSE of 144% and relative MAE of 102%. The least accurate results for the mean balancing process, on the other hand, were realized in scene 17/38 where the RMSE was 353% of the mean and the MAE was 280%. There were two instances, 17/39 and 18/38, where the mean balancing process produced lower RMSE and MAE results (Tab. 4.13).

Estimates for conifer age were on average 23% for the mean balancing process and 25%, for the initial KNN method, below the FIA's. Scene 18/39 in southern Georgia yielded the most accurate estimates the initial KNN method according to the relative RMSE, 71%, and relative MAE, 57%, where the FIA estimate is 31 years and the KNN estimate is 27 years (Tab. 4.14). The most accurate mean balancing estimates come from scene 17/38 whose RMSE is 87% of the mean (18 years) and

Table 4.12: Comparison of deciduous basal area estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	98	80	56	40	71	60	44
17 / 38	105	89	63	46	86	67	49
17 / 39	121	77	87	63	79	91	65
18 / 38	111	84	59	42	73	67	49
18 / 37	97	78	53	38	68	61	44
18 / 38	96	71	66	50	63	74	55
18 / 39	81	63	62	45	62	64	44
19 / 36	114	94	50	36	87	57	42
19 / 37	101	81	54	40	72	60	45
19 / 38	92	81	51	38	74	55	40
19 / 39	70	76	15	6	74	33	20
20 / 36	105	75	62	45	67	65	48

Table 4.13: Comparison of mixed type basal area estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	78	66	45	34	72	56	43
17 / 38	77	47	58	43	21	74	59
17 / 39	80	68	50	38	74	46	34
18 / 38	118	79	67	46	79	71	46
18 / 37	94	71	59	45	67	64	49
18 / 38	77	60	49	40	69	48	38
18 / 39	91	50	72	51	35	84	60
19 / 36	105	63	62	47	73	68	49
19 / 37	93	78	50	38	86	55	42
19 / 38	77	61	41	34	49	46	40
19 / 39	110	98	17	12	83	27	27
20 / 36	113	56	78	62	57	83	64

relative MAE is 65% of the mean (16 years) (Tab. 4.14). Scene 19/36, on the other hand, yields the largest RMSE and MAE relative to the mean for the initial KNN process, 138% and 98%, respectively. However, the largest relative RMSE result for the mean balancing process was found in scene 17/37 and the largest relative MAE in 18/38 (101%). The initial KNN process in all scenes yielded the lower RMSE and MAE (Tab. 4.14).

Table 4.14: Comparison of conifer age estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	28	21	20	13	19	24	17
17 / 38	25	21	18	13	21	18	14
17 / 39	24	19	17	12	19	19	14
18 / 38	33	22	26	20	25	30	25
18 / 37	28	23	19	14	22	20	15
18 / 38	26	19	20	15	19	22	18
18 / 39	31	27	19	15	22	22	17
19 / 36	30	18	25	18	26	26	22
19 / 37	31	21	22	15	23	23	16
19 / 38	29	22	21	16	20	23	18
19 / 39	39	42	13	3	37	12	9
20 / 36	25	18	20	15	21	24	21

Deciduous scene mean age estimates were 15% below the FIA for the initial KNN and 7% below for the mean balanced process. Scene 19/36 in extreme north Georgia produced the most accurate estimates using both methods. Root-mean squared error for19/36 was 61% of the mean (31 years) for the initial KNN process and 56% (33 years) for mean balancing (Tab. 4.15). Mean absolute error relative to the mean was 47% for KNN and 43% for mean balancing. The largest relative errors were found in the southeastern scenes 17/39 and 18/38. In all cases, the initial KNN process yielded the lowest RMSE and MAE measures.

Mean mixed type age estimates for each scene using the initial KNN method were on average 26% below the FIA's measure and 28% below using the mean balancing method. The 17/39 scene in southeastern Georgia yielded the most accurate results for both methods and each measure of error. The scene just to the north, 17/38, produced the least initial KNN estimates according to

Table 4.15: Comparison of deciduous age estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	39	31	28	20	32	30	22
17 / 38	39	32	26	20	35	27	22
17 / 39	41	25	30	23	30	31	23
18 / 38	59	42	38	29	49	40	30
18 / 37	43	33	28	23	36	31	25
18 / 38	42	29	33	25	31	35	27
18 / 39	45	33	30	22	34	30	23
19 / 36	62	51	31	24	60	33	25
19 / 37	46	37	27	20	39	31	23
19 / 38	42	39	24	18	40	27	20
19 / 39	44	45	3	1	40	19	15
20 / 36	56	42	33	25	48	34	25

RMSE (145%, 27 years) and for mean balancing (455%, 8 years) (Tab. 4.16). The north Georgia scene 20/46produced the largest MAE for the initial KNN estimation process (120% of the mean) and scene 17/38 for the mean balancing process. There were five instances where the mean difference between the initial KNN estimates and the FIA were larger than the differences between the mean balancing estimates and the FIA.

A comparison of the FIA scene means and the initial conifer cubic-foot volume per acre estimates for each scene summarized for the LCOV types are shown in table 4.17. On average, scene-wide estimates of conifer volume were 28% below the mean calculated from the FIA measurements. The path 19, row 37 scene located in west-central part of the state yielded the lowest mean conifer volume per acre (ft^3) (Fig. 4.7). The path 19, row 36 scene immediately to the north yielded the second smallest mean conifer volume per acre, 862 ft^3. These two scenes were 57% and 56% the FIA's mean estimates for the same area. Pixel estimates ranged from zero to 4,897ft^3 and had a mean of 785 ft^3 and a standard deviation of 956. The scene with the maximum conifer volume

Table 4.16: Comparison of mixed type age estimates from the FIA, KNN, and Mean balanced for all entries in the training list.

Path/	FIA	KNN			Mean Balance		
Row	Mean	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)	Mean (ft^3)	RMSE (ft^3)	MAE (ft^3)
17 / 37	34	27	26	20	26	30	23
17 / 38	34	19	27	21	8	35	28
17 / 39	32	26	21	15	28	20	15
18 / 38	44	31	32	24	32	33	24
18 / 37	35	27	25	19	26	27	21
18 / 38	29	22	25	19	25	25	20
18 / 39	38	21	26	20	20	27	23
19 / 36	41	28	24	18	32	29	23
19 / 37	30	26	21	15	29	22	16
19 / 38	24	24	25	20	17	22	17
19 / 39	86	62	35	25	63	23	23
20 / 36	32	17	24	20	18	29	25

per acre was located in the central part of the state (path 17, row 37) (mean: 1,577, sd 853) with individual pixel estimates ranging from zero to 4,442 ft^3.

scene-wide estimates of deciduous volume per acre were on average 6% below their respective FIA estimates. The satellite-derived estimates for four scenes were above the FIA's by approximately 3% while the remaining were below by an average of 11%. Path 19, row 37 yielded the smallest mean deciduous volume per acre (Tab. 4.18) and the largest underestimate (Fig. 4.7). Pixel estimates within the scene ranged from 0 to 4,892 ft^3 and had a mean and standard deviation of 1,256 ft^3 and 1,043 ft^3, respectively. The northern scene, path 19, row 36, produced the largest mean deciduous volume per acre (mean: 2,274 ft^3, sd: 759 ft^3) for a scene. Pixel estimates of hardwood volume per acre in this area ranged from 0 to 5,654 ft^3.

Estimates of the mixed cover type volume per acre were on average 28% below the FIA's estimate. Path 18, row 39 was the only scene to yield a mean above the FIA's. South Georgia scenes path 18, row 39 and path 17, row 39 yielded produced the extreme mean estimates of mixed type volume per acre (Fig. 4.7). Pixel estimates within path 18, row 39 ranged from zero to 4,286

Table 4.17: scene-wide summaries of conifer volume per acre generated during the initial KNN estimation process.

	FIA				Initial TM		
Path/Row	Mean	SD	# Plots	Rank	Mean	SD	Max
17 / 37	1,794	115	358	2	1,544	880	5,323
17 / 38	1,590	77	682	5	1,244	765	5,168
17 / 39	1,390	109	264	8	1,023	564	4,900
18 / 36	2,153	252	107	3	1,470	1,011	3,876
18 / 37	1,777	86	613	1	1,577	853	4,442
18 / 38	1,508	71	689	7	1,119	743	4,434
18 / 39	1,453	129	183	4	1,370	717	4,130
19 / 36	1,982	210	143	10	862	757	5,654
19 / 37	1,791	116	345	11	785	956	4,897
19 / 38	1,503	113	253	6	1,203	710	4,139
19 / 39	1,546	310	37	NA	1,582	1,198	3,939
20 / 36	1,641	225	84	9	937	695	4,017

Table 4.18: scene-wide summaries of deciduous volume per acre generated during the initial KNN estimation process.

	FIA				Initial TM		
Path/Row	Mean	SD	# Plots	Rank	Mean	SD	Max
17 / 37	1,916	174	217	3	1,989	1,065	7,335
17 / 38	1,853	132	366	5	1,918	994	7,002
17 / 39	1,746	240	117	9	1,686	1,025	6,226
18 / 36	2,488	176	267	2	2,160	614	4,200
18 / 37	1,955	115	481	6	1,898	741	4,910
18 / 38	1,812	114	478	8	1,799	849	5,865
18 / 39	1,605	202	106	10	1,670	718	4,300
19 / 36	2,452	130	461	1	2,274	759	5,654
19 / 37	2,070	132	358	11	1,256	1,043	4,892
19 / 38	1,811	164	218	7	1,807	817	5,250
19 / 39	1,795	447	25	NA	1,825	1,008	3,939
20 / 36	2,172	207	148	4	1,925	618	4,182

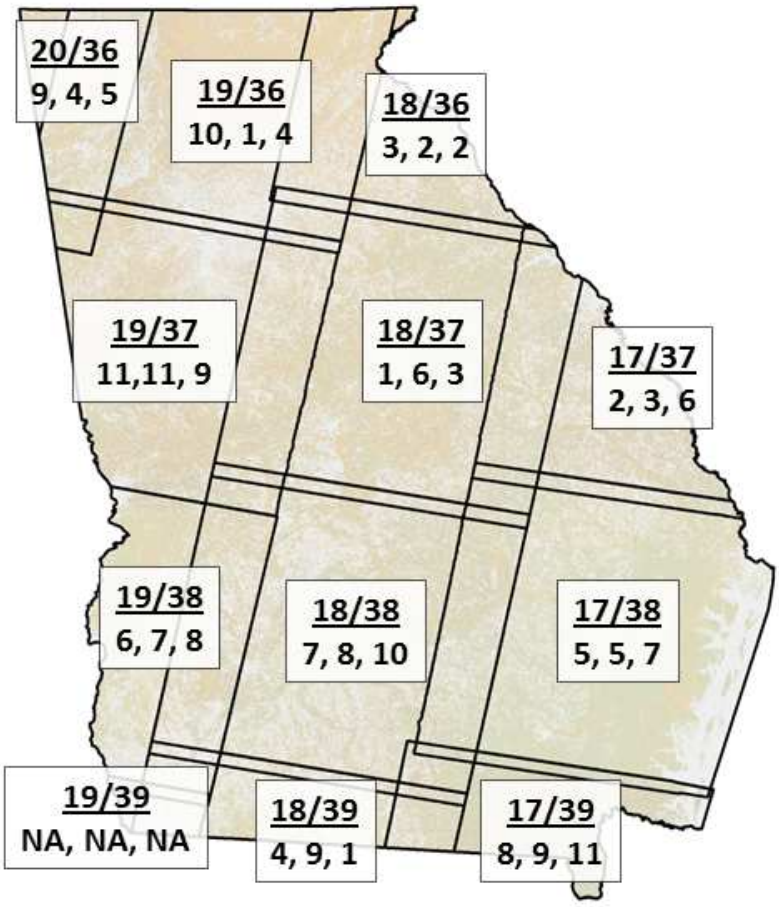

Figure 4.7: Volume per acre (ft^3) ranks for each Landsat scene for the conifer, deciduous, and mixed cover types represented in LCOV.

ft^3 and the scene's mean was 1,517 ft^3 (sd 685 ft^3) (Tab. 4.19). On the low side, pixel estimates in the path 17, row 39 scene ranged from 0 to 4,806 ft^3 and had a mean of 958 ft^3 (sd 580).

Table 4.19: scene-wide summaries of mixed volume per acre generated during the initial KNN estimation process.

	FIA				Initial TM		
Path/Row	Mean	SD	# Plots	Rank	Mean	SD	Max
17 / 37	1,937	245	88	6	1,261	861	5,348
17 / 38	1,486	168	138	7	1,236	804	6,149
17 / 39	1,385	265	49	11	958	580	4,806
18 / 36	2,427	319	79	2	1,497	781	4,071
18 / 37	1,798	146	230	3	1,492	782	4,644
18 / 38	1,449	144	170	10	1,150	736	4,662
18 / 39	1,436	258	50	1	1,517	685	4,286
19 / 36	2,334	224	143	4	1,437	812	5,654
19 / 37	1,808	170	166	9	1,157	831	4,641
19 / 38	1,294	188	75	8	1,195	636	4,919
19 / 39	1,634	769	7	NA	1,858	1,112	3,939
20 / 36	1,993	330	49	5	1,308	608	4,182

The simultaneous estimates of conifer basal area ranged from 0 ft^2 in all scenes to 207 ft^2 (Tab. 4.20) in the eastern Georgia path 17, row 37 scene which also yielded the lowest mean basal area (53 ft^2) (Tab. 4.20). Path 18, row 37 located in the center of the state returned the largest mean, 89 ft^2. Deciduous estimates ranged from 0 in all scenes to 269 ft^2 in path 17, row 38 in southeastern Georgia. Path 19, row 36 scene in the north gave the largest scene-wide mean (102 ft^2) and the south Georgia path 18, row 39 scene yielded the smallest (81 ft^2). scene-wide means for the mixed type ranged from 64 ft^2 in path 18, row 38 in the south central part of Georgia to 80 ft^2 in path 18, row 37.

Maximum estimates of conifer age ranged from 0 in all scenes to 112 (Tab. 4.21) in path 18, row 39. Maximum estimates of deciduous age ranged from 0 to 119 in path 19, row 36 and from 0 to 112 for the mixed type in path 18, row 39. scene-wide mean estimates for conifer age ranged from 17 years in path 19, row 36 to 31 years in path 18, row 39 (Tab. 4.21).; maximum scene-wide estimates of deciduous age ranged from 31 years (path 17, row 39) to 55 years (path 19, row 36) and from 64 (path 17, row 39) to 112 (path 18, row 39) for the mixed type.

Table 4.20: scene-wide summaries of basal area for each scene and type.

	Conifer			Deciduous			Mixed		
Path/Row	Max	Ave	Sd	Max	Ave	Sd	Max	Ave	Sd
17/37	190	82	42	222	96	38	198	68	41
17/38	207	71	36	269	100	39	246	70	35
17/39	179	64	29	258	93	38	202	65	28
18/36	165	78	51	162	99	22	162	76	37
18/37	180	89	41	178	93	26	178	80	32
18/38	186	62	39	221	90	32	193	64	36
18/39	178	66	31	198	81	29	182	73	28
19/36	196	53	44	195	102	26	196	69	32
19/37	193	71	49	185	88	32	178	74	34
19/38	175	71	40	201	91	31	191	66	30
19/39	154	65	41	154	81	36	154	72	39
20/36	197	65	49	191	89	23	191	69	30

Table 4.21: scene-wide summaries of age for each scene and type.

	Conifer			Deciduous			Mixed		
Path/Row	Max	Ave	Sd	Max	Ave	Sd	Max	Ave	Sd
17/37	98	23	12	100	39	17	98	27	15
17/38	88	22	12	94	37	15	88	27	13
17/39	67	20	11	70	31	11	64	23	11
18/36	83	24	15	104	49	16	91	29	14
18/37	75	24	12	81	40	13	77	30	13
18/38	72	21	11	90	37	13	77	25	11
18/39	112	31	16	115	44	17	112	38	16
19/36	96	17	13	119	55	19	108	31	14
19/37	74	19	13	88	39	16	85	27	12
19/38	81	22	12	95	42	14	92	31	13
19/39	125	32	26	125	51	34	125	50	27
20/36	75	17	12	87	48	15	83	28	11

4.4.3 Scene-wide summaries

Summaries of the entire initial KNN and mean balanced estimated surfaces follow. All forested pixels are included in these results. The cover type specifications were assigned by the LCOV data layer.

Mean conifer volume per acre estimates before mean balancing were on average 26% lower than the mean derived from the FIA. Thirty-nine percent of all the conifer-classified pixels in the state required adjustment to attain equalization. Four scenes needed adjustments to 100% (Tab. 4.22) of their conifer-classified areas, while the other eight scenes required adjustments to 20% or fewer. In total, 5,693,316 conifer acres across the 12 scenes (Tab. 4.22) were scaled. After mean balancing, conifer volume per acre estimates ranged 0 ft^3 to 11,366 ft^3 and a mean of 1,657 ft^3, compared to 0-5,654 and a mean of 1,226 before processing

The initial conifer mean in the northern scene, path 19, row 36, was 57% below the FIA's (Tab. 4.22). In order to raise that scene's conifer mean to the appropriate level, 100% of the conifer pixels (927,580 acres) were adjusted. The range of the adjusted data in this scene increased from 0-5,654 ft^3 to 0-11,366 ft^3 with a mean of 1,983 ft^3, which is equal to the FIA's. The conifer mean for the scene immediately to the south, path 19, row 37 was 56% below its associated FIA mean. However, only 20% of the pixels (327,212 acres) required adjustment.

Table 4.22: Amount of coniferous area adjusted during the mean balancing process.

Path/Row	Initial Mean (% of FIA Mean)	Adjusted Area (ac)	Adjusted Area (%)	Adjusted Mean (sd)	Mim	Max
17 / 37	-14%	18,487	2%	1,794 (980)	1	9,098
17 / 38	-22%	194,449	6%	1,590 (944)	1	8,934
17 / 39	-26%	153,955	12%	1,390 (716)	1.5	5,441
18 / 36	-32%	401,973	100%	2,154 (943)	0	6,746
18 / 37	-11%	104,270	4%	1,776 (967)	1	8,657
18 / 38	-26%	2,922,854	100%	1,506 (2639)	0	3,483
18 / 39	-6%	82,201	12%	1,454 (717)	1.2	4,291
19 / 36	-57%	927,580	100%	1,983 (718)	0	11,366
19 / 37	-56%	327,212	20%	1,792 (1184)	1	8,146
19 / 38	-20%	133,814	12%	1,503 (866)	1	5,915
19 / 39	2%	16	<1%	1,546 (856)	1	3,985
20 / 36	-43%	426,505	100%	1,641 (2875)	0	3,895

Before mean balancing, deciduous volume per acre estimates were on average 5% below the FIA-derived mean volume per acre (Tab. 4.23), and required adjustments to approximately 10% of the pixels to equalize the scene-wide deciduous volume per acre means . Approximately 1,052,000 acres of this type were scaled across the 12 scenes (Tab. 4.23). After adjustment, the range of volumes increased from 0 -7,335ft^3 to 0-12,024 ft^3 and the mean increased from 1,851 to 2,053 ft^3.

Path 19, row 37 was the scene with the largest overall deciduous underestimate (39%) (Tab. 4.23). During the mean balancing process, adjustments were made to 15,187 acres, approximately 1% of the deciduous-typed pixels in the scene, before its mean was equalized with the FIA's. The range of the processed data was raised from 0-4,982 ft^3 to 2 8,590 ft^3 (mean:2,069, sd: 808 ft^3) (Tab. 4.23). The eastern Georgia, path 17, row 37 scene required an adjustment of 100% of the deciduous-classified pixels, approximately 822,000 acres, to achieve mean balancing,. The remaining scenes required adjustments to 5% or fewer of their deciduous-classified pixels to achieve equilibrium.

Table 4.23: Amount of deciduous area adjusted during the mean balancing process.

Path/Row	Initial Mean (% of FIA Mean)	Adjusted Area (ac)	Adjusted Area (%)	Adjusted Mean (sd)	Mim	Max
17 / 37	4%	822,320	100%	1,917 (936)	0	6,941
17 / 38	4%	81,203	5%	1,853 (931)	1	6,643
17 / 39	-3%	17	<1%	1,747 (929)	1	11,724
18 / 36	-13%	23,232	2%	2,487 (1043)	1	10,226
18 / 37	-3%	607	<1%	1,953 (791)	2.1	12,024
18 / 38	-2%	54	<1%	1,812 (781)	38.5	11,200
18 / 39	4%	49,769	11%	1,605 (664)	1	4,218
19 / 36	-7%	40,344	2%	2,450 (1002)	1	9,852
19 / 37	-39%	15,187	1%	2,069 (808)	1.8	8,590
19 / 38	<-1%	21	<1%	1,811 (786)	750	9,346
19 / 39	2%	11	<1%	1,795 (811)	1	3,998
20 / 36	-11%	19,383	3%	2,172 (760)	1.5	5,975

On average, the mixed cover type scene-wide estimates were 21% below the FIA estimates before the mean balancing process was initiated (Tab. 4.24). Thirty-seven percent, 670,812, of the mixed type area was adjusted. Seven scenes required the scaling of 13% or fewer of the mixed type

pixels in their respective scenes (Tab. 4.24). Three scenes required an adjustment to 100% of the pixels in this classification. Two of them, path 19, row 36 and path 20, row 26 are located in the northern extreme of the state. The third northern scene, path 18, row 36, required an adjustment of 60% of the mixed-type pixels. Estimates in the southern Georgia path 18, row 39 scene, on the other hand, were reduced by 6%. Before processing, the range of volume per acre (ft^3) estimates was 0-6,149 with a mean of 1,339 ft^3, and afterwards the range was 0-8,718 with a statewide mean of 1,774.

Table 4.24: Amount of mixed type area adjusted during the mean balancing process.

Path/Row	Initial Mean (% of FIA Mean)	Adjusted Area (ac)	Adjusted Area (%)	Adjusted Mean (sd)	Mim	Max
17 / 37	-35%	102,791	43%	1,936 (1219)	1.2	5,558
17 / 38	-17%	1,314	1%	1,486 (752)	1.5	6,778
17 / 39	-31%	30,498	13%	1,385 (808)	1.4	4,719
18 / 36	-38%	72,452	60%	2,427 (1070)	1.5	6,659
18 / 37	-17%	24,890	6%	1,798 (899)	1.6	6,978
18 / 38	-21%	5,275	1%	1,449 (810)	1.7	6,615
18 / 39	6%	3	<1%	1,436 (601)	363	2,193
19 / 36	-38%	288,599	100%	2,334 (736)	0	8,718
19 / 37	-36%	49,520	11%	1,808 (895)	1.6	6,410
19 / 38	-8%	2,954	3%	1,294 (650)	1.4	5,147
19 / 39	14%	16,603	100%	1,634 (711)	0	3,746
20 / 36	-34%	75,912	100%	1,993 (448)	0	6,011

4.4.4 Fusion of industrial data in path 18, row 37

The path 18, row 37 mean of the initial KNN-based estimates for conifer volume per acre were more than 11% below the FIA's estimate (Tab. 4.25). The deciduous mean was almost 3% below and the mixed forest type was 17%below the FIA. By design, after scaling the scene-wide means for each type were all near equal to the mean derived from the 18/38 FIA measures (+/- 0.2%). The maximum pixel estimate for both the conifer (8,604 ft^3) and deciduous types (8,280 ft^3) from the industry-infused data were almost twice their FIA and initial KNN counterparts. However, the pixel maximum mixed type estimate was slightly lower, 4,130 ft^3, than the FIA measurements and the Initial KNN processed data and 40% below the mean balanced data.

Table 4.25: Comparison of mean conifer, deciduous and mixed volume per acre estimates reported by the FIA and generated from the three remote sensing methods for the path 18, row 37 scene.

	Conifer			Deciduous			Mixed		
Method	MEAN	STD	Max	MEAN	STD	Max	MEAN	STD	Max
FIA	1,777	86	4,442	1,955	115	4910	1,798	146	4644
Initial KNN	1,577	853	4,442	1,898	741	4910	1,492	782	4644
Mean balanced	1,776	967	8,657	1,953	791	12024	1,798	917	6978
Industry Fused	1,780	995	8,604	1,962	140	8280	1,797	912	4130

The mean stand volume per acre produced by the initial KNN estimation routine was 27% below the mean calculated from the industry ground measurements (Tab. 4.26) and the range of predicted stand means was half. The average from the mean balancing process was 18% below the industry's measure with a compressed range of estimates of almost 17%. By design, the average stand cubic foot volume per acre and the industry measures are nearly equal. There is a 2% difference in means and a compression of stand mean estimates of 10%. The initial KNN and the mean balancing process yielded similar RMSE measures of 1,322 ft^3 and 1,382 ft^3, and MAE measures of 1,052 ft^3 and 1,126 ft^3, respectively (Tab. 4.26). The RMSE and MAE from the industry-fused process was nearly 60% lower, 541 ft^3 and 406 ft^3, respectively.

Table 4.26: Stand-level comparison of mean conifer volume per acre generated from the three estimates based on remote sensing.

Measure	Mean	SD	Min	Max	RMSE	MAE
Industry	2,261	1,544	0	6,093		
Initial KNN	1,652	657	29	3,024	1,322	1,052
Mean-balanced	1,850	770	23	5,102	1,382	1,126
Industry Fused	2,218	1,149	15	5,495	541	406

The scatter plots in Figure 4.8 show the weak positive relationship between the industry observed stand's cubic foot volume and its remotely sensed estimates using the initial KNN procedure (Fig. 4.8A) and the mean balancing routine (Fig. 4.8B). By design, there is a strong positive relationship with the industry measures and the industry-fused estimates (Fig. 4.8C). The effects of the scaling that occurred during the mean balancing process (Fig. 4.8B) are apparent throughout

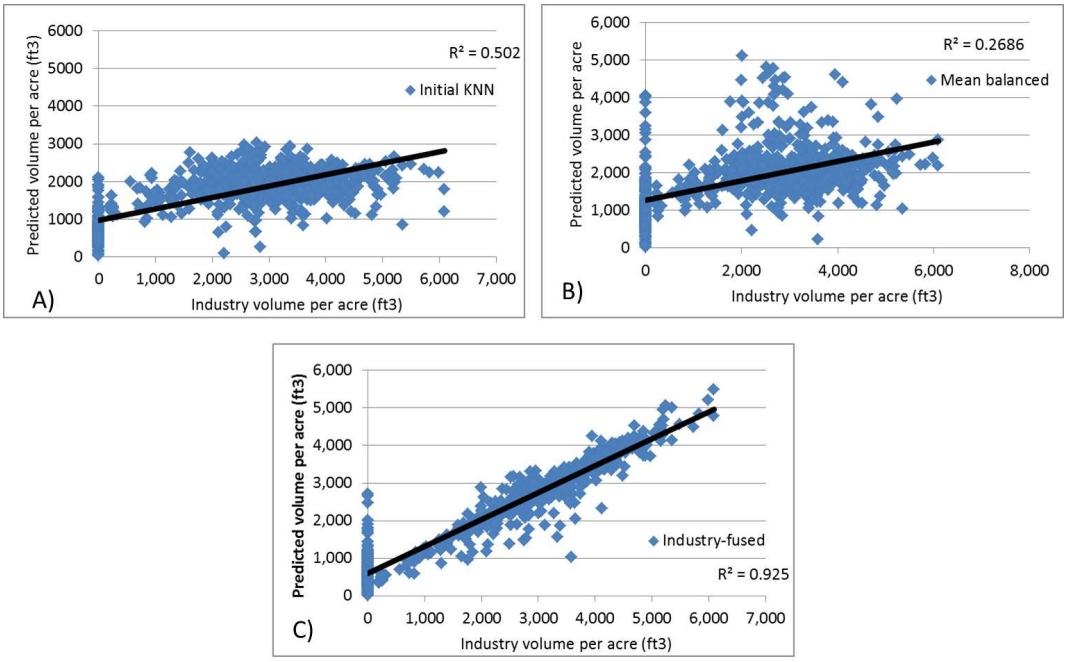

Figure 4.8: Scatter plots reflecting the volume per acre (ft^3) estimates for each industry stand from the A) initial KNN, the B) mean balancing, and the C) industry-fused processes.

the extent of the industry measurements. The range of estimates for the zero-volume samples (i.e. Harvested sites) expanded from 0 to just above 2,000 ft^3 (Fig. 4.8A) to 0 to approximately 4,000 ft^3 (Fig. 4.8B). Additional effects of the mean balancing are visible for the industry stands whose estimates range from approximately 2,000 ft^3 to 5,000 ft^3.

Mean stand basal area produced during the initial KNN and mean balancing processes were within 5% and 8% of the industry mean (Tab. 4.27), however the industry-fused mean was over estimated by 19%. The RMSE measures ranged from 40 ft^2 for the initial KNN process and 43 ft^2 for the other two measures and their MAEs ranged from 34 for the initial KNN to 36 for the mean balanced process (Tab. 4.27). The expansion of the estimates that occur in the mean balancing and industry-fused processes is apparent in the scatter plots below (Fig. 4.9). The largest estimates for the zero-basal area industry stands, those stands that were harvested, from the initial KNN process was approximately 140 ft^2, which increased to above 150 for the industry-fused routine.

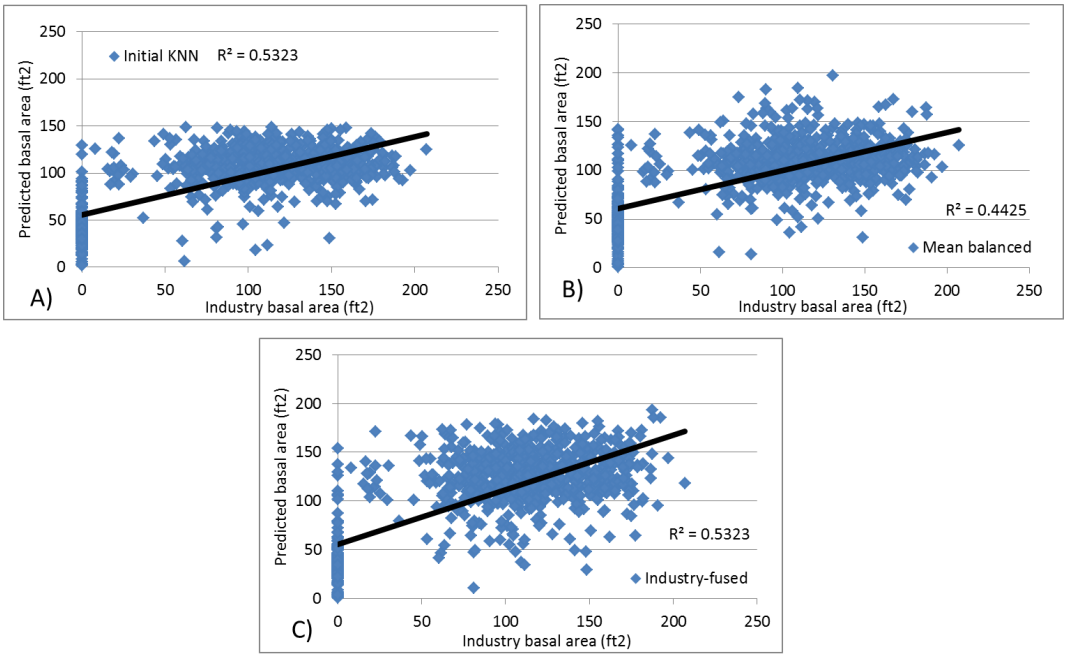

Figure 4.9: Scatter plots reflecting the basal per acre (ft^2) estimates for each industry stand from the A) initial KNN, the B) mean balancing, and the C) industry-fused processes.

Similar scatter is apparent in the mid and far ranges of the industrial basal area measurements, too (Fig. 4.9).

Table 4.27: Stand-level comparison of mean basal area per acre generated from the three estimates based on remote sensing.

Measure	Mean	SD	Min	Max	RMSE	MAE
Industry	88	58	0	207	NA	NA
Initial KNN	92	33	1	151	40	34
Mean-balanced	95	34	1	197	43	36
100% Industry Fused	105	45	1	193	43	35

Estimates of stand age differed between 44% for the initial KNN process and 72% for the mean balanced routine. The initial KNN routine generated the most accurate predictions of stand age yielding a RMSE of 11 years and a MAE of 8 years. The industry-fused process yielded the largest RMSE, 16 years, and MAE, 14 years (Tab. 4.28).

Table 4.28: Stand-level comparison of mean stand age generated from the three estimates based on remote sensing.

Measure	Mean	SD	Min	Max	RMSE	MAE
Industry	18	12	0	61	NA	NA
Initial KNN	26	8	1	48	11	9
Mean-balanced	28	9	1	59	14	11
100% Industry Fused	31	11	1	60	16	14

The range of errors for the industrial stands with zero value increased in the mean balancing process to a maximum of approximately 55 years (Fig. 4.10) and then decreased to near 40 years during the industry-fused process. Ages in the mid-range of the industry data were compressed during both the mean balancing and the industrial-fused estimations (Fig. 4.10).

4.4.5 Multi-scale queries

Total conifer area reported by the FIA and the area of conifer represented in LCOV for Hancock County, Georgia are nearly identical. The FIA reports 138,886 acres of coniferous forestland while LCOV represents 138,862 total conifer acres. Each of the remotely sensed processes yielded a mean conifer volume per acre larger than what the FIA reported. The initial KNN process yields a mean volume per acre of 1,798 ft^3 (Tab. 4.29), 17% more than the FIA; mean balancing estimates 1,908 ft^3,25% more, and the industry-fused process yields an estimate of 1,933 ft^3 (Tab. 4.29), 26% more than the FIA. The difference between the FIA's estimate, 213 million ft^3 (Tab. 4.29) and the remote sensing estimates for total conifer volume ranged from 16% to 25%. The initial KNN process yields 247 million ft^3, mean balancing 262 million ft^3, and the industry-fused process 266 million ft^3.

The total area of deciduous forestland reported by the FIA was 71,927 (Tab. 4.30) acres, nearly 40% below the 114,466 acres of deciduous forest represented in LCOV. The FIA reported that there was an average volume in the county of 1,533 ft^3 per acre while the initial KNN process produced 1,945 ft^3 per acre, mean balancing 2,002 ft^3 per acre, and the industry infused 2,029 ft^3

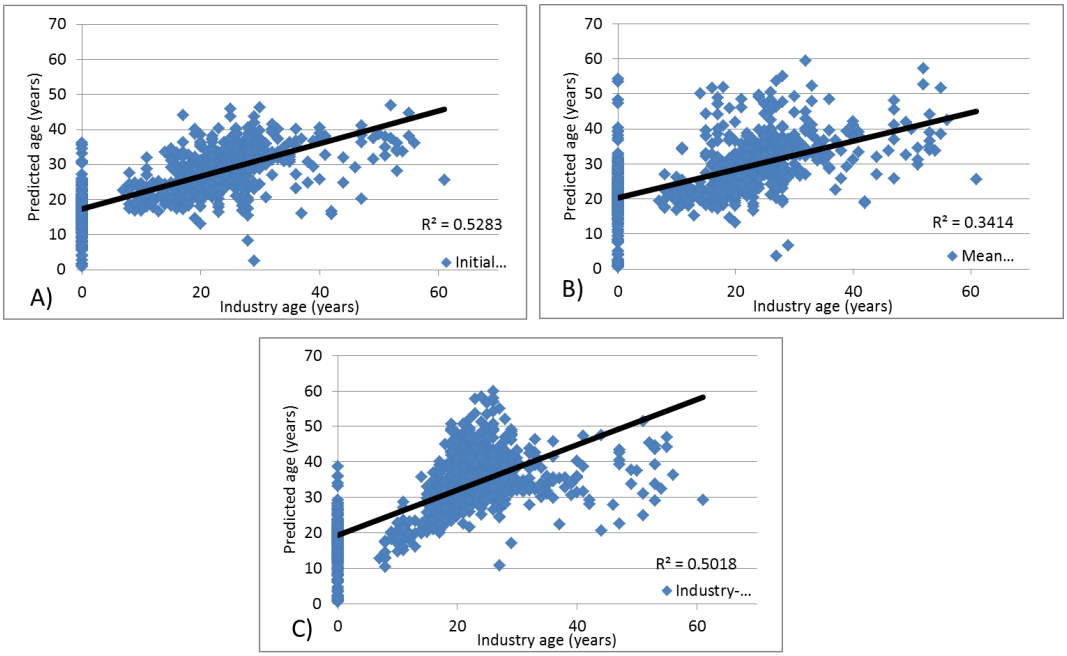

Figure 4.10: Scatter plots reflecting the age (years) estimates for each industry stand from the A) initial KNN, the B) mean balancing, and the C) industry-fused processes.

Table 4.29: Conifer volume per acre estimates generated from the FIA, the initial KNN, mean balancing, and the industry-fused methods for Hancock County, Georgia.

Measure	Type	Area (ac)	Min	Max	Mean	SD	Volume (Mil. ft^3)
FIA Db (Hancock)	Conifer	138,886	NA	NA	1533	NA	213
Initial KNN	Conifer	138,862	0	4,133	1,798	768	247
Mean balanced	Conifer	138,862	1	8,490	1,908	860	262
Industry-fused	Conifer	138,862	1	8,439	1,933	911	266

per acre (Tab. 4.30), 44% to 51% more. Due primarily to the discrepancy in total area represented, the initial KNN process yields a total volume estimate, 220 million ft^3 (Tab. 4.30), 2.2 times more than what is reported by the FIA. The mean balancing process yields a total of 227 million ft^3, 2.3 times the FIA, and the industry-fused process gives 230 million ft^3, 2.4 times the FIA.

Table 4.30: Deciduous volume per acre estimates generated from the FIA, the initial KNN, mean balancing, and the industry-fused methods for Hancock County, Georgia.

Measure	Type	Area (ac)	Min	Max	Mean	SD	Volume (Mil. ft^3)
FIA Db (Hancock)	Decid.	71,927	NA	NA	1,346	NA	97
Initial KNN	Decid.	114,466	43	4,577	1,945	751	220
Mean balanced	Decid.	114,466	43	10,654	2,002	808	227
Industry-fused	Decid.	114,466	0	3,930	2,029	786	230

The FIA reports 67% more mixed type area than what is represented in LCOV. Mean volume per acre (ft^3) estimates for the initial KNN, 1,582, the mean balancing process, 1,833 ft^3 (Tab. 4.31), and the industry-fused process, 1,950 ft^3 were 71% to 111 % above the FIA's reported 926 ft^3 per acre. Differences in estimates of total volume, again due to the discrepancy in total area reported, ranged from 55% to 68% below what was reported by the FIA. Remotely sensed total volume estimates ranged from 29 million ft^3 to 35 million ft^3, while the FIA reported 52 million ft^3.

Table 4.31: Mixed type volume per acre estimates generated from the FIA, the initial KNN, mean balancing, and the industry-fused methods for Hancock County, Georgia.

Measure	Area (ac)	Min	Max	Mean	SD	Volume (Million ft^3)
FIA Db (Hancock)	56,450	NA	NA	926	NA	52
Initial KNN	18,350	13	3,953	1,582	791	29
Mean balanced	18,350	20	6,312	1,833	922	33
Industry-fused	18,350	15	8,965	1,950	997	35

The FIA reported 5,607 (Tab. 4.32) acres of conifer, 5,891 acres of deciduous and no existing mixed cover type area in the 3.5-mile radius query area. However, the LCOV layer contains 11,041 acres of conifer, 7,950 acres of deciduous, and 1,275 acres of mixed woods. Remotely sensed conifer volume estimates were below the FIA's reported value by 50% from the initial KNN and industry-fused data to 54% in the mean balanced data. FIA reports a volume per acre of 3,706 ft^3

while the TM-derived data reports a volume per acre of 1,719 to 1,861 ft^3 (Tab. 4.32). Deciduous estimates of volume per acre differed by 751% to 782% from the FIA's reported 263 ft^3.

Table 4.32: Query results from the 3.5-mile radius query to the FIA database, the initial KNN, mean balancing, and the industry-fused methods for the conifer, deciduous, and mixed types.

Measure	Type	Area (ac)	Min	Max	Mean	SD	Volume (million cu.ft)
FIA Db Query	Conifer	5,607	NA	NA	3,706	NA	21
Initial KNN	Conifer	11,041	1	7,432	1,849	940	20
Mean balanced	Conifer	11,041	0	3,809	1,719	811	18
Industry-fused	Conifer	11,041	1	7,694	1,861	1,021	20
FIA Db Query	Deciduous	5,891	NA	NA	263	NA	2
Initial KNN	Deciduous	7,950	73	10,654	2,041	844	16
Mean balanced	Deciduous	7,950	55	4,450	1,974	752	15
Industry-fused	Deciduous	7,950	56	3,915	2,057	786	16
FIA Db Query	Mixed	0	NA	NA	0	NA	0
Initial KNN	Mixed	1,275	20	5,656	1,905	980	2
Mean balanced	Mixed	1,275	13	3,881	1,632	808	2
Industry-fused	Mixed	1,275	0	6,009	2,012	1,021	3

FIA reports no forestland area or volume in the half-mile query (Tab. 4.33). The remotely sensed estimates in this query area ranged from a mean conifer volume per acre of 2.036 ft^3 from the initial KNN estimate to 2,251 ft^3 from the industry-fused data. Deciduous estimates ranged from 2,168 ft^3 to 2,253 ft^3 (Tab. 4.33) and mixed type from 1,734 to 2,117 ft^3.

4.5 Discussion

I based the approach used in this research on the rationalization for balancing an inventory to an unbiased total presented by Iles (2009). In essence, he states that any process resulting in the same total as an unbiased estimate is itself unbiased. Though Dr. Iles balanced on the total volume reported from a large-area timber inventory, we, however, balanced on the means reported by the FIA. In this approach, I allowed individual pixel estimates to adjust upward or downward until the remote sensing-based mean volume per acre (ft^3) equalized with the mean derived from the FIA plot measurements.

Table 4.33: Query results from the 0.5-mile radius query to the FIA database, the initial KNN, mean balancing, and the industry-fused methods for the conifer, deciduous, and mixed types.

Measure	Type	Area (ac)	Min	Max	Mean	SD	Volume (million cu.ft)
FIA Db Query	Conifer	0	NA	NA	0	NA	0
Initial KNN	Conifer	299	64	3,621	2,036	747	607
Mean balanced	Conifer	299	91	6,608	2,103	795	627
Industry-fused	Conifer	299	91	7,695	2,252	1,093	672
FIA Db Query	Deciduous	0	NA	NA	0	NA	0
Initial KNN	Deciduous	132	249	3,771	2,168	716	286
Mean balanced	Deciduous	132	253	5,939	2,215	746	292
Industry-fused	Deciduous	132	257	3,906	2,245	742	296
FIA Db Query	Mixed	0	NA	NA	0	NA	0
Initial KNN	Mixed	31	179	3,704	1,734	784	53
Mean balanced	Mixed	31	285	4,816	1,952	945	60
Industry-fused	Mixed	31	0	3,950	2,024	959	62

This inventory of Georgia differentiates itself from other large-area, remote sensing-based inventories in the northeastern United States and abroad (R. McRoberts, Tomppo, and Naesset 2010; Tomppo et al. 2008; R. E. McRoberts et al. 2009) in the manner is addressed. Recommendations for minimizing bias are the incorporation of a weighting factor during the nearest neighbor process (Katila, 2006; McRoberts, 2009), generalization or segmentation (Hyvonen, Pekkarinen, & Tuominen, 2005; Woodcock, Macomber, Pax-lenney, & Cohen, 2001) and the careful selection of the optimal K (McRoberts et al., 2002) and method of estimation (Labrecque, Fournier, Luther, & Piercey, 2006). I on the other hand, accept the statistical integrity of the FIA's large-area reports and conform our measurements.

Scaling estimates based solely on the ESD to attain equalization decreased the local accuracy of our stand-level volume per acre (ft^3) estimates (Fig.4.8A and 4.8B). Root-mean squared error decreased by 4% and MAE by 7% (Tab. 4.26) when compared to the initial KNN estimates. However, after incorporating the small area forest inventory, our local accuracy increased by nearly 2.5 times, all the while, maintaining large area conformity with the FIA (Tab. 4.25).

Neither mean balancing (Fig. 4.9)(Tab. 4.27) nor the industry-fusion process (Fig. 4.10) (Tab. 4.28) increased the accuracies of stand basal area or age. Stand-level basal area and age were most accurately estimated using the initial KNN estimation process.

Several issues requiring further assessment were identified throughout this research. I did not explore balancing to the total volume. My rationalization for balancing to the mean is the fact that volume per acre is invariant to total area. On the other hand, total volume is a product of forestland area and, unlike volume per acre, fluctuates as that area changes. However, total volume is the measure the FIA reports, so the issue should be addressed

Second, there is room for more complete utilization of the small-area measurements. This study only leveraged the information from our industry partners within their stand boundaries. The high resolution ground information, however, can be used for estimates across the entire scene. For instance, Sivianpillai (2006) used similar high resolution forest measurements in conjunction with remote sensing and multivariate regression to estimate age and density for a site in eastern Texas and Meng (2009) used high resolution forest information and satellite imagery with geostatistical techniques for a forest Several issues requiring further assessment were identified throughout this research. I did not explore balancing to the total volume. My rationalization for balancing to the mean is the fact that volume per acre is invariant to total area. On the other hand, total volume is a product of forestland area and, unlike volume per acre, fluctuates as that area changes. However, total volume is the measure the FIA reports, so the topic should be explored.

4.6 Conclusions

Natural resource managers have a growing amount of data available for incorporation into their decision-making and management processes. Regardless of the source, whether it is the product of a small forest inventory designed for a locally accurate estimate, a report based on sparsely located plots adequate for large area approximations, or even if it is a bit of information your foreman 'knows' is true and, for that reason alone, must be included in the analysis, they all contain useful bits of information. We used the total balancing concept to assimilate those seemingly unrelated,

yet useful bits of information into our high resolution, spatially explicit inventory for the state of Georgia. The inventory retains the FIA's unbiased nature across large areas for volume per acre (ft^3), however, unlike the FIA, our inventory also maintains the local accuracies provided by our forest industry partners.

4.7 REFERENCES

[1] Bedard, F., G. Reichert, R. Dobbins, and I. Trepanier. 2008. *Evaluation of Segment-based Gap-filled Landsat ETM+ SLC-off Satellite Data for Land Cover Classification in Southern Saskatchewan, Canada.* International Journal of Remote Sensing 29 (7) (April): 20412054. doi:10.1080/01431160701281064.

[2] Chen, Jin, Xiaolin Zhu, James E. Vogelmann, Feng Gao, and Suming Jin. 2011. *A Simple and Effective Method for Filling Gaps in Landsat ETM+ SLC-off Images.* Remote Sensing of Environment 115 (4) (April 15): 10531064. doi:10.1016/j.rse.2010.12.010.

[3] Chen, M, and Josef Cihlar. 2000. *Retrieving Leaf Area Index of Boreal Conifer Forests Using Landsat TM Images* 162 (August 1995): 153162.

[4] Cieszewski, CJ and RC Lowe. 2008. Generic gapfilling method for reconstructions of missing data using KNN approach on multitemporal scene pairings. Fiber Supply Assessment Technical Report 2008-1. WSFNR, University of Georgia. Athens GA 30602. 2p.

[5] Cieszewski, CJ., RC Lowe. *Biomass InFORM (Interactive Fast Online Reports & Maps).* GrowthAndYield. 2007. Available At: http://www.growthandyield.com/maps/InFORMB/GA/state_tmplate.html

[6] Curran, Paul J, Jennifer L Dungan, and Henry L Gholz. 1992. *Seasonal LAI in Slash Pine Estimated with Landsat TM.* Remote Sensing of Environment 39 (1) (January): 313. doi:10.1016/0034-4257(92)90136-8.

[7] EPA (United States Environmental Protection Agency), 2000. Multi-Resolution Land Characteristics Consortium (MRLC) database. Available At: http://www.epa.gov/mrlcpage.

[8] Gillis, Mark D. 2001. *Canada's National Forest Inventory (Responding to Current Information Needs).* Environmental Monitoring and Assessment 67 (1) (February): 121129. doi:10.1023/A:1006405820244.

[9] Gjertsen, Amt Kristian. 2007. *Accuracy of Forest Mapping Based on Landsat TM Data and a kNN-based Method.* REMOTE SENSING OF ENVIRONMENT 110 (4) (October): 420430. doi:10.1016/j.rse.2006.08.018.

[10] Howard, By Stephen M, James M Lacasse, and S M Howard. 2004. *An Evaluation of Gap-Filled Landsat SLC-Off Imagery for Wildland Fire Burn Severity Mapping.* Photogrammetric Engineering Remote Sensing 70 (August): 877880.

[11] Hyvonen, Pekka, Anssi Pekkarinen, and Sakari Tuominen. 2005. *Segment-level Stand Inventory for Forest Management*. Scandinavian Journal of Forest Research 20 (1) (February 1): 7584. doi:10.1080/02827580510008220.

[12] Jin, Suming, Collin Homer, Limin Yang, George Xian, Joyce Fry, Patrick Danielson, and Philip A. Townsend. 2013. *Automated Cloud and Shadow Detection and Filling Using Two-date Landsat Imagery in the USA*. International Journal of Remote Sensing 34 (March 10): 3741. doi:10.1080/01431161.2012.720045.

[13] Katila, Matti. 2006. *Correcting Map Errors in Forest Inventory Estimates for Small Areas.* In Forest Inventory, 40:225233. http://www.springerlink.com/content/gk067k0u773555g5.

[14] Labrecque, S., R.A. Fournier, J.E. Luther, and D. Piercey. 2006. *A Comparison of Four Methods to Map Biomass from Landsat-TM and Inventory Data in Western Newfoundland.* Forest Ecology and Management 226 (1-3) (May): 129144. doi:10.1016/j.foreco.2006.01.030.

[15] Lowe, RC, CJ Cieszewski. 2009. *KNN Gapfill method to repair SLC-off Landsat ETM+ satellite imagery.* Presentation given at the 7th Southern Forestry and Natural Resources GIS Conference, held on Dec. 7-9, 2009, Athens, Georgia.

[16] Maselli, F., and M. Chiesi. 2006. *Integration of Multi-source NDVI Data for the Estimation of Mediterranean Forest Productivity.* International Journal of Remote Sensing 27 (1) (January 10): 5572. doi:10.1080/01431160500329486.

[17] Maxwell, and S. 2004. *Filling Landsat ETM Plus SLC-off Gaps Using a Segmentation Model Approach* 70 (10): 11091111.

[18] Maxwell, S. K., G. L. Schmidt, J. C. Storey, Maxwell, K. S., Schmidt, L. G., Storey, and C. J. 2007. *A Multi-scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-off Images.* International Journal of Remote Sensing 28 (23) (January): 53395356. doi:10.1080/01431160601034902.

[19] McRoberts, R. 2009. *Diagnostic Tools for Nearest Neighbors Techniques When Used with Satellite Imagery.* Remote Sensing of Environment 113 (3) (March 16): 489499. http://linkinghub.elsevier.com/retrieve/pii/S0034425708003052.

[20] McRoberts, R., and E. Tomppo. 2007. *Remote Sensing Support for National Forest Inventories.* Remote Sensing of Environment 110 (4) (October): 412419.

[21] McRoberts, R.E. Ronald E, Daniel G Wendt, M.D. Mark D Nelson, M.H. Mark H Hansen, McRoberts R.E., Nelson M.D., and Wendt D.G. 2002. *Stratified Estimation of Forest Area Using Satellite Imagery, Inventory Data, and the k-Nearest Neighbors Technique.* Remote Sensing of Environment 81 (2-3) (July): 457468. doi:10.1016/S0034-4257(02)00064-0. http://linkinghub.elsevier.com/retrieve/pii/S0034-4257(02)00064-0.

[22] McRoberts, Ronald E., Erkki Tomppo, Klemens Schadauer, Claude Vidal, Goran Stahl, Gherardo Chirici, Adrian Lanz, Emil Cienciala, Susanne Winter, and W. Brad Smith. 2009. *Harmonizing National Forest Inventories.* Journal of Forestry (June): 179187. http://www.ingentaconnect.com/content/saf/jof/2009/00000107/00000004/art00005.

[23] McRoberts, Ronald, Erkki Tomppo, and Erik Nsset. 2010. *Advances and Emerging Issues in National Forest Inventories.* Scandinavian Journal of Forest Research 25 (4) (August): 368381. doi:10.1080/02827581.2010.496739.

[24] Meng, Q, C Cieszewski, and M Madden. 2009. *Large Area Forest Inventory Using Landsat ETM+: A Geostatistical Approach.* ISPRS Journal of Photogrammetry and Remote Sensing 64 (1) (January): 2736. http://linkinghub.elsevier.com/retrieve/pii/S0924271608000609.

[25] Meng, Qingmin, Bruce E Borders, Chris J Cieszewski, and Marguerite Madden. 2009. *Closest Spectral Fit for Removing Clouds and Cloud Shadows.* Photogrammetric Engineering & Remote Sensing 75 (5): 569576.

[26] NARSAL (Natural Resources Spatial Analysis Laboratory), 2001. Georgia Land Use Trends (GLUT) project data. Athens: University of Georgia Institute of Ecology.

[27] Reese, Heather, Mats Nilsson, Tina Granqvist Pahln, Olle Hagner, Steve Joyce, Ulf Tingelf, Mikael Egberth, and Hkan Olsson. 2010. *Countrywide Estimates and Data Satellite Using Inventory Forest of Forest Data Field Variables the From.* Environment.

[28] Remmel, T. K., F. Csillag, S. Mitchell, and M. A. Wulder. 2005. *Integration of Forest Inventory and Satellite Imagery:a Canadian Status Assessment and Research Issues.* Forest Ecology and Management 207 (3) (March): 405428. doi:10.1016/j.foreco.2004.11.023. http://linkinghub.elsevier.com/retrieve/pii/S0378112704008205.

[29] Sader, Steven A, Robert B Waide, and William T Lawrence. 1989. *Tropical Forest Biomass and Successional Age Class Relationships to a Vegetation Index Derived from Landsat TM Data.* Biomass 156: 143156.

[30] Sivanpillai, R, C Smith, R Srinivasan, M Messina, and X Wu. 2006. *Estimation of Managed Loblolly Pine Stand Age and Density with Landsat ETM+ Data.* Forest Ecology and Management 223 (1-3) (March): 247254. doi:10.1016/j.foreco.2005.11.013. http://linkinghub.elsevier.com/retrieve/pii/S0378112705007140.

[31] Tomppo, E, H Olsson, G Stahl, M Nilsson, O Hagner, and M Katila. 2008. *Combining National Forest Inventory Field Plots and Remote Sensing Data for Forest Databases.* Remote Sensing of Environment 112 (5) (May): 1982-1999. http://linkinghub.elsevier.com/retrieve/pii/S0034425708000242.

[32] Trotter, C. M., J. R. Dymond, and C. J. Goulding. 1997. *Estimation of Timber Volume in a Coniferous Plantation Forest Using Landsat TM.* International Journal of Remote Sensing 18 (10) (July): 22092223. doi:10.1080/014311697217846.

[33] Turner, David P, Warren B Cohen, Robert E Kennedy, Karin S Fassnacht, and John M Briggs. 1999. *Relationships Between Leaf Area Index and Landsat TM Spectral Vegetation Indices Across Three Temperate Zone Sites* 68 (April 1998): 5268.

[34] USGS (United States Geological Survey), 2003. *Preliminary Assessment of Landsat 7 ETM+ Data Following Scan Line Corrector Malfunction.* http://landsat.usgs.gov/documents/SLC_off_Scientific_Usability.pdf (accessed 3/4/2010)

[35] Vieira, I. 2003. *Classifying Successional Forests Using Landsat Spectral Properties and Ecological Characteristics in Eastern Amaznia.* Remote Sensing of Environment 87 (4) (November 15): 470481. doi:10.1016/j.rse.2002.09.002.

[36] Wilson, Emily Hoffhine, and Steven a Sader. 2002. *Detection of Forest Harvest Type Using Multiple Dates of Landsat TM Imagery.* Remote Sensing of Environment 80 (3) (June): 385396. doi:10.1016/S0034-4257(01)00318-2.

[37] Woodcock, Curtis E, Scott A Macomber, Mary Pax-lenney, and Warren B Cohen. 2001. *Monitoring Large Areas for Forest Change Using Landsat: Generalization Across Space , Time and Landsat Sensors.* Remote Sensing of Environment 78: 194 203.

[38] Xiao, Xiangming, Stephen Boles, Jiyuan Liu, Dafang Zhuang, and Mingliang Liu. 2002. *Characterization of Forest Types in Northeastern China, Using Multi-temporal SPOT-4 VEGETATION Sensor Data.* Remote Sensing of Environment 82 (2-3) (October): 335348. doi:10.1016/S0034-4257(02)00051-2.

[39] Zhu, Xiaolin, Desheng Liu, and Jin Chen. 2012. *A New Geostatistical Approach for Filling Gaps in Landsat ETM+ SLC-off Images.* Remote Sensing of Environment 124 (September): 4960. doi:10.1016/j.rse.2012.04.019.

CHAPTER 5

CONCLUSIONS

Under a multi-use sustainable management regime, the provision of timely, reliable, and accurate information about forests, their forested ecosystems and adjacent areas is essential for maintaining their sustained productivity and ecological balance. This is especially important where forests tend to be fast growing and changing, highly fragmented in area and ownership, and where there are multiple demands for their use, such as those in the southeastern United States. Competition for the use of the forests throughout the southeast due to population growth and migration, urban and rural development, and climate change, to name a few, are likely to increase. These issues are compounded since these stressors, their magnitude, focus, and their effect on the current and future ecosystems will vary depending on geography, thus, our responses to them will require regional and local strategies to address these management issues.

In this Dissertation, I have presented three research projects in an attempt to demonstrate various methods of multi-source data fusion based on the Forest Inventory and Analysis (FIA), remotely sensed imagery, and various other sources for improving the FIA inventory data and the level of detail they describe for different locations in the state of Georgia. In the first study, I addressed an issue with the usability of the satellite imagery from the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor. Since the scan line corrector failure in 2003, information in approximately 20% of the total pixels were absent and a majority of the remaining data were not suitable for assessment across a continuous landscape. I used a nearest neighbor approach based on multi-temporal and multi-sensor data for their repair. Understandably, the best data pairs are ones captured close together in time. However, possibly not as apparent, results suggest matching seasons is more important than nearness in acquisition date. The next best choice in image pairs are ones captured in the same season a year before or after.

The second study demonstrates an approach to small area assessments throughout the state using satellite imagery. In an attempt to reduce variability in the data, I used various methods to distribute FIA plot data segmented ETM+ imagery in a manner that accurately distributed the field measures throughout the landscape. I demonstrated an application of the segmented inventory by intersecting it with a series of road beautifying buffers and stream side management zones corresponding to Georgia's Best Management Practices. After long-term analysis, we concluded the magnitude of mandatory stream side protection and visually pleasing roadside buffers will delay harvest of mature forests.

The third study is an implementation of the framework for a higher resolution forest inventory described in Cieszewski et al. (2005) and Iles (2009) in which data from various sources with incompatible variances were combined to create a pixel-level forest inventory of Georgia. The K-nearest neighbor methodology was once again employed to distribute the FIA field measurements across multiple Landsat 5 Thematic Mapper satellite images. This study differentiates itself from other large-area, remote sensing-based inventories in the northeastern United States and abroad (R. McRoberts, Tomppo, and Naesset 2010; R. McRoberts and Tomppo 2007; Tomppo et al. 2008; R. E. McRoberts et al. 2009) in the manner in which these incompatibilities are addressed. We simply accept the statistical integrity of the FIA's large-area reports and conform our measurements for the same area to theirs. There is, however, minimal gain when the large-area means are simply equalized proportionally. However, by incorporating locally-accurate cruise information from our industrial sponsors, precision rose by 2.5 times

In these three studies, we were able to provide more detailed information with a higher resolution than we could achieve otherwise by leveraging the FIA inventory in a manner for which it was not designed. Using the approaches described herein, we provide the natural resource managers spatially explicit, high resolution and statistically valid information across large areas to address the critical issues of the day. Nevertheless, there are areas that should be explored further. The Landsat 5 satellite was decommissioned in January, 2013, so alternative companion data sources for the Landsat 7 gapfilling described in Chapter 2, and natural resource modeling pro-

cesses described in Chapters 3 and 4 need to be explored. The next generation Landsat satellite, the Landsat Data Continuity Mission spacecraft (LCDM), was launched in February of 2013. Reports from the USGS suggest the testing and calibration phases should be completed near the beginning of the summer (2013) and images delivered soon thereafter. The LCDM has similar spatial and spectral characteristics to Landsat 5 and 7, so expectations are it will be a suitable replacement.

5.1 REFERENCES

[6] Cieszewski, C.J., K. Iles, R.C. Lowe, and M. Zasada. 2005. *Proof of concept for an approach to a finer resolution inventory.* In 2003 Proc. of the Fifth Annual Forest Inventory and Analysis Symposium, McRoberts, R.E., Reams, G.A., Van Deusen, P.C., and McWilliams, W.H. (eds.). November 18 20, 2003, New Orleans, LA. Gen. Tech. Rep. WO-69. US For. Serv., Washington, DC. 222 p.

[14] Iles, K. 2009. *Total-Balancing an inventory: A method for unbiased inventories using highly biased non-sample data at variable scales.* Math. Comput. For. Nat. Resour. Sci. 1(1):10 13. Available online at mcfns.com/index.php/Journal/ article/view/MCFNS-1:10/18; last accessed April 13, 2009.

[17] McRoberts, R., and E. Tomppo. 2007. *Remote Sensing Support for National Forest Inventories.* Remote Sensing of Environment 110 (4) (October): 412419. doi:10.1016/j.rse.2006.09.034.

[26] Tomppo, E, H Olsson, G Stahl, M Nilsson, O Hagner, and M Katila. 2008. *Combining National Forest Inventory Field Plots and Remote Sensing Data for Forest Databases.* Remote Sensing of Environment 112 (5) (May): 19821999. doi:10.1016/j.rse.2007.03.032.

[18] McRoberts, Ronald E., Erkki Tomppo, Klemens Schadauer, Claude Vidal, Goran Stahl, Gherardo Chirici, Adrian Lanz, Emil Cienciala, Susanne Winter, and W. Brad Smith. 2009. *Harmonizing National Forest Inventories.* Journal of Forestry (June): 179187.

[45] McRoberts, R.E., E.O. Tomppo, and E. Naesset. 2010. Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research 25(4):368-381.