The Influence of Multi-Season Imagery on Models of Canopy Cover: A Case Study

Christopher R. King
M.S. Candidate – University of Tennessee-Knoxville
Forestry Tech., USFS Southern Research Station

Co-Authors:
John W. Coulston, USFS, SRS-FIA
Dennis M. Jacobs, USFS, SRS-FIA
Ivey C. Elmore, USFS, SRS-FIA

Description

- Statistical comparison of random forest models of percent-tree canopy cover developed from multi-season vs. leaf-on only Landsat imagery
  - Canopy Cover: Why it matters
  - Existing Research: What was lacking
  - Methods
  - Results
  - Discussion

Canopy Cover

- "The area covered by the vertical projection of tree crowns." (Jennings 1999)
  - A Primary Component of Ecosystems:
    - Habitat Suitability
    - Fire Behavior
    - Aesthetics
    - Carbon Dynamics
    - Forest Management

Canopy Cover

- Additionally, percent canopy cover present is used in creating:
  - Forest Land-Use Definitions
  - Forest Land Cover Definitions

Canopy Cover

- Quantifying Canopy Cover Spatially
  - Ecosystem Monitoring (broad-scale)
  - Natural Resource Management

Existing Research

- Researchers have developed empirical models of tree canopy cover to produce geospatial products.
  - For subpixel models, percent tree canopy cover estimates (derived from fine-scale imagery) serve as the response variable.
  - The explanatory variables are developed from reflectance values and derivatives, elevation and derivatives, and other ancillary data.

Existing Research

- Lack of guidance in the literature regarding the use of **leaf-on only imagery** vs. **multi-season imagery** for the explanatory variables.


Existing Research

- Available Literature includes examples suggesting that multi-season imagery is appropriate…
  - Lopez et al. 2001
  - Hansen et al. 2003

- And others suggesting that only single-season imagery is appropriate…
  - Carreiras et al. 2006
  - Sen et al. 2011

The Question:

- Does the inclusion of multi-season imagery as an explanatory variable significantly improve empirical models of percent tree canopy cover?
- The research objective was to answer this and provide guidance as to where the results are relevant.

Methods

- We compared models developed from leaf-on only Landsat imagery with models developed from multi-season imagery for a study area in Georgia, US.
  - Study Area
  - Sampling Methods
  - Explanatory data
  - Statistics

Study Area

- The study area was approximately the size of one Landsat scene.
- It covered central and northern Georgia in the southeastern United States, and was specifically selected to capture the south to north environmental gradient.
- The Piedmont was the dominant (77 percent) ecoregion (USEPA, 2011) in the study area.

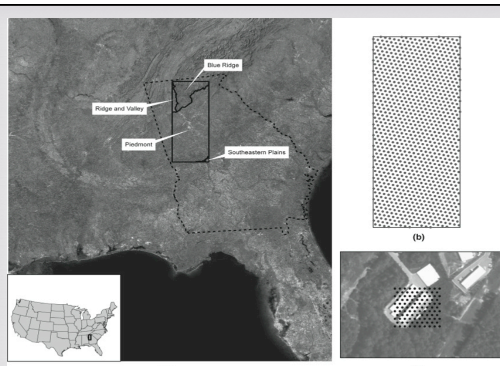
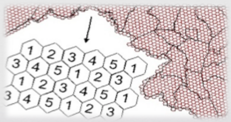


Figure 1. (a) Georgia study area and (b) and (c) sample design. One of the four systematic samples (PSU) is displayed in (b). For each PSU, a 105 point secondary sampling unit was used for photo interpretation of canopy cover (c).

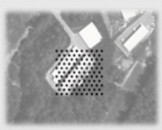
Methods

- Percent tree canopy cover was estimated for 4,125 sample locations (PSUs) across the study area and these estimates served as the response data.
- Sample locations: Identified based on a 4X intensification of the USDA Forest Service FIA sampling grid using the procedures described by White *et al*. (1992).

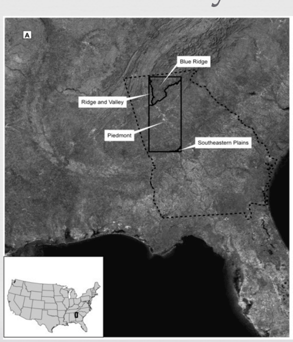


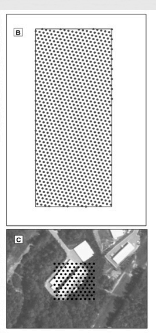
Methods

- At each PSU, a 105 point triangular-grid that filled a 90m by 90m (0.81 ha) area served as the basis for photo-interpretation.
- Each of the 105 points was manually interpreted as either "tree canopy" or "no tree canopy" using leaf-on 2009 NAIP imagery.



Study Area

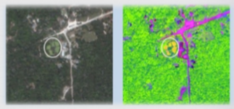




Explanatory Data

- Landsat-5 data and derivatives
  - (NDVI, tasseled cap)

True Color — Tasseled Cap



- Digital elevation data and derivatives
  - (slope, aspect, sine and cosine of aspect, compound topographic index)
- 2001 NLCD land cover data

Explanatory Data

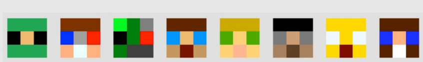
- Six Landsat-5 scenes were downloaded from MRLC (Multi-Resolution Land Characteristics) – 2011

Landsat-5 Acquisition Dates for Leaf-On, Leaf-Off, & Spring

Image	Date
Landsat-5 (path 19 row 36)	
leaf-on	24-Jul-08
leaf-off	16-Jan-09
spring	9-Apr-10
Landsat-5 (path 19 row 37)	
leaf-on	9-Aug-08
leaf-off	16-Jan-09
spring	9-Apr-10

Explanatory Data

- Explanatory variables for modeling were developed by calculating the **mean** and **standard deviation** of each variable for each PSU.
  - This was done using 3x3 pixel window focal statistics.
- In total, there were 73 explanatory variables.



Statistics

- We used the "random forest" algorithm (Breiman, 2001) to construct empirical models of percent tree canopy cover.
- Uses bootstrap sampling to develop multiple models and improve prediction (without replacement)
  - Random = bootstrap sampling of the data
  - Forest = an ensemble of regression trees

Statistics

- For modeling, we used the **R** ver. 2.12 (R Development Core Team, 2010) random forest library (Liaw and Wiener, 2002) to construct empirical models of percent tree canopy cover.
- Three random forest models were developed, each using 25% of the observations.

Statistics

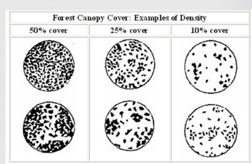
- This was done using the 4x grid, with subsample 4 being a hold-out for model comparison.
  - Subsample 1: multi season model
  - Subsample 2: leaf-on
  - Subsample 3: reduced
  - Subsample 4: hold-out

Statistics

- We performed two principal component analyses:
  - One for standardized Landsat data & derivatives
  - One for standardized elevation data & derivatives
- This retains *n* components that accounted for approximately 90% of the variation.
- Models were then compared using the hold-out dataset.

Results

- Based on photo-interpretation of the 4X sample, the average percent canopy cover (across all 2001 NLCD land cover classes) was 66 percent in the GA study area.
- Land cover types:
  - Ag. – 34% canopy cover
  - Forest – 84% canopy cover
  - Urban – 41% canopy cover



Results

- **Landsat**: Principal components analysis results indicated that 90% of the variance across all 60 variables was explained by the first 10 principal components.
- **Digital elevation models:** PCA results indicated that 90% of variance across the 12 variables was explained by the first 7 components.

Results

- Each component was interpreted and a representative variable was selected.
- The following 10 Landsat variables were retained:
  - Leaf-off TM band 3
  - Standard deviation of spring TM band 3
  - Standard deviation of leaf-off greenness
  - Standard deviation of leaf-on TM band 6
  - Spring NDVI
  - Leaf-on NDVI
  - Standard deviation of spring wetness
  - Standard deviation of spring TM band 4
  - Spring TM band 5
  - Leaf-off brightness

Results

- The following 7 D.E.M. variables were retained:
  - Slope
  - Aspect
  - Sine aspect
  - Standard deviation of slope
  - Standard deviation of aspect
  - Standard deviation of sine aspect
  - Standard deviation of compound topographic index
- These 10 Landsat variables and 7 digital elevation variables, along with 2001 NLCD land cover, served as the explanatory variables for the reduced model.

Model Results

- The empirical models of percent tree canopy cover had similar pseudo R^2's.
- All Three models produced distributions that were statistically different ($p<0.001$) than the observed distribution.
  - Overall, models under-predicted the amount of "no tree" and "100%" canopy cover.

Table 2. Model Fit Statistics for Each Model

Model	RMSE	R2
Leaf-on	15.01	0.81
Multi-season	14.02	0.83
Reduced	14.1	0.83

Model Results

- Overall, models under-predicted the amount of "no tree" and "100%" canopy cover.

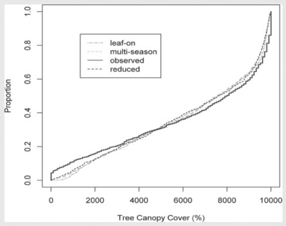


Figure 2. Cumulative distribution of observed tree canopy and predicted tree canopy cover based on the leaf-on, multi-season, and reduced models.

Model Results

- While all three distributions were significantly different from the observed distribution, there was <u>no significant difference</u> (α=0.05) among models.

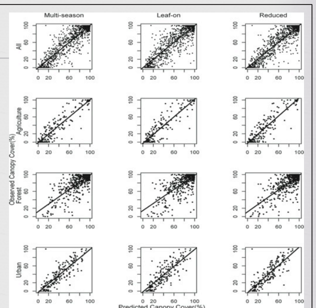


Figure 3. Observed versus predicted canopy percent canopy cover, based on a hold-out dataset, for all land cover classes, agricultural classes, forest classes, and urban classes based on the multi-season model, the leaf-on model, and the reduced model.

Discussion

- The goal of this research was to identify whether using multi-season imagery for explanatory variables resulted in more accurate tree canopy cover models.
  - When models are equally accurate, we generally choose the least complex.
  - The leaf-on model is the simplest in terms of data acquisition, storage, and processing.

Discussion

- We suggest that leaf-on imagery is adequate for the development of empirical models of percent tree canopy cover in the Piedmont of the Southeastern United States.
- We also recommend this model for better efficiency while maintaining accuracy.

Discussion

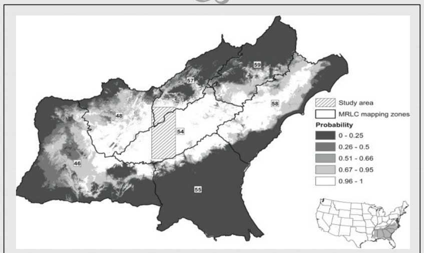


Figure 4. The probability of similar monthly minimum temperature profiles to the study area. The probabilities were from the Wald test which compared Fourier regression parameters of areas outside the study area to the most similar Fourier regression parameters from within the study area.

Acknowledgements

- Co-authors:
  - Dr. John Coulston – FIA, Southern Research Station
  - Dennis Jacobs - FIA, Southern Research Station
  - Ivey Elmore - FIA, Southern Research Station
- Remote Sensing Applications Center – USDA
- USFS – Forest Inventory and Analysis

References

- Breiman, L., 2001. Random Forests, *Machine learning*, 45(1):5–32.
- Brocklebank, J.C., and D.A. Dickey, 1986. *SAS System for Forecasting Time Series*, SAS Institute, Inc., Cary, North Carolina, 240 p.
- Brooks, E.B., V.A. Thomas, R.H. Wynne, and J.W Coulston, 2012. Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis, *Transactions on Geoscience and Remote Sensing*, 50(9):3340–3353.
- Carreiras, J.M.B, J.M.C Pereira, and J.S. Pereira, 2006. Estimation of tree canopy cover in evergreen oak woodlands using remote sensing, *Forest Ecology and Management*, 223:45–53.
- Chavez, P.S., 1996. Image-based atmospheric corrections – Revisited and improved, *Photogrammetric Engineering & Remote Sensing*, 62(9):1025–1036.
- Cochran, W.G., 1977. *Sampling Techniques*, Third edition, John Wiley & Sons, New York, 428 p.
- Coulston, J.W., G.G. Moisen, B.T. Wilson, M.V. Finco, W.B. Cohen, and C.K. Brewer, 2012. Modeling percent tree canopy: A pilot study, *Photogrammetric Engineering & Remote Sensing*, 78(7):715–727.
- Duane, M.V., W.B. Cohen, J.L. Campbell, T. Hudiburb, D.P Turner, and D.L. Weyermann, 2010. Implications of alternative field sampling designs for Landsat-based mapping of stand age and carbon stocks in Oregon forests, *Forest Science*, 56(4):405–416.
- Efron, B., and R.J. Tibshirani, 1993. *An Introduction to the Bootstrap*, Chapman & Hall, CRC, Boca Raton, Florida, 456 p.
- FAO, 2001. *Global Forest Resources Assessment 2000 Main Report*, FAO Forestry Paper 140.
- Franco-Lopez, H., A.R. Ek, and 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.
- Hansen, M.C., R.S. DeFries, J.R.G. Townshend, M. Carroll, C. Dimiceli, and R.A. Sohlberg, 2003. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fi elds algorithm, *Earth Interactions*, 7(010):1–15.
- Harrell, F.E, 2001. *Regression Modeling Strategies*, Springer-Verlag, New York, 568 p.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N.
- Herold, A. McKerrow, J.N. VanDriel, and J. Wickham, 2007. Completion of the 2001 National Land Cover Database for the conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 73(4):337–341.
- Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 National Land Cover Database for the United States, *Photogrammetric Engineering & Remote Sensing*, 70(7):829–840.
- Homer, C.G., and A. Gallant, 2001. Partitioning the conterminous United States into mapping zones for Landsat TM land cover mapping, USGS (last date accessed: 31 January 2013).
- Huang, C., L. Yang, B. Wylie, and C. Homer, 2001. A strategy for estimating tree canopy density using Landsat 7 ETM and high resolution images over large areas, *Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry*, 05–07 November, Denver, Colorado, unpaginated CD-ROM.
- Hwang, T., C. Song, J.M. Vose, and L.E. Band, 2011. Topography mediated controls on local vegetation phenology estimated from MODIS vegetation index, *Landscape Ecology*, 26: 541–556.
- Jennings, S.B., N.D. Brown, and D. Sheil, 1999. Assessing forest canopies and understory illumination: Canopy closure, canopy cover and other measures, *Forestry*, 72(1):59–73.
- Johnson, R.A., and D.W. Wichern, 2002. *Applied Multivariate Statistical Analysis*, Fifth edition, Prentice-Hall, Upper Saddle River, New Jersey, 767 p.
- Liaw, A., and M. Wiener, 2002. Classification and regression by random forest, *R News*, 2(3):18–22.
- MRLC, 2011. Multi-Resolution Land Characteristics Consortium, URL: *http://www.mrlc.gov/*, US Geological Survey, Sioux Falls, South Dakota (last date accessed: 31 January 2013).
- PRISM Climate Group, 2012. PRISM Climate Group, Oregon State
- University, URL: *http://prism.oregonstate.edu*, Oregon State University, Corvallis, Oregon (last date accessed: 31 January 2013).
- R Development Core Team, 2010. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL: *http://www.R-project. org* (last date accessed: 31 January 2013).
- Riemann, R., B.T. Wilson, A. Lister, and S. Parks, 2010. An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data, *Remote Sensing of Environment*, 114:2337–2352.
- Rollins, M.G., and C.K Frame (editors), 2006. *The LANDFIRE Prototype Project: Nationally consistent and locally relevant geospatial data for wildland fire management*, USDA Forest Service, General Technical Report RMRS-GTR-175, Rocky Mountain Research Station, Fort Collins, Colorado, 416 p.
- Satorra, A., and P.M. Bentler, 2001. A scaled difference chi-square test statistic for moment structure analysis, *Psychometrika*, 66(4):507–514.
- Schroeder, T.A., W.B. Cohen, C. Song, M.J. Canty, and Z. Yang, 1006. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon, *Remote Sensing of Environment*, 103:16–26.
- Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration, *International Journal of Remote Sensing*, 6(8):1335–1372. Sen, S., C.E. Zipper, R.H. Wynne, P.F. Donovan, and J.W. Coulston,
- 2011. Using satellite imagery to characterize locations, ages, and woody canopy cover of reclaimed surface mines in Appalachian USA, *Proceedings of the 2011 National Meeting of the American Society of Mining Reclamation*, 11-16 June, Bismarck, North Dakota, unpaginated CD-ROM.
- Song, C., C.E. Woodcock, K.C. Seto, M. Pax-Lenney, and S.A.
- Macomber, 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects, *Remote Sensing of Environment*, 75:230–244.
- Steel, R.G.D., J.H. Torrie, and D.A. Dickey, 1997. *Principles and Procedures of Statistics: A Biometrical Approach*, Third edition, The McGraw-Hill Companies, Inc., New York, 666 p.
- US Department of Agriculture, 2009. National Agriculture Imagery Program, USDA Farm Service Agency, Aerial Photography Field Office, Salt Lake City, Utah (last date accessed: 31 January 2013).
- US Environmental Protection Agency, 2011. Level III ecoregions of the continental United States, Map revised December 2011, (last date accessed: 31 January 2013).
- USGS, 2005. Elevation derivatives for national applications, URL: *http://edna.usgs.gov/*, US Geological Survey, (last date accessed: 31 January 2013).
- USDA, 2012. USDA plant hardiness zone map. US Department of Agriculture (last date accessed: 31 January 2013).
- Walton, J.T., 2008. Sub-pixel urban land cover estimation: Comparing cubist, random forests, and support vector regression, *Photogrammetric Engineering & Remote Sensing*, 74(10):1213–1222.
- White, D., J. Kimerling, and S.W. Overton, 1992. Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartographic and Geographic Information Systems*, 19(1):5–22.
- White, M.A., R.R. Nernani, P.E. Thornton, and S.W. Running. 2002. Satellite evidence of phonological differences between urbanized and rural areas of the eastern United States deciduous broadleaf forest, *Ecosystems*, 5:260–277.

Photo Credits

- http://www.gettyimages.com/detail/photo/multi-colored-aspen-leaves-with-rain-drop-high-res-stock-photography/90053696
- http://www.archatlas.dept.shef.ac.uk/occpaper/Ferrand.php
- USFS, Southern Research Station – Forest Inventory and Analysis
- Coulston, J.W., D.M. Jacobs, C.R. King, and I.C. Elmore 2013. The Influence of Multi-Season Imagery on Models of Canopy Cover: A Case Study, *Photogrammetric Engineering and Remote Sensing*. Vol. 79, No. 5, May 2013, pp. 469-477.
- http://www.blm.gov/or/districts/medford/plans/evans/photos_forest.php
- http://www.reddit.com/r/mindcrack/comments/1iwr4k/
- http://hwwff.cce.cornell.edu/learning3377.html?unit=2&section=Site%20Assessment%20Part%20I%20(Climate)&subsection=Climate%20III