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
Additionally, percent canopy cover present is used in creating:

- Forest Land-Use Definitions
- Forest Land Cover Definitions

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

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

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 150 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)

- 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

<table>
<thead>
<tr>
<th>Image</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landsat-5 (path 19 row 30)</td>
<td>24-Jul-08</td>
</tr>
<tr>
<td>leaf-on</td>
<td>24-Jul-08</td>
</tr>
<tr>
<td>leaf-off</td>
<td>16-Jan-09</td>
</tr>
<tr>
<td>spring</td>
<td>9-Apr-10</td>
</tr>
<tr>
<td>Landsat-5 (path 19 row 37)</td>
<td>9-Aug-08</td>
</tr>
<tr>
<td>leaf-on</td>
<td>9-Aug-08</td>
</tr>
<tr>
<td>leaf-off</td>
<td>16-Jan-09</td>
</tr>
<tr>
<td>spring</td>
<td>9-Apr-10</td>
</tr>
</tbody>
</table>

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.
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

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.

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

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.

Results

- The empirical models of percent tree canopy cover had similar pseudo R²'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>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf-on</td>
<td>15.01</td>
<td>0.81</td>
</tr>
<tr>
<td>Multi-season</td>
<td>14.02</td>
<td>0.83</td>
</tr>
<tr>
<td>Reduced</td>
<td>14.1</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Model Results

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

Model Results

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

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.
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.

Figure 4. The probability of similar monthly minimum temperature profiles in 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.

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


Brooks, E.B., V. A. Thomas, R.H. Wynne, and J.W Coulston, 2012. Fitting the


Breiman, L., 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects,


Weyermann, N. Hossain, C. Larson, N.

International Journal of Remote Sensing

http://hwwff.cce.cornell.edu/learning3377.html?unit=2&section=Site

sen, S., C.E. Zipper, R.H. Wynne, P .F. Donovan, and J.W. Coulston,

http://www.reddit.com/r/mindcrack/comments/1iwr4k/

Coulston, J.W., D.M. Jacobs, C.R. King, and I.C. Elmore 2013. The Influence of

USFS, Southern Research Station – Forest Inventory and Analysis

http://www.mrlc.gov/

http://www.archatlas.dept.shef.ac.uk/occpaper/Ferrand.php


Franco-Lopez, H., A.R. Harrell, F.E, 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects,

Brooks, E.B., V .A. Thomas, R.H. Wynne, and J.W Coulston, 2012. Fitting the


Breiman, L., 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects,


Weyermann, N. Hossain, C. Larson, N.

International Journal of Remote Sensing

http://hwwff.cce.cornell.edu/learning3377.html?unit=2&section=Site

sen, S., C.E. Zipper, R.H. Wynne, P .F. Donovan, and J.W. Coulston,

http://www.reddit.com/r/mindcrack/comments/1iwr4k/

Coulston, J.W., D.M. Jacobs, C.R. King, and I.C. Elmore 2013. The Influence of

USFS, Southern Research Station – Forest Inventory and Analysis

http://www.mrlc.gov/

http://www.archatlas.dept.shef.ac.uk/occpaper/Ferrand.php


Franco-Lopez, H., A.R. Harrell, F.E, 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects,

Brooks, E.B., V .A. Thomas, R.H. Wynne, and J.W Coulston, 2012. Fitting the


Breiman, L., 2001. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects,


Weyermann, N. Hossain, C. Larson, N.

International Journal of Remote Sensing

http://hwwff.cce.cornell.edu/learning3377.html?unit=2&section=Site

sen, S., C.E. Zipper, R.H. Wynne, P .F. Donovan, and J.W. Coulston,