2006

Comparing the Accuracy of Multi-Source Data Integration for Two Supervised Image Classification Methods: Maximum Likelihood and Artificial Neural Network

Hillary Tribby

Daniel Unger
Arthur Temple College of Forestry and Agriculture, Stephen F. Austin State University, unger@sfasu.edu

Follow this and additional works at: https://scholarworks.sfasu.edu/spatialsci_facultypres

Part of the Forest Sciences Commons
Tell us how this article helped you.

Repository Citation
Tribby, Hillary and Unger, Daniel, "Comparing the Accuracy of Multi-Source Data Integration for Two Supervised Image Classification Methods: Maximum Likelihood and Artificial Neural Network" (2006). Faculty Presentations. 24.
https://scholarworks.sfasu.edu/spatialsci_facultypres/24

This Presentation is brought to you for free and open access by the Spatial Science at SFA ScholarWorks. It has been accepted for inclusion in Faculty Presentations by an authorized administrator of SFA ScholarWorks. For more information, please contact cdsscholarworks@sfasu.edu.
Comparing the Accuracy of Multi-Source Data Integration for Two Supervised Image Classification Methods: Maximum Likelihood and Artificial Neural Network

Hilary Tribby, Research Assistant
Daniel Unger, Associate Professor of Spatial Science
Arthur Temple College of Forestry and Agriculture
Stephen F. Austin State University
Nacogdoches, Texas
Land Cover Classification

• Reasons for Creating a Land Cover Map
  – Provide quantitative analysis for use in a GIS
  – Track landscape changes over time
  – Assess landscape components
  – Monitor wildlife habitat
Accuracy Assessment

• Error Matrix
  – Overall accuracy
  – User’s accuracy
  – Producer’s accuracy

• K Hat
  – Measure of percent agreement beyond random chance

• Z Statistic
  – Measure of statistical significance
Satellite Resolution & Accuracy

• Spectral
  – Number and regions of wavelengths
• Spatial
  – Pixel Size
• Radiometric
  – Saturation of reflective energy
• Temporal
  – Frequency of satellite capturing imagery at the same location
## Satellite Sensor Comparison

<table>
<thead>
<tr>
<th></th>
<th>Landsat ETM+</th>
<th>IKONOS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectral</strong></td>
<td>B, G, R, NIR, MIR, FIR</td>
<td>B, G, R, NIR</td>
</tr>
<tr>
<td><strong>Spatial</strong></td>
<td>30 meter</td>
<td>4 meter</td>
</tr>
<tr>
<td><strong>Radiometric</strong></td>
<td>8 bit (0-255)</td>
<td>11 bit (0-2047)</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td>16 days</td>
<td>3-5 days off-nadir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>144 for true nadir</td>
</tr>
</tbody>
</table>
Other Possible Data Sources

• LIDAR Data
  – Surface elevation model
  – Digital terrain model
  – Canopy height model
Temple-Inland’s
Forest Lake Research Facility
Tyler County, Texas
• Introduction
• Literature
• Methods
• Results
• Conclusions
Classification Schemes

• U.S.G.S. 1976
  – Anderson et al. developed continental land cover classes to be used with the Earth Resources Technology Satellite

• T.G.I.C. 1999
  – Developed land cover classes based on the Anderson continental classification scheme, but for Texas cover types
Liu et al. 2003 tested five classes using Landsat TM
- ML overall accuracy 81.35%
- ANN overall accuracy 83.47%

Peddle et al. 1994 used SPOT, texture, and an elevation model
- ML KHAT 0.79
- ANN KHAT 0.96
Additional Classification Sources

• Lidar Derived Components
  – CHM
    • Zimble et al. 2003
      – Single story and multi-story canopy height boundaries were delineated with 97% accuracy.
  – DTM
    • Lee & Shan 2003
      – Found that the inclusion of elevation data increased the accuracy of the land cover maps with KHAT statistics of 0.903 (without) to 0.934 (with).

• Long Wavelength Components
  – Thermal Imagery
    • Southworth 2004
      – Found no statistical difference between land cover maps generated with and without 120 meter thermal data.
• Introduction
• Literature
• Methods
• Results
• Conclusions
Traditional Classification Methods

Image Acquisition

Radiometric Correction

Geometric Correction

Image Classification

Accuracy Assessment
Project Classification Methods

- Landsat ETM+ PCA
- CHM, DTM, and Thermal
- Maximum Likelihood
- Artificial Neural Network
- IKONOS PCA
- CHM
- DTM
- Thermal
- CHM, DTM, and Thermal
Supervised Classification

- Training Points
  - Taken from insitu identification
- Code according to classification scheme
  - Anderson 1976
    - Level I and II
  - Texas Geographic Information Councils (TGIC) 1999
    - Level II and IV
Training Points

Classification of points into two schemes

<table>
<thead>
<tr>
<th>Plot</th>
<th>SAF</th>
<th>Cover</th>
<th>TX 1</th>
<th>TX 2</th>
<th>TX 3</th>
<th>TX 4</th>
<th>TX 5</th>
<th>Description with regeneration species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>Majority brush and shrubbery.</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>Million trees and shrubbery.</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>Mixed forest with coniferous and deciduous</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>do not overlap.</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>Majority brush and shrubbery.</td>
</tr>
</tbody>
</table>

IKONOS image

Landsat ETM+ image
Maximum Likelihood
Artificial Neural Network

- Input Pixel
- Weighting Process
- Hidden Layer
- Weighting Process
- Range
- Forest
- Wetland
## Principal Component Analysis (PCA)

<table>
<thead>
<tr>
<th>Band</th>
<th>Eigenvalue</th>
<th>% Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>674.4268</td>
<td>75.1596</td>
</tr>
<tr>
<td>2</td>
<td>208.8710</td>
<td>23.2771</td>
</tr>
<tr>
<td>3</td>
<td>12.1916</td>
<td>1.3587</td>
</tr>
<tr>
<td>4</td>
<td>2.1364</td>
<td>0.2380</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Band</th>
<th>Eigenvalue</th>
<th>% Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2425.1383</td>
<td>90.3187</td>
</tr>
<tr>
<td>2</td>
<td>139.8310</td>
<td>5.2077</td>
</tr>
<tr>
<td>3</td>
<td>110.0388</td>
<td>4.0981</td>
</tr>
<tr>
<td>4</td>
<td>6.5709</td>
<td>0.2447</td>
</tr>
<tr>
<td>5</td>
<td>2.5431</td>
<td>0.0947</td>
</tr>
<tr>
<td>6</td>
<td>0.9666</td>
<td>0.0360</td>
</tr>
</tbody>
</table>
PCA
(Data layer 1)

Landsat ETM+        PCA        IKONOS        PCA
99.6245 % Variability
July 21, 2000        July 20, 2000
99.7954 % Variability
LIDAR sensors send thousands of light pulses per second. When the pulses return to the altimetry device, distance values are calculated from the point of dispersion measured by the GPS to the reflected surface with orientation data from the IMU/INS. These distances are transformed into elevation values for each point producing a data cloud.
Surface Elevation Model

- Moving windows
  - Neighborhood Point Statistics (Located within the Spatial Analyst Tools)
- Raster Output with a spatial resolution of the search radius, which was 3 meters.
Point shapefiles of terrain were extracted from the data cloud.

TerraPoint, LLC produced level 2 filtered lidar
- Points delineated according to bare earth and not
Digital Terrain Model
(Data layer 2)

- Surface Interpolation
  - Geostatistical Analyst Wizard
- Radial Basis Function
  - Multiquadric
- Lowest Root Mean Square Error
  - 0.2253 meters
- Export as a raster with 4 meter spatial resolution
Canopy Height Model (Data layer 3)

- The DTM was subtracted from the surface elevation model.
- Negative values are surface water returns and interpolation errors. Over water the DTM overestimated river elevation.
Thermal Data
(Data layer 4)

• Landsat ETM+
  – 60 meter spatial resolution
  – May create ambiguity
  – Interference may mask true thermal properties
Maps

- **40 Maps**
  - Phase I - IKONOS T.G.I.C. Level 4
    - 10 maps
  - Phase II - Landsat ETM+ U.S.G.S. Level 2
    - 10 maps
  - Phase III - IKONOS T.G.I.C. Level 2
    - 10 maps
  - Phase IV - Landsat ETM+ U.S.G.S. Level 1
    - 10 maps
Accuracy Assessment

• Using QuickBird imagery (November 24, 2002) with the highest satellite spatial resolution of Forest Lake, 900 simple random points were generated and coded to the differing levels of land cover classes for each satellite.

• The results of which were recorded in an error matrix and tested for significance.
Reference Points

- Landsat ETM+ Random Points
- IKONOS Random Points

Legend:
- **Study Area Boundary**
- **County Line**
## Error Matrix

### Reference Data

<table>
<thead>
<tr>
<th></th>
<th>Forest</th>
<th>Range</th>
<th>Water</th>
<th>Wetland</th>
<th>Total</th>
<th>User’s %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest</td>
<td>463</td>
<td>93</td>
<td>2</td>
<td>36</td>
<td>594</td>
<td>77.95%</td>
</tr>
<tr>
<td>Range</td>
<td>29</td>
<td>127</td>
<td>2</td>
<td>6</td>
<td>164</td>
<td>77.44%</td>
</tr>
<tr>
<td>Water</td>
<td>2</td>
<td>5</td>
<td>21</td>
<td>2</td>
<td>30</td>
<td>70.00%</td>
</tr>
<tr>
<td>Wetland</td>
<td>50</td>
<td>21</td>
<td>1</td>
<td>40</td>
<td>112</td>
<td>35.71%</td>
</tr>
<tr>
<td>Total</td>
<td>544</td>
<td>246</td>
<td>26</td>
<td>84</td>
<td>900</td>
<td></td>
</tr>
</tbody>
</table>

Produce’s % 85.11% 51.63% 80.77% 47.62% 72.33%
Accuracy Assessment

- **Error Matrix**
  - Overall accuracy: 72.33%
  - User’s accuracy: 35.71 - 77.95%
  - Producer’s accuracy: 47.62 - 85.11%

- **KHAT**
  - Value of 0.4864 indicates the map is 48.64% better than random chance assignment of pixels.

- **Z Statistic**
  - Z value of 20.6836
  - Reject the null hypothesis that KHAT = 0
  - Indicates that map is statistically better than random chance assignment.
Phase I

IKONOS
T.G.I.C.  Level 4

Land Cover Map of Forest Lake, Texas

- Classification Method: Maximum Likelihood
- Classification System: T.G.I.C. Level 4
- Satellite Imagery: IKONOS June 7, 2000
- Data Source: PCA
- Post Process: None

Land Cover Classes:
- Tall Grass
- Swamp
- Shrub Wetland
- Short Grass
- River
- Mixed Woodland
- Mixed Shrub
- Mixed Forest
- Marsh
- Lake
- Forest Wetland
- Evergreen Woodland
- Evergreen Shrub
- Evergreen Forest
- Deciduous Woodland
- Deciduous Shrub
- Deciduous Forest

Scale: 1:52,000

Meters

0 500 1,000
## Phase I

### IKONOS T.G.I.C. Level 4

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Accuracy</th>
<th>KHAT</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>36.89%</td>
<td>0.3038</td>
<td>26.5027</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>20.89%</td>
<td>0.1503</td>
<td>17.2351</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>19.44%</td>
<td>0.1319</td>
<td>14.6103</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>35.78%</td>
<td>0.2930</td>
<td>28.7369</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>22.67%</td>
<td>0.1503</td>
<td>15.1933</td>
</tr>
<tr>
<td><strong>ANN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>23.67%</td>
<td>0.1891</td>
<td>24.6298</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>28.22%</td>
<td>0.1509</td>
<td>13.9583</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>17.89%</td>
<td>0.1192</td>
<td>15.2786</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>34.00%</td>
<td>0.2670</td>
<td>26.0242</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>13.67%</td>
<td>0.0876</td>
<td>12.5116</td>
</tr>
</tbody>
</table>
Phase II

Landsat ETM+
U.S.G.S. Level 2

Land Cover Map of Forest Lake, Texas

Figure: A-24
Classification Method: Maximum Likelihood
Classification System: U.S.G.S. LULC Classification Level 2
Satellite Imagery: Landsat ETM+ July 20, 2000
Data Source: PCA and Thermal
Post Process: None

Land Cover Class:
- Wetland non forest
- Wetland Forest
- Shrub and Brush
- River
- Mixed Rangeland
- Mixed Forest
- Lake
- Herbaceous
- Evergreen Forest
- Deciduous Forest

Legend:
1:52,000
0 500 1,000 Meters

Map showing land cover classes in Forest Lake, Texas.
## Phase II

**Landsat ETM+ U.S.G.S. Level 2**

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Accuracy</th>
<th>KHAT</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>54.00%</td>
<td>0.4445</td>
<td>31.8148</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td><strong>55.56%</strong></td>
<td>0.4628</td>
<td>33.0192</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>55.11%</td>
<td>0.4609</td>
<td>33.2561</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td><strong>57.56%</strong></td>
<td>0.4868</td>
<td>34.9607</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td><strong>56.67%</strong></td>
<td>0.4731</td>
<td>33.6354</td>
</tr>
<tr>
<td><strong>ANN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>53.56%</td>
<td>0.3966</td>
<td>26.5374</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>34.22%</td>
<td>0.1909</td>
<td>13.8224</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td><strong>30.33%</strong></td>
<td>0.1195</td>
<td>8.6447</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>37.56%</td>
<td>0.1783</td>
<td>12.1076</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>37.67%</td>
<td>0.2257</td>
<td>15.7683</td>
</tr>
</tbody>
</table>
Phase III

IKONOS
T.G.I.C.  Level 2

Land Cover Map of Forest Lake, Texas

- Figure: A-42
- Classification Method: Maximum Likelihood
- Classification System: T.G.I.C. Level 2
- Satellite Imagery: IKONOS June 7, 2000
- Data Source: PCA and CHM
- Post Process: None

Land Cover Class:
- Woody Wetland
- Woodland
- Water
- Shrubland
- Natural Herbaceous
- Forested
- Emergent Wetland

Scale: 1:52,000

Meters
0 500 1,000
## Phase III

### IKONOS T.G.I.C. Level 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Accuracy</th>
<th>KHAT</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>48.44%</td>
<td>0.3126</td>
<td>19.2137</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td><strong>57.22%</strong></td>
<td>0.4470</td>
<td>27.2018</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>48.56%</td>
<td>0.3257</td>
<td>19.7464</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>45.77%</td>
<td>0.2265</td>
<td>13.7522</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td><strong>49.67%</strong></td>
<td>0.3855</td>
<td>26.4058</td>
</tr>
<tr>
<td><strong>ANN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td><strong>29.40%</strong></td>
<td>0.1351</td>
<td>14.2971</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>35.44%</td>
<td>0.1813</td>
<td>14.9880</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>30.44%</td>
<td>0.1804</td>
<td>17.4799</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>34.86%</td>
<td>0.1451</td>
<td>9.2679</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>40.98%</td>
<td>0.2263</td>
<td>13.5573</td>
</tr>
</tbody>
</table>
Phase IV

Landsat ETM+
U.S.G.S. Level 1

Land Cover Map of Forest Lake, Texas

Figure: A-85
Classification Method: Maximum Likelihood
Classification System: U.S.G.S. LULC Classification Level 1
Satellite Imagery: Landsat ETM+ July 20, 2000
Data Source: PCA, CHM, DTM and Thermal
Post Process: None

Land Cover Class
- Wetland
- Water
- Rangeland
- Forest

1:52,000

[Map of Forest Lake, Texas with land cover classification]
# Phase IV

## Landsat ETM+ U.S.G.S. Level 1

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Accuracy</th>
<th>KHAT</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>68.67%</td>
<td>0.4134</td>
<td>19.3566</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>70.44%</td>
<td>0.4443</td>
<td>20.4265</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>72.00%</td>
<td>0.4840</td>
<td>20.8969</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>69.67%</td>
<td>0.4355</td>
<td>20.0346</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>72.33%</td>
<td>0.4864</td>
<td>20.6836</td>
</tr>
<tr>
<td><strong>ANN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>64.89%</td>
<td>0.4328</td>
<td>20.1101</td>
</tr>
<tr>
<td>PCA and CHM</td>
<td>42.89%</td>
<td>0.1984</td>
<td>11.1062</td>
</tr>
<tr>
<td>PCA and DTM</td>
<td>39.56%</td>
<td>0.0813</td>
<td>4.3250</td>
</tr>
<tr>
<td>PCA and Thermal</td>
<td>26.00%</td>
<td>0.0828</td>
<td>6.3673</td>
</tr>
<tr>
<td>PCA, CHM, DTM, and Thermal</td>
<td>64.33%</td>
<td>0.3287</td>
<td>12.1545</td>
</tr>
</tbody>
</table>
• Introduction
• Literature
• Methods
• Results
• Conclusions
Conclusions

• Multiple Sources of data did not statistically increase land cover classification accuracy consistently.
• Maximum Likelihood performed statistically better than the Artificial Neural Networks consistently.
Conclusions

- Landsat ETM+ classifications performed statistically better than IKONOS.
- Lower classification scheme levels performed similarly for Landsat, but statistically better for IKONOS.
Questions and Comments