Inventory & Monitoring
Remote Sensing
Report

Analysis of Canopy Cover and Impervious Surface Cover of Zone 41

Bonnie Ruefenacht, Alex Hoppus, Jule Caylor
Remote Sensing Applications Center
Salt Lake City, Utah

David Nowak, Jeff Walton
Northeast Research Station
Syracuse, New York

Limin Yang, Collin Homer
US Geological Survey
EROS Data Center
Sioux Falls, South Dakota

Greg Koeln
Earth Satellite Corporation
Rockville, Maryland

Report Prepared for:
Inventory & Monitoring Steering Committee
Bob Simonson
San Dimas Technology & Development Center
San Dimas, California
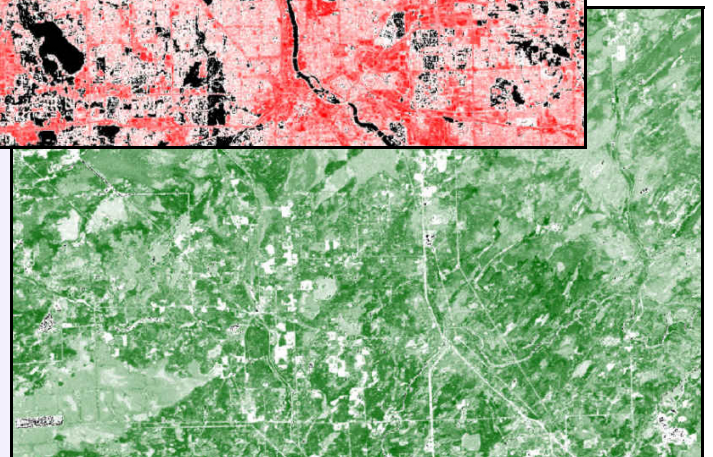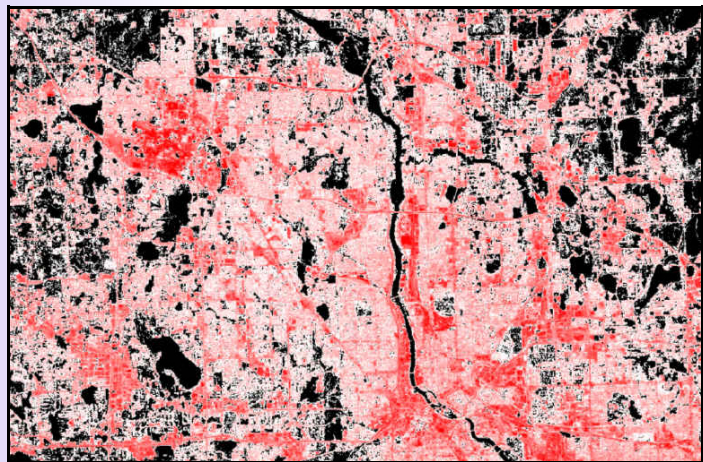
Foreword

This report was initiated and funded by the Inventory and Monitoring Technology Development (IMTD) Steering Committee. This committee was chartered to identify emerging issues and provide oversight to the USDA Forest Service, Technology and Development (T&D) Program. The Remote Sensing Applications Center wishes to acknowledge the committee and the San Dimas Technology and Development Center for guidance, direction, and oversight on the project reported in this document.

Abstract

Satellite imagery, such as Landsat Enhanced Thematic Mapper (ETM+), is typically classified through unsupervised and supervised classification techniques. While these techniques are often very accurate, they can be time consuming and expensive to perform especially over large areas, where other methods are more efficient. Using regression-tree analysis with a variety of imagery to classify large areas produces highly accurate results in a relatively short time and is also inexpensive (Lawrence and Wright, 2001). Regression-tree analysis involves classifying high-resolution imagery by standard techniques (i.e., unsupervised/supervised classification), combining the classified high-resolution imagery with lower resolution images and ancillary data, using regression-tree software to build prediction models, and deriving a classified layer from the prediction models. This project used regression-tree analysis to derive percent canopy-cover and impervious surface-cover layers for U.S. Geological Survey (USGS) Zone 41 (18 million hectares). The project also developed automated methods to derive percent cover layers more efficiently.

Table of Contents

Introduction

The Multiresolution Land Characterization (MRLC) consortium is developing a second-generation national land cover database (NLCD) (Homer *et al.,* 2002). In addition to the land-cover classification data layer, layers will be created for percent canopy cover and impervious surface cover for all 50 states. The US Geological Survey, EROS Data Center (EDC) is directing the formulation of these data layers. The USDA Forest Service (USFS) is interested in developing similar data layers for urban areas. These layers can be used to inventory and monitor the structure and health of vegetation, analyze sprawl, assess changes in tree cover, and ascertain the distribution of vegetation on the local and regional environment.

The method EDC has developed for deriving the percent cover layers involve (1) classifying a few relatively expensive high-resolution images; (2) converting the classifications to percentages to create percent cover layers; (3) combining these percent cover layers with Landsat ETM+ imagery and ancillary data, such as tasseled-cap transformations and information on soil, elevation, slope, and aspect; and (4) using regression-tree classifications to derive a cover layer for the entire area with relatively inexpensive, lower resolution imagery such as Landsat ETM+ (Huang *et al.,* 2001).

The objective of this effort was to design and codify procedures that would enable people at various skill levels to derive these percent cover layers in a timely and cost effective manner. The Remote Sensing Applications Center (RSAC) created procedures to do this using models and programs designed in Leica Geosystems ERDAS Imagine software. RSAC tested these procedures by deriving the percent canopy cover and impervious surface cover for Zone 41 in central and northern Minnesota (figure 1). This paper describes the analysis of Zone 41.

Study Area

Zone 41 encompasses most of Minnesota and consists of water (7.1 million hectares), forest (5.7 million hectares), agricultural tracts (4.4 million hectares), urban/ developed areas (0.4 million hectares), and shrub land and grasslands (0.1 million hectares), according to the 1992 NLCD classification. Elevation ranges from 177 to 701 meters. Urban and agricultural areas occur mainly in the south and west. Forest land occupies primarily the north and east. Numerous water bodies are scattered throughout the area.
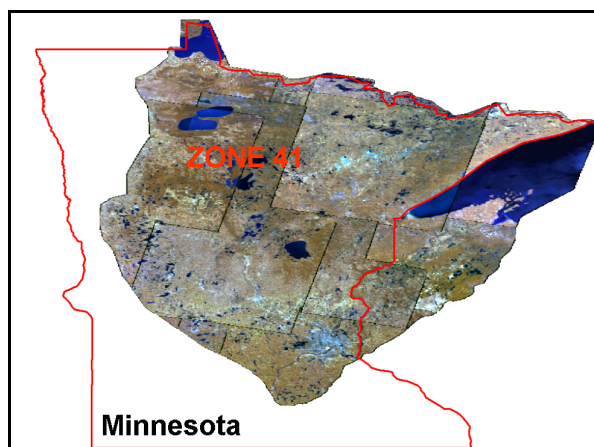

Figure 1—Zone 41 and Minnesota.

Methods and Results

Classification of High-Resolution Images
Training data were derived from 28 scanned natural-color aerial photographs obtained from the Natural Resources Conservation Service (NRCS). Eleven images were of urban/ developed areas. The remaining 17 depicted forest. Fourteen images were located in the southern section of Zone 41 while eight images were located in the western part. The remaining six images were scattered around the central and eastern areas. Each image covered approximately 289 hectares and had a spatial resolution of approximately 0.3 meters. Only the effective area of each photograph was used in this analysis due to geometric distortions at the edges. Occasionally, other areas in the images were excluded due to poor scan quality.

The 28 digital aerial photographs were classified using unsupervised classification techniques into three categories: impervious surface cover, canopy cover, and background. High amounts of confusion occurred among classes in the impervious surface-cover classifications. To separate these classes, area-of-interest (AOI) polygons were drawn around impervious surface areas, and the confused classes were recoded. In the canopy-cover classification, high amounts of confusion occurred between trees and shrub lands/grasslands. To refine the classification of canopy cover, texture was combined with the digital aerial photographs. Modeling using texture also helped eliminate the confusion.

Visually, the classifications appear to be 90 to 100 percent accurate for the impervious surface-cover classifications and 70 to 80 percent accurate for canopy-cover. Examples of the classifications appear in figures 2A and 2B. The digital aerial photography cannot be shown due to plot confidentially agreements.

Conversion of Classifications to Percent Cover Images
Classifications derived from the digital aerial photographs were converted to percent cover layers with a spatial resolution of 30 meters, which matched the resolution of the imagery used for the analysis (figures 2C and 2D). An ERDAS Imagine program was developed to perform this procedure (figure 3A). The program counts the number of cover pixels within the lower resolution size specified by the user and converts this number to a percentage. The image is rescaled to 30 meters in a separate program, which changes the cell size of the imagery (figure 3B).

Combining Lower Resolution Images with Percent Cover Layers
The next step in obtaining continuous cover layers for Zone 41 involved combining the percent cover layers with available lower resolution imagery, including Landsat ETM+, state soil geographic (STATSGO) soil-data imagery, tasseled-cap transformations, and elevation, slope, and aspect information. All of these images were obtained from the EDC and were 30 meter spatial resolution. The STATSGO imagery consisted of soil-carbon, soil-quality, and soil-available-water-capacity data layers. The Landsat ETM+ data comprised three composite images from spring (March–May), leaf-on (July–August), and leaf-off (October– November) time periods (figure 4). The specific dates and paths/rows of the imagery are shown in table 1.

Figure 2A—Impervious surface-cover classification used for training data to analyze Zone 41.

Figure 2B—Tree canopy-cover classification used for training data to analyze Zone 41.

Convert to 30-meters spatial resolution

Convert to 30-meters spatial resolution

Figure 2C—Percent impervious surface cover at 30-meter spatial resolution developed from a classification.
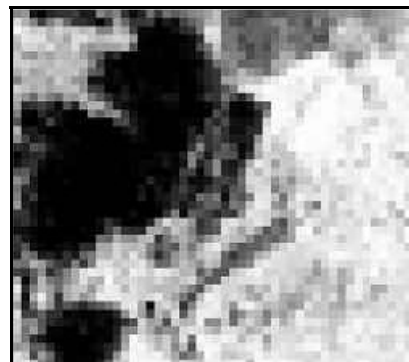
Figure 2D—Percent canopy cover at 30-meter spatial resolution developed from a classification.
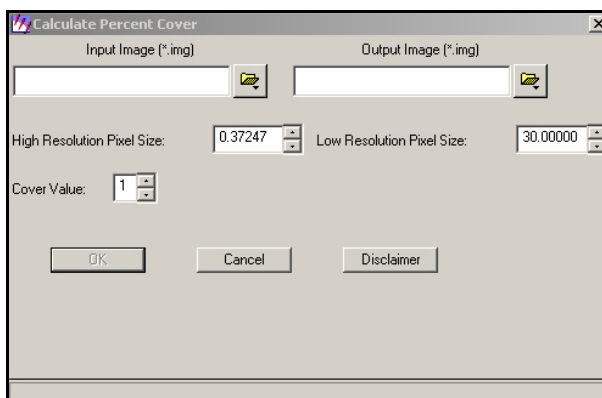
Figure 3A—ERDAS Imagine graphical-user interface which converts a high-resolution classified image into a low-resolution percent cover image.
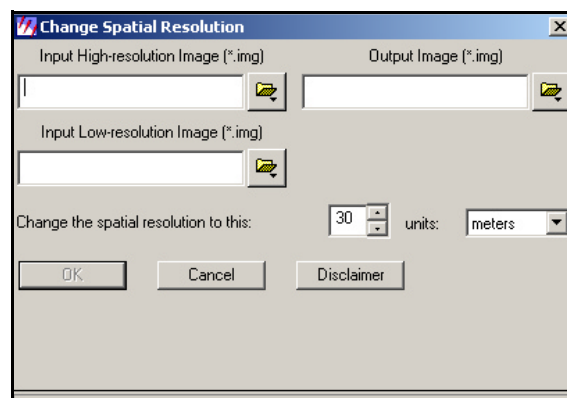
Figure 3B—ERDAS Imagine graphical-user interface which changes the spatial resolution of an image.

3
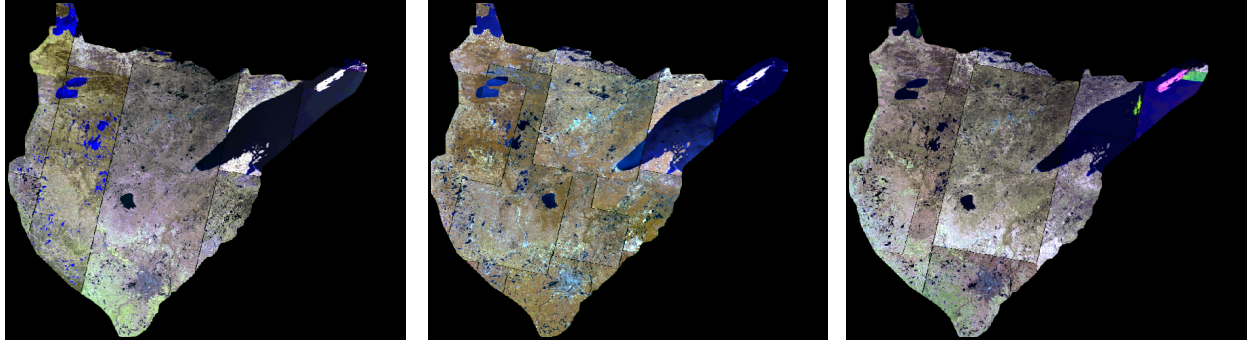
Figure 4—Landsat EMT+ composites used to analyze percent impervious cover and canopy cover for Zone 41.

Table 1—Dates and Paths/Rows for the Landsat ETM+ Imagery Used to Analyze Zone 41

Index	Path/Row	Date	Index	Path/Row	Date	Index	Path/Row	Date
1	25/26	5/4/2001	1	25/26	7/4/2000	1	25/26	10/24/2000
2	25/27	4/23/2000	1	25/27	7/4/2000	2	25/27	10/11/2001
3	26/27	4/28/1999	2	26/27	7/3/2000	3	26/27	11/14/1999
4	26/28	4/25/2001	3	26/28	6/28/2001	3	26/28	11/14/1999
4	26/29	4/25/2001	4	26/29	6/9/2000	3	26/29	11/14/1999
5	27/26	4/29/2000	5	27/26	7/18/2000	4	27/26	10/22/2000
5	27/27	4/29/2000	6	27/27	7/5/2001	5	27/27	11/5/1999
5	27/28	4/29/2000	7	27/28	7/10/2000	5	27/28	11/5/1999
5	27/29	4/29/2000	8	27/29	7/24/1999	6	27/29	11/10/2001
6	28/26	4/20/2000	9	28/26	6/26/2001	7	28/26	11/17/2001
7	28/27	3/3/2000	10	28/27	7/23/1999	8	28/27	10/27/1999
7	28/28	3/3/2000	10	28/28	7/23/1999	8	28/28	10/27/1999
7	28/29	3/3/2000	11	28/29	8/10/2000	8	28/29	10/27/1999
8	29/26	3/10/2000	12	29/26	7/16/2000	9	29/26	10/20/2000
9	29/27	3/2/2000	13	29/27	8/1/2000	9	29/27	10/20/2000
10	29/26	5/8/2001	13	29/28	8/1/2000	9	29/28	10/20/2000
Spring			**Leaf-on**			**Leaf-off**		

The Landsat ETM+ scenes used for the composite images were the most cloud-free ones available. However, some clouds remained in the images. RSAC developed an automated process to eliminate clouds by predicting what the digital-number (DN) values would be if there were no clouds. This process is described in more detail in the Appendix. RSAC used this method to "remove" clouds from imagery for Zone 41. Originally, RSAC attempted to perform this procedure on the entire composite image for each time period. The results were poor quality because of radiometric differences between individual dates within the composites. To improve the quality, each composite image was split into individual dates and the eliminate-cloud procedure was performed on each one. Seven of the individual date images were completely cloud free, and 24 of them had clouds. An example of a result appears in figure 5.
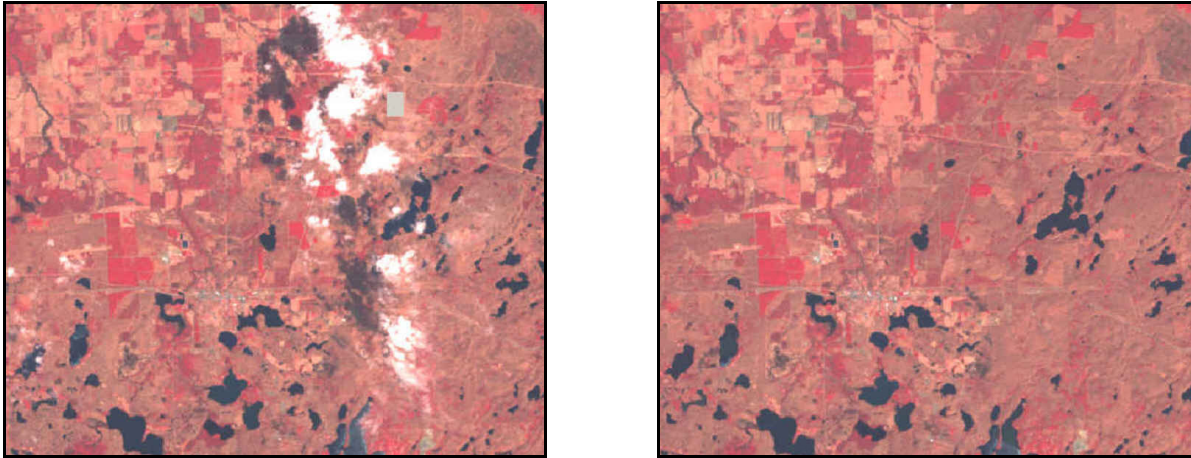
Figure 5—On the left is the original Landsat ETM+ image with clouds. In the image on the right the clouds have been removed by using the eliminate-clouds procedure developed at RSAC. Bands 5, 6, and 3 (RGB) are displayed. These images are displayed unstretched, with brightness and contrast enhancement. The difference between the two images illustrates the cloud removal process.

To test the accuracy of this method, the DN values of a small subset of completely cloud-free imagery were predicted (figure 6). The original values were subtracted from the predicted ones. The results are shown in table 2. The maximum difference between the original and predicted DN values was 77 (band 5). The mode and median differences for all bands except 5 were zero. Band 5 had a mode of -1 and a median of 1. The means of the differences ranged from a low of 0.037 (band 3) to a high of 0.758 (band 5). Based on the standard deviations, 68 percent of the predicted values were within $\pm$ 7 of the original DN values, and 95 percent of the predicted values were within $\pm$ 13 of the original DN values. This test concluded that the predicted values would not significantly affect further analyses and the slight errors in the predictions were more acceptable than holes in the final results due to cloud and cloud-shadow cover.
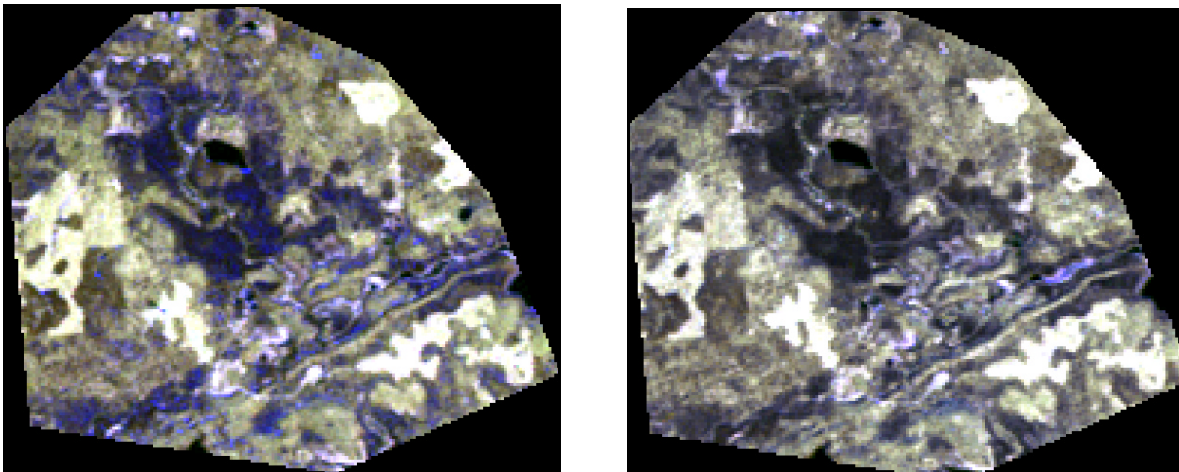
Figure 6—The RSAC eliminate-clouds procedure was tested on a small subset of Landsat ETM+ imagery. The image on the left was used for the test. The image on the right is the predicted image. Bands 5, 6, and 3 (RGB) are displayed. These images are displayed stretched without brightness or contrast enhancement. The similarity of the two images illustrates the ability of the model to accurately predict the DN values of the pixel layers.

Table 2—Summary of the Difference Results Between Original and Predicted Digital Number Values

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Maximum Difference	21	25	25	31	60	32
Minimum Difference	-14	-23	-23	-53	-77	-55
Mode	0	0	0	0	-1	0
Median	0	0	0	0	1	0
Mean Difference	0.176	0.038	0.037	-0.118	0.758	0.435
Standard Deviation	1.915	2.787	2.778	4.678	6.658	4.820

After the clouds had been removed from the Landsat ETM+ data, all the lower resolution and ancillary data layers were stacked into one image and subsetted so that their geographical extent was slightly smaller than that of the high-resolution imagery. Next, the percent cover layers were subsetted so that the pixels on the borders barely overlapped the pixel boundaries of the subsetted lower resolution imagery. The subsetted training percent cover layers were then stacked into one image with this lower resolution imagery. Every layer in the stacked image was carefully checked for georeferencing accuracy.

Regression-Tree Classification
These images were then used as training data to build the prediction models, which were created by Cubist (www.rulequest.com). Cubist does not read remote sensing imagery directly so RSAC developed a program that converts imagery to a format that Cubist can decipher (see the Appendix for a detailed description).

Cubist can be run three different ways: "rules alone," "instances and rules," and "let Cubist decide." Models were created using all three of these options. The "instances and rules" and "let Cubist decide" options produced models with higher correlation coefficients and lower average errors than the "rules alone" option. However, due to the complexity of the models generated, it would have taken more than 250 days to apply them to the entire dataset. Therefore, models using the "rules alone" option were applied to the entire dataset, which took 18 only hours. Several Cubist trials were run with different variables to derive the best possible models. The results of these trials appear in tables 3 and 4. The factors that determined which model to use to build the final percent cover layers were low average error, low relative error, and high correlation.

Originally, model 8 was chosen to chart percent canopy cover (table 3); it used all the Landsat ETM+ bands and the three STATSGO bands. However, when the model was applied, the canopy-cover image had visible unnatural soil boundaries and seam lines (figure 7). To improve the canopy-cover image, model 14 was chosen (table 3); it eliminated the soil data but used all the Landsat ETM+ bands and the elevation information. The soil data were omitted because of their strong influence on the final models. Additionally, date images (thematic layers representing the dates of the individual dates in the composite Landsat ETM+ images) provided by the EDC were used to reduce the seam lines.

For percent impervious surface cover, model 10 was originally chosen (table 4); it used all the Landsat ETM+ bands, all the STATSGO bands, and all the elevation, slope, and aspect information. However, like the original canopy-cover image, the impervious surface-cover image had visible unnatural soil boundaries, although it did not have problems with seam lines. To improve this impervious surface-cover image, model 11 was selected (table 4); it eliminated the soil data but used all the Landsat ETM+ bands, and the elevation, slope, and aspect information.

To apply the Cubist models to the Zone 41 datasets, which consisted of the data layers chosen in the Cubist models, RSAC developed a program that interfaces with the CubistSam software, which is recommended by RuleQuest for applying Cubist models (see the Appendix). The final canopy-cover image is shown in figure 8. The final impervious surface-cover image appears in figure 9.

Accuracy Assessment
To assess the accuracy of the models, RSAC did a tenfold cross validation using Cubist. This technique divides the data into 10 equal groups. With each iteration, nine groups build models, and the remaining group tests them. At the end of each iteration, error rates and correlation coefficients are calculated. At the end of the tenfold cross validation, the error rates and correlation coefficients are averaged. These averages are nearly unbiased predictors of the true error rates and correlation coefficients of the model built with all the data. The cross validation for the percent canopy-cover model had an average error of 14.5, a relative error of 0.55, and a correlation coefficient of 0.77, while the one for the percent impervious surface-cover analysis had an average error of 4.6, a relative error of 0.36, and a correlation coefficient of 0.82.

Conclusion
Compared to other classification techniques that could have been used to derive these percent cover layers, these accuracy levels are considered high. Additionally, these methods are much faster and less expensive than traditional approaches (Lawrence and Wright, 2001). These procedures provide geospatial data users an improved method to derive percent cover layers.

The project objective was to develop procedures to construct percent cover layers. Analyzing Zone 41, accomplished this objective. RSAC formulated a nine-step method, which is described in Appendix A, to derive percent cover layers. This method involves easy-to-use ERDAS Imagine programs that simplify processes and automate procedures. Using this approach, users can derive highly accurate percent cover layers in a timely, cost effective manner.
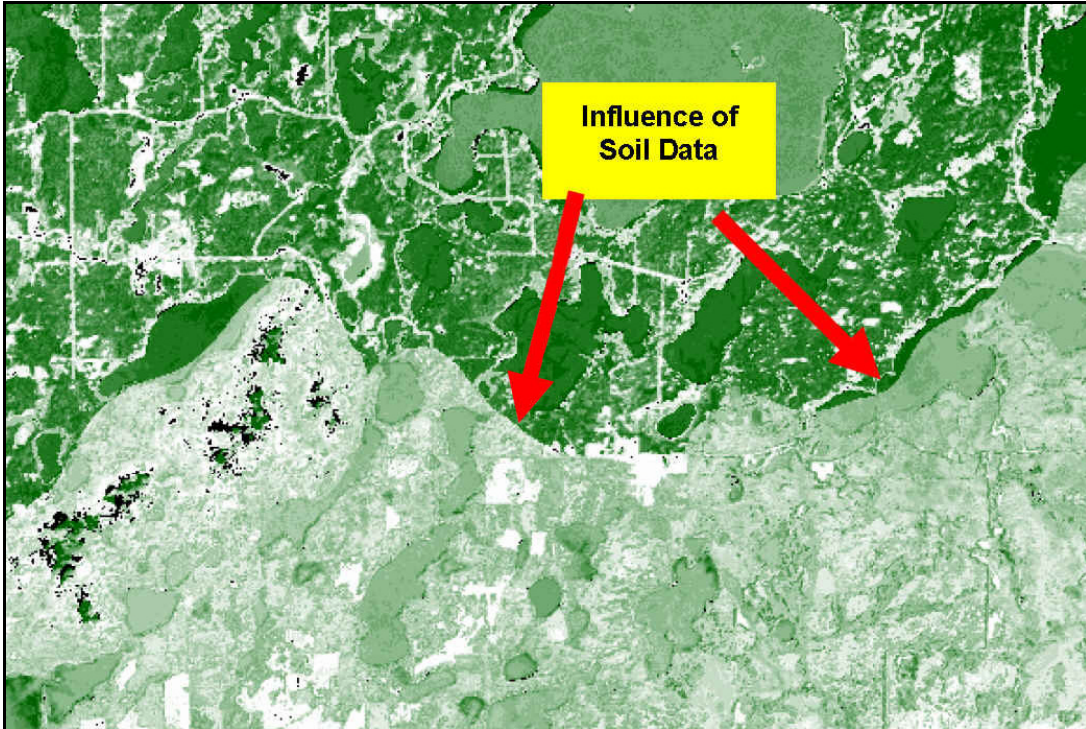
Table 3—Cubist Trials for Analysis of Canopy Cover Zone 41

#	Spring 1	Spring 2	Spring 3	Spring 4	Spring 5	Spring 7	Leaf-on 1	Leaf-on 2	Leaf-on 3	Leaf-on 4	Leaf-on 5	Leaf-on 7	Leaf-off 1	Leaf-off 2	Leaf-off 3	Leaf-off 4	Leaf-off 5	Leaf-off 7	Spring TC 1	Spring TC 2	Spring TC 3	Leaf-on TC 1	Leaf-on TC 2	Leaf-on TC 3	Leaf-off TC 1	Leaf-off TC 2	Leaf-off TC 3	Soil AWC	Soil C	Soil Qual	DEM	Slope	Aspect	Spring Date	Leaf-on Date	Leaf-off Date	average	relative	correlation
1	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×				12.10	0.47	0.82
2	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×																			14.30	0.55	0.76
3		×		×	×	×	×	×	×	×	×	×										×	×	×													13.30	0.51	0.79
4	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×										×	×	×	×	×	×				12.10	0.47	0.82
5			×	×	×	×				×	×	×				×	×	×		×			×			×		×	×	×	×	×	×				12.20	0.47	0.82
6		×	×	×	×	×	×	×	×	×	×	×			×	×	×	×										×	×	×	×	×	×				12.10	0.47	0.82
7	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×										×	×	×							12.70	0.49	0.81
8	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×										×	×	×	×						12.20	0.47	0.82
9	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×										×	×	×	×		×				12.10	0.47	0.82
10	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×										×	×	×	×	×					12.10	0.47	0.82
11	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×													×	×	×				12.80	0.50	0.80
12	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×						12.30	0.48	0.81
13	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×										14.30	0.55	0.76
14	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×													×			×	×	×	14.10	0.54	0.78

Note: The highlighted row is the model ultimately used

Table 4—Cubist Trials for Analysis of Impervious Surface Cover in Zone 41

| | Trials | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Training | | |
	Spring 1	Spring 2	Spring 3	Spring 4	Spring 5	Spring 7	Leaf-on 1	Leaf-on 2	Leaf-on 3	Leaf-on 4	Leaf-on 5	Leaf-on 7	Leaf-off 1	Leaf-off 2	Leaf-off 3	Leaf-off 4	Leaf-off 5	Leaf-off 7	Spring TC 1	Spring TC 2	Spring TC 3	Leaf-on TC 1	Leaf-on TC 2	Leaf-on TC 3	Leaf-off TC 1	Leaf-off TC 2	Leaf-off TC 3	Soil AWC	Soil C	Soil Qual	DEM	Slope	Aspect	average	relative	correlation
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	3.80	0.31	0.88
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X										X	X	X	X	X		3.70	0.30	0.89
3	X	X		X	X	X	X			X	X	X				X	X	X	X			X			X			X	X	X		X		4.00	0.33	0.87
4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																4.30	0.35	0.85
5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X							4.30	0.35	0.85
6	X	X	X	X	X	X							X	X	X	X	X	X										X	X	X		X		4.50	0.36	0.84
7	X	X	X	X	X	X							X	X	X	X	X	X										X	X	X		X		4.30	0.35	0.86
8	X	X	X	X	X	X													X	X	X							X	X	X	X	X		4.70	0.38	0.83
9	X	X	X	X	X	X																						X	X	X	X	X		4.90	0.40	0.80
10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X										X	X	X	X	X	X	3.70	0.30	0.89
11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X													X	X	X	4.60	0.36	0.83

Figure 7—Examples of bad results obtained for the canopy-cover layer of Zone 41. These artifacts of modeling were eliminated using the highlighted model shown in table 3.
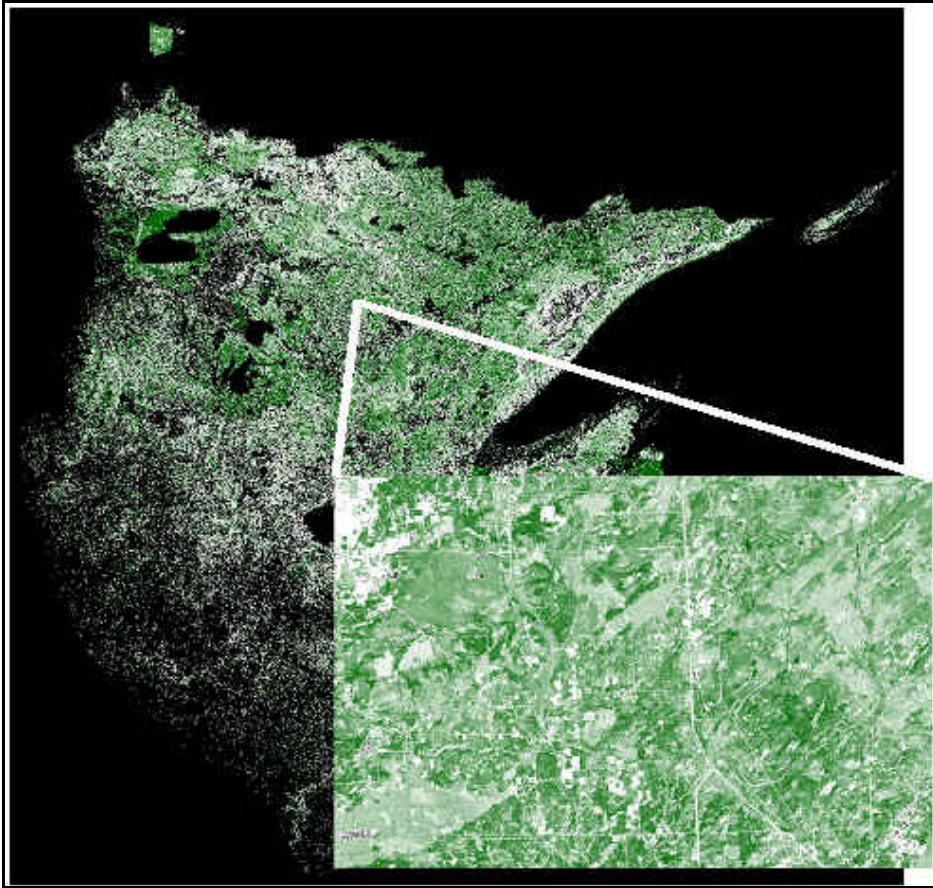
Figure 8. Final canopy-cover layer for Zone 41. Colors vary from white to dark green, indicating an increase in percent cover. Black areas have been masked or have no tree cover.
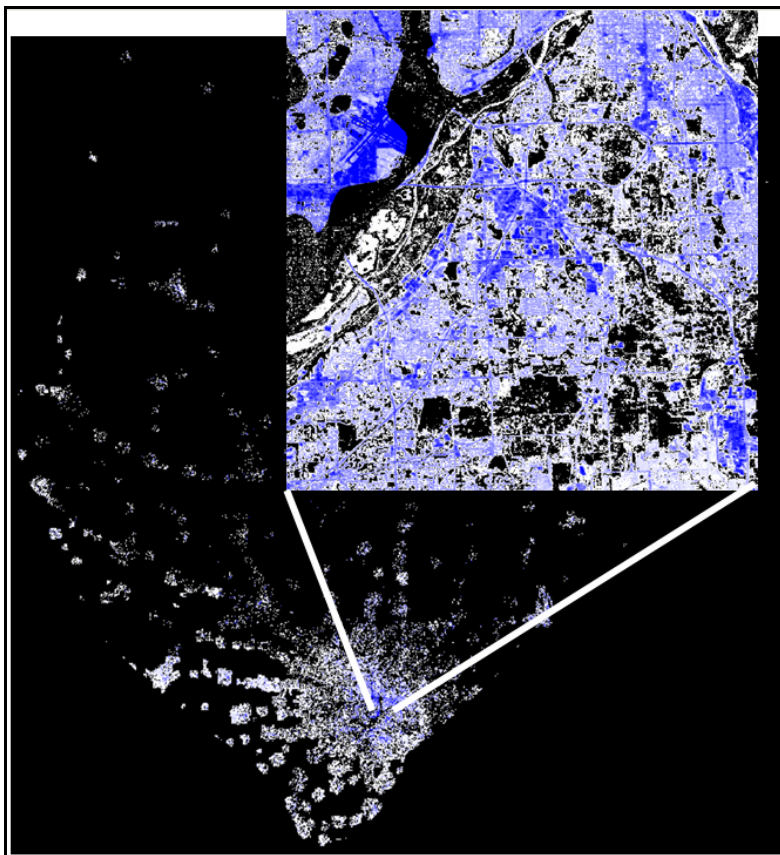

Figure 9. Final impervious surface-cover layer for Zone 41. Colors vary from white to dark blue, indicating an increase in percent cover. Non-urban areas have been masked in black.

11

References

Homer, C.G., C. Huang, L. Yang and B. Wylie. 2002. "Development of a Circa 2000 Land-Cover Database for the United States," in Proceedings of the American Society of Photogrammetry and Remote Sensing Annual Conference, April, 2002, Washington DC.

Huang, C., L. Yang, B. Wylie and C. Homer. 2001. "A Strategy for Estimating Tree Canopy Density Using Landsat & ETM+ and High-resolution Images over Large Areas," in Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, 5-7 November 2001, Denver, CO, CD-Rom.

Lawrence, R.L., and A. Wright. 2001. "Rule-Based Classification Systems Using Classification and Regression-Tree (CART) Analysis," Photogrammetric Engineering and Remote Sensing, 67:1137–1142.

Appendix A: Description of Programs

The programs outlined in this appendix were developed by RSAC to analyze percent canopy cover and impervious surface cover but are also useful for determining other percent cover layers. The nine steps are as follows:

1. Eliminate clouds from lower resolution images
2. Select high-resolution images
3. Classify high-resolution images
4. Subset lower resolution images
5. Convert classifications to percentages
6. Select training/testing blocks
7. Prepare data for Cubist
8. Run Cubist
9. Apply Cubist results spatially

The materials needed to derive the percent cover layers include the following:

- High-resolution images
- Lower resolution images
- Ancillary data
- Cubist software (www.rulequest.com)
- ERDAS Imagine (www.erdas.com)
- ERDAS Imagine processing tools developed by RSAC. The software can be downloaded from the RSAC web site (http://fsweb.rsac.fs.fed.us)

Step 1: Eliminate Clouds from the Lower Resolution Images

There are three different ways to perform the eliminate-clouds procedure. The method chosen depends upon the data and available software. The one-step and the three-step procedure allows a user only three images along with their associated cloud and cloud-shadow masks. The four-step procedure allows an unlimited number of images. The three-step procedure is recommended only if the user does not have a local Cubist license.

One-Step Eliminate-Clouds Procedure
The ERDAS Imagine graphical-user interface (GUI) developed by RSAC for the one-step eliminate-clouds procedure is shown in figure A1. The input image in the upper left is the clouded image. The two input images below this clouded image are to be used to build the prediction cloud-free1 and cloud-free2. If the cloud-free1 and cloud-free2 images have clouds and cloud shadows, these areas must not overlap the shadowed areas of the clouded image. If they do, not all of the clouds and cloud shadows can be removed from the clouded image. To resolve this problem, run the program again and select other images for the cloud free1 and cloud free2 inputs.

Each of these input images must have a cloud and cloud shadow mask even if there are no clouds or shadows in the images. These masks must be coded so that a value of one signifies clouds and cloud shadows while a value of zero indicates there are none.

The program may take a long time to complete. Only one of these processes should be running at a time. The eliminate-clouds output consists of only the areas covered by the clouded-image cloud and cloud shadow mask. To construct a full image, combine the eliminate-clouds output with the clouded image by creating a model in ERDAS Imagine's Model Maker. Use the conditional statement to create a function that chooses the eliminate-clouds output if the coded value is greater than one or assigns the clouded image if the coded value is zero.

Three-step Eliminate-clouds Procedure
The ERDAS Imagine menu for the three-step eliminate-clouds procedure is shown in figure A2. The GUIs for Step 1: Prepare Data for Cubist and Step 3: Apply Cubist Results are identical to the GUI for the one-step eliminate-clouds procedure (figure A1). So use the instructions already outlined. **All of the paths and filenames selected in the GUIs for Step 1 and Step 3 must be identical.**

Step 1: Prepare Data for Cubist combines the clouded, cloud free1, and cloud free2 images and converts them to comma-delimited-text ASCII files for each band. When Step 2: Run Cubist is clicked, nothing happens. At this point, a user must find a computer with access to a Cubist license and run Cubist. Once Cubist bas been completed, Step 3: Apply Cubist Results can be performed using the **exact paths and filenames as Step 1**. The output should look the same way it does in the one-step procedure.
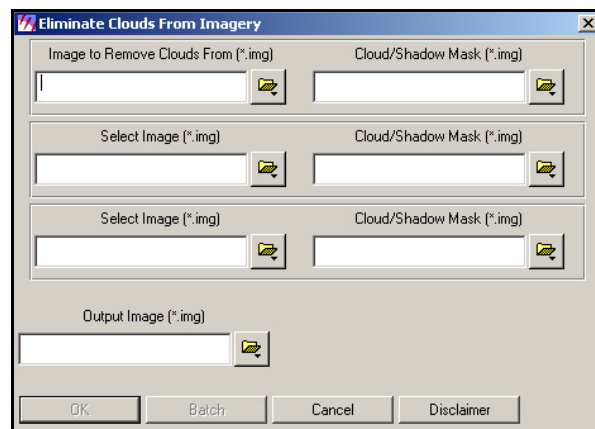

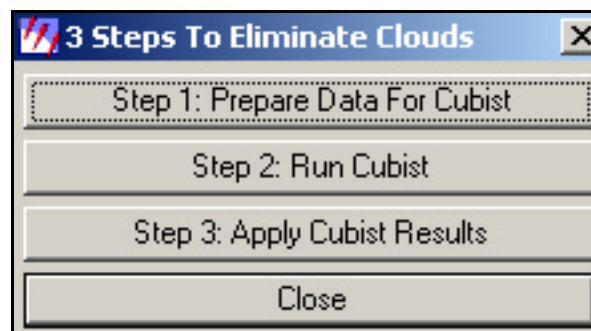Figure A1—ERDAS Imagine GUI which activates the eliminate-clouds procedure.

Figure A2—List of steps for the ERDAS Imagine three-step eliminate-clouds procedure.

14

Four-Step Eliminate-Clouds Procedure

The ERDAS Imagine menu for the four-step eliminate-clouds procedure is shown in figure A3. This suite of programs should be chosen only if the user wants to include more than two images in building the prediction models.

The first step, generate masked and unmasked subsets, involves identifying areas that are cloud free in all images. The ERDAS Imagine GUI that accomplishes this is shown in figure A4. The image to use for the dependent variable is the clouded image. The images to use for the independent variables are cloud-free1 and cloud-free2. If any of the images chosen for the independent variables have areas that should not be used for the training data (eg. clouds, cloud shadows, landscape variations), these areas should be masked. In the masked images, a value of one signifies areas to be eliminated and a zero indicates areas that should be used. The masked images should be input into the edit text field on the right.

Figure A3—List of steps for the ERDAS Imagine four-step eliminate-clouds procedure.

The program outputs two sets of images. One set is named Output_Root_Name_unmasked*. img. These output files become the input files for Step 2: Prepare Data for Cubist. The other set is named Output_Root_Name_masked*.img. These output files become the input files for Step 4: Apply Cubist Results. The number of images in each set of output files equals the number of bands in the dependent-variable input image.

Figure A4—ERDAS Imagine GUI that identifies areas that are cloud free in all images.

15

The next step involves converting all the Output_Root_Name_unmasked*.img output files to comma-delimited-text ASCII files. The ERDAS Imagine GUI for Step 2: Prepare Data for Cubist is shown in figure A5. For every input image, the program creates two output files, a .data file and a .names file. These files become the input files for the next step.

The ERDAS Imagine GUI for Step 3: Run Cubist is shown in figure A6. This GUI was to enable Cubist to run in batch mode. Cubist is perhaps the slowest part of the whole process. Additionally, Cubist uses a lot of the central processing unit and can dramatically slow down the computer. With this GUI, numerous Cubist processes can be run overnight or during the weekend. This GUI runs Cubist using the "rules alone" option, which is the most appropriate way to derive prediction models for the eliminate-clouds procedure. If Cubist is run outside of ERDAS Imagine, be sure to choose the "rules alone" option.

After the Cubist models have been developed, the next step applies them to the clouded portions of the dependent-variable image selected in Step 1. The ERDAS Imagine GUI used for Step 4: Apply Cubist Results is shown in figure A7. The input images are the Output_Root_Name_masked*.img files from Step 1. This program must be run for each individual Output_Root_Name_masked*.img and uses a modified version of the CubistSam (www.rulequest.com) software, which is recommended by RuleQuest for applying Cubist models. The number of output files should equal the number of bands in the dependent-variable image. These output files must be stacked using the Interpreter -> Utilities -> Layer Stack function in ERDAS Imagine.
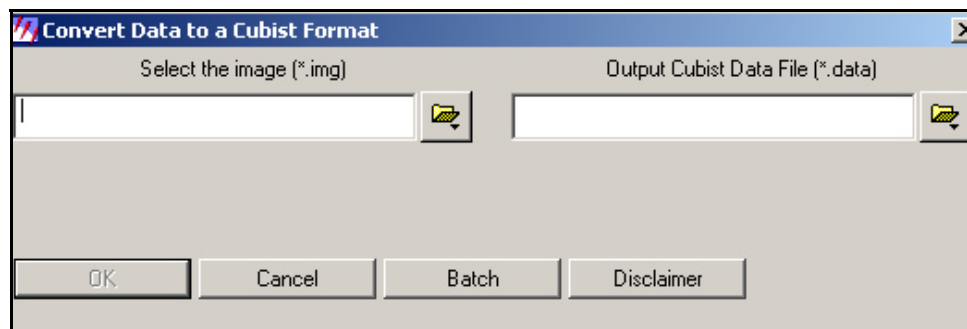
Figure A5—ERDAS Imagine GUI that convert an image into a comma-delimited text ASCII format that can be read by Cubist.
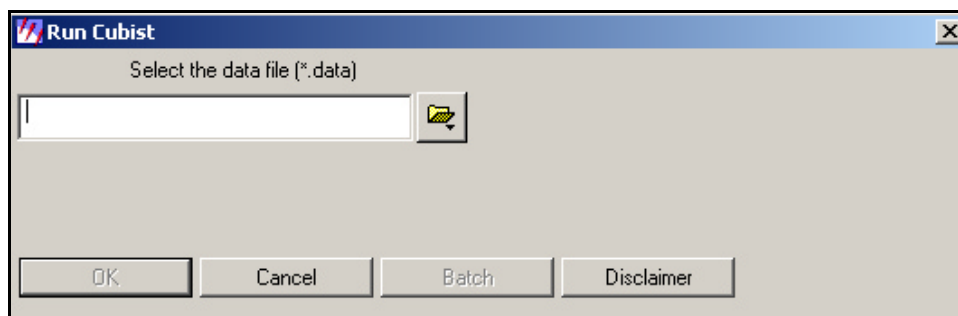
Figure A6—ERDAS Imagine GUI, for Run Cubist (www.rulequest.com). RSAC developed this GUI to run Cubist in batch mode.
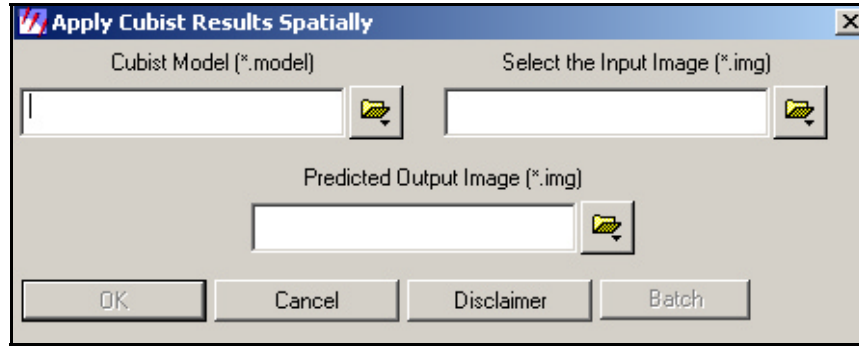
Figure A7—ERDAS Imagine GUI which applies Cubist models spatially.

Once the bands have been stacked, this stacked image can be combined with the dependent-variable image to create a cloud-free one. To combine the images, use a conditional statement in ERDAS Imagine's Model Maker program. The conditional statement should choose the predicted layer-stack image if assigned value is greater than zero or the dependent-variable image if the predicted layer stack image is zero.

Step 2: Select High-Resolution Images
RSAC developed ERDAS Imagine programs to aid in selecting high-resolution images (figure A8). The images recommended by these programs are viewed as suggestions. The user can make other choices.

Create a Water Mask
This program creates a water mask if the user has no access to one (figure A9). The water mask is used in the Select Block program . A water mask is recommended because this program selects blocks of imagery with high variance. Since boundaries between water and land always have high variance because of the large spectral difference between the two, the Select Block program always chooses areas with water. This may be undesirable.
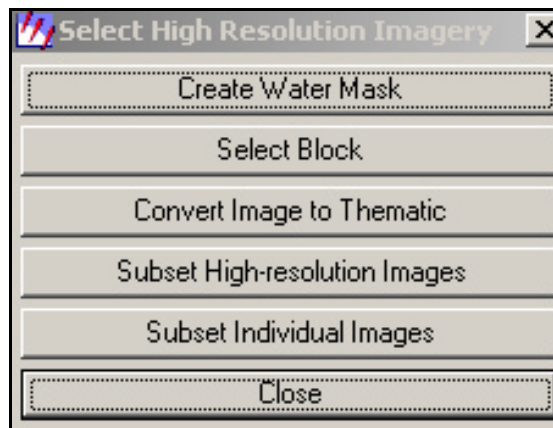
Figure A8—List of programs to aid in selecting high-resolution imagery.
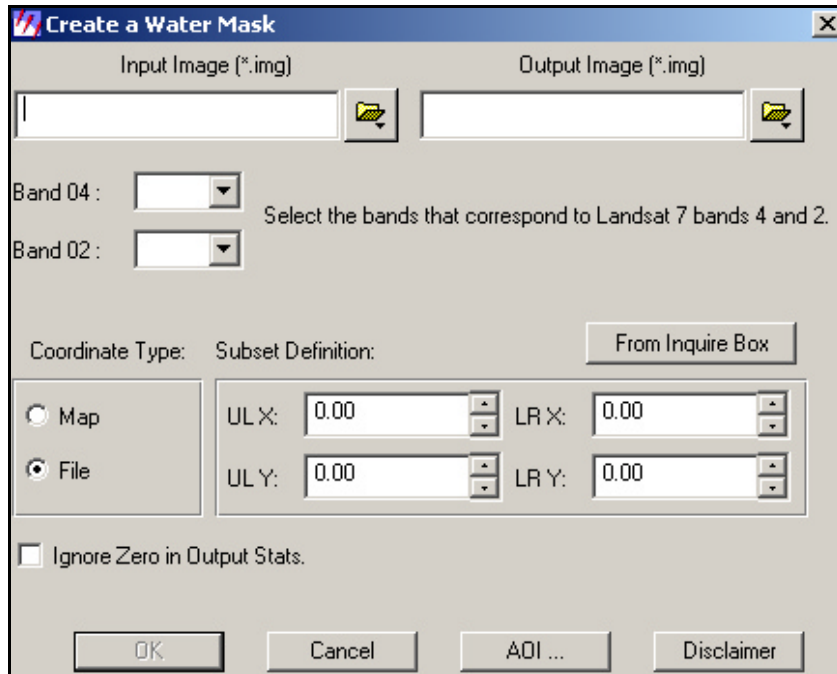
Figure A9—The ERDAS Imagine Create Water Mask program
assigns a value of one if Band 04 is greater than Band 02; otherwise,
the value is zero.

To use the Create Water Mask program, select an input and an output image. If the input image is Landsat ETM+ imagery, set Band 04 to four and Band 02 to two. The output of the program will consist of ones and zeros. A value of one signifies land, and zero indicates water. The output is calculated by assigning a one if band 4 is greater than band 2, otherwise, assign a zero. An example of a water mask appears in figure A10.
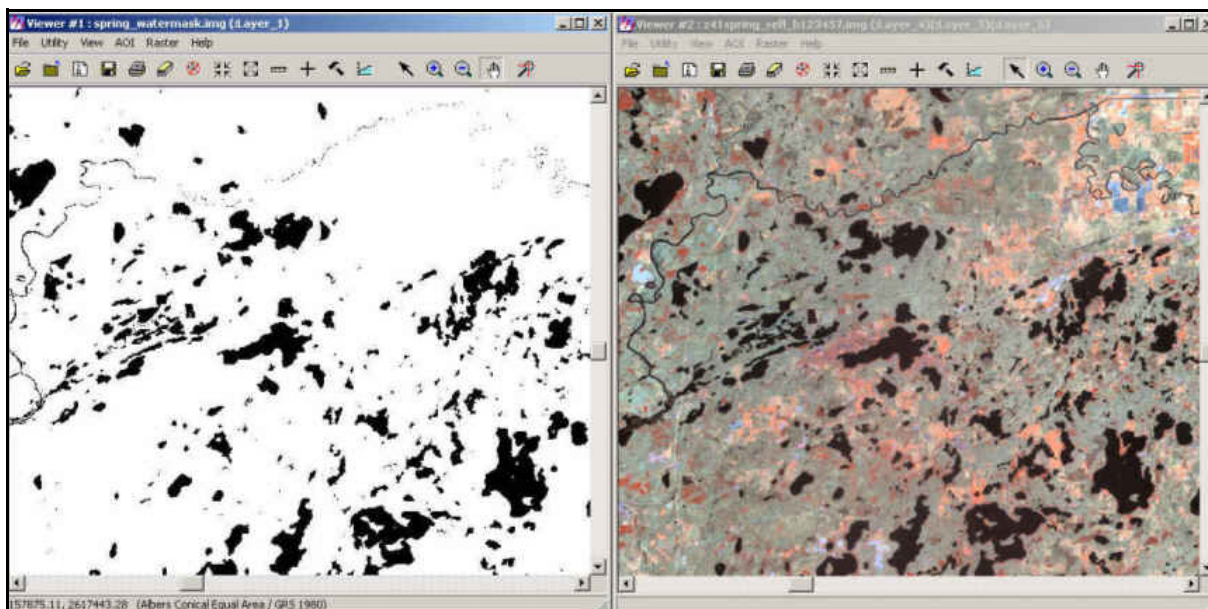
Figure A10—The image on the left is a water mask created with the Create Water Mask program. The input image is shown on the right. This image is Landsat ETM+. Shown are bands 4, 5, and 7.
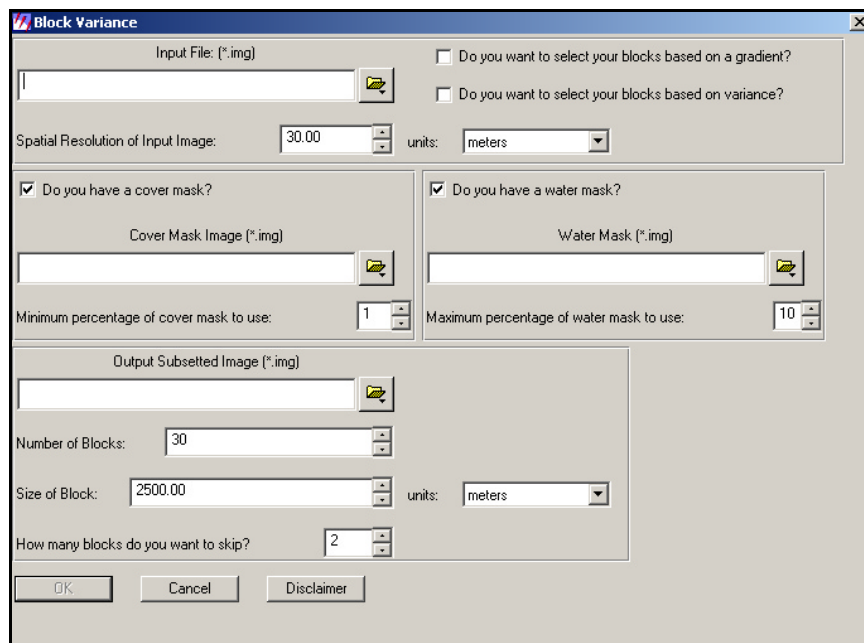
Figure A11—The ERDAS Imagine Select Block program, chooses
blocks based upon variance or a gradient of values.

Select Block

The objective of the Select Block program (figure A11) is to choose small blocks from a large image (see figure A12A). These small blocks will be further analyzed, and will eventually produce a classified cover image of the large image. Therefore, it is important that these small blocks adequately represent the large image.

The Select Block program chooses blocks based either upon variance or a gradient. The variance option chooses blocks of imagery with the highest variance. If areas of high variance are chosen, probably most of the information in the image is being captured. Blocks can also be selected based upon a gradient of values, which may better represent the data. If the gradient option is chosen, the input image must contain gradients (e.g., tasseled-cap, NDVI). Better results may be obtained if extreme values are eliminated and the gradient image is stretched. RSAC has developed a model called "trim_tails_stretch.gmd" to accomplish this.

A water mask can be selected from the Create Water Mask program. This water mask should have a value of zero for water areas and a value of one for land. If a water mask is chosen, the program will only analyze a block if its percentage of water pixels is less than or equal to the Maximum Percentage of Water Mask to Use.

If certain areas (e.g., urban regions) or cover types (e.g., forest) are the focus of the analysis, cover-mask these areas. In the cover-mask, zeros represent those areas to be masked; any other number can be used for non-masked areas. The block will be analyzed only if the percentage of non-masked cover pixels is greater than the Minimum Percentage of Cover Mask to Use.
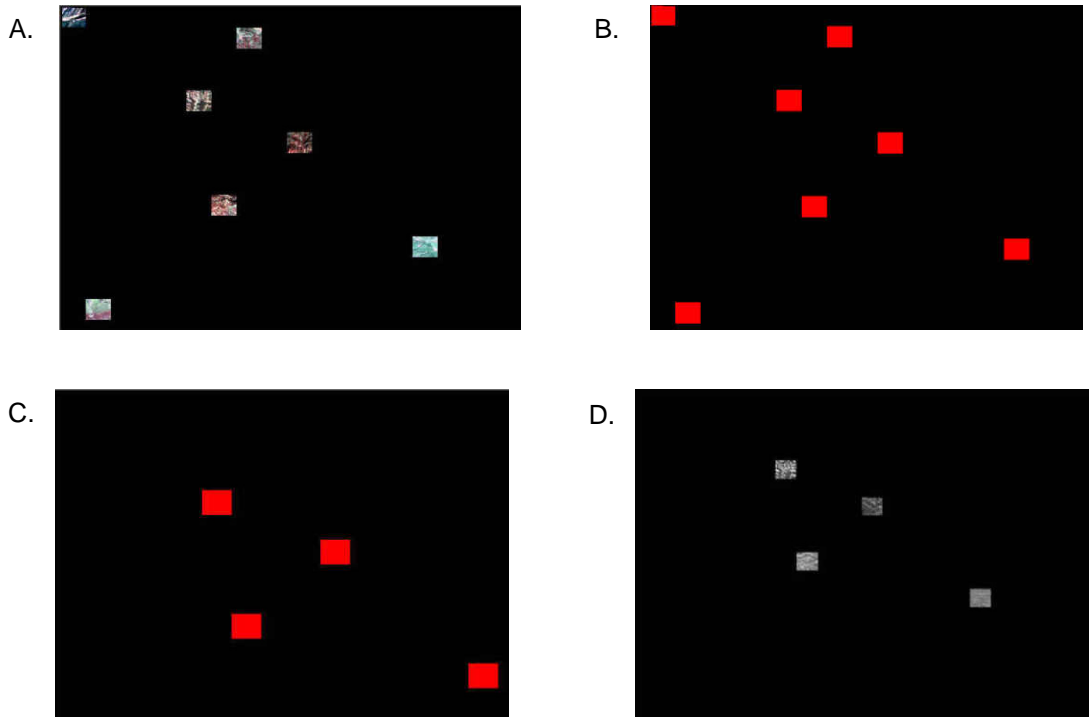
19

A.   B.

C.   D.

Figure A12—Examples of block files: *A.* shows the output, from the Select Block program; *B.* shows the blocks after they have been converted to a thematic image using the Convert Image to Thematic program; *C.* shows only those blocks to use. The other blocks are recoded by drawing area-of-interest polygons around them and changing the values to zero; *D.* Shows the blocks of high-resolution images created with the Subset High-Resolution Images program.

Blocks are likely to be next to others with similar variance or gradient values. To avoid clumps of blocks in the output, set a number of blocks to skip.

The Select Block program is designed to guide the selection process. If additional blocks are needed, open the output from the Select Block program in an ERDAS Imagine viewer. Go to Utility -> Inquire Box. Move the Inquire Box so that it completely encompasses a single block. Make sure no background data is included. Move the Inquire Box to an area you wish to include. Be careful when moving the Inquire Box not to change its size.

On the main ERDAS Imagine toolbar, click on Data Prep. Select Subset Image. Choose the input image for the Select Block program as the one to subset. Click on the From Inquire Box button. After this procedure has been completed, open the add_block.gmd model, which was created by RSAC, in ERDAS Imagine's Model Maker. Select the inputs and outputs and run this model.

Convert Image to Thematic
The Convert Image to Thematic program converts the output from the Select Block program to a binary thematic image (figure A13) . The blocks are recoded to a value of one, and everything else receives a value of zero (see figures A12B and A12C).

If blocks are not to be used in the analysis, recode these by drawing area-of-interest polygons (AOIs) around them. Go to Raster -> Recode. Change the value of these blocks from one to zero.
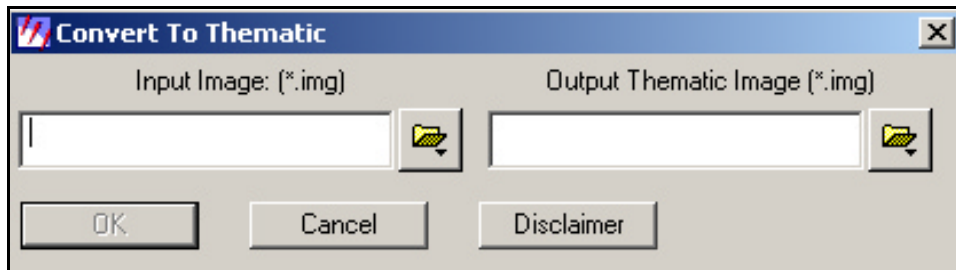

Figure A13—The ERDAS Imagine Convert Image to Thematic program, changes the input image to a binary one where a value of one indicates areas of nonbackground data and zero represents background data.

Subset High-Resolution Images
Once the blocks have been selected, the high-resolution images need to be subsetted. To run the Subset High-Resolution Images program (figure A14), select all the high-resolution images for the input files by clicking on the Add button for each one. If you make a mistake, click on the Clear button. Select Block Image, which is the output from the Convert Image to Thematic program. Select an output image and click OK. Depending upon the input imagery, this program may take at least two to three hours to run. An example of the output appears in figure A12D.

Subset Individual Images
Next, each individual block in the output from the Subset High-Resolution Images program must be subsetted. Load these output images into a viewer. Go to the Utility -> Inquire Box. Position the box around a block of imagery. Make sure to include no background data. Click on Subset Individual Images or go to Data Prep -> Subset Image (figure A15). Select the output from the Subset High-Resolution Images program as the input file. Choose an output file. Click on the From Inquire Box button and press OK. Repeat these steps to subset each individual block.
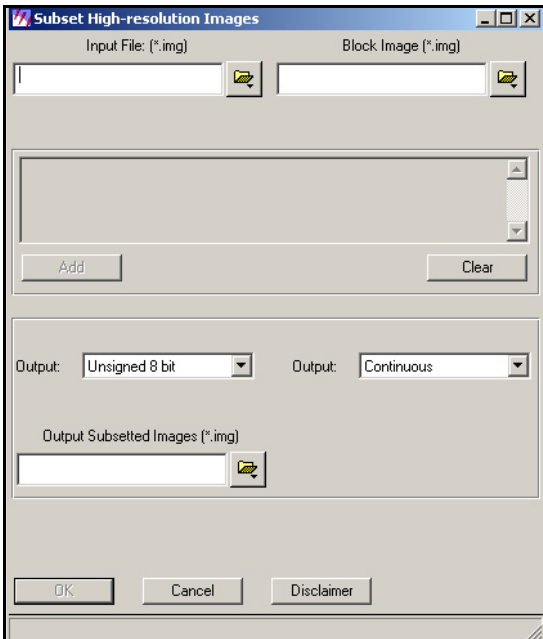
Figure A14—The ERDAS Imagine Subset High-Resolution Images program, subsets the input images to fit the blocks in the Block Image.
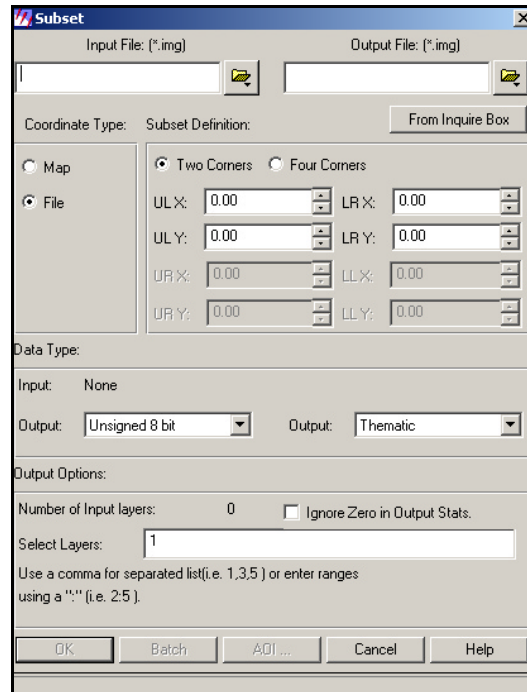

Figure A15—The ERDAS Imagine Subset program.

Step 3: Classify High-resolution Images

A variety of techniques can be used to classify high-resolution imagery. RSAC used ISODATA within ERDAS Imagine to analyze Zone 41, but other classification software exists.

The high-resolution imagery should only be categorized into classes of interest. For example, for the analysis of Zone 41, since the classes of interest were impervious surface cover and canopy cover, the high-resolution images were put into only these two classes.

Step 4: Subset Lower Resolution Images

Step four involves subsetting the high-resolution and lower resolution imagery so that they fit together. If blocks were chosen with the Select High-Resolution Imagery programs already described, use the subset Lower Resolution Image program (figure A17) to prepare the lower resolution images. All the lower resolution imagery and ancillary data should be layer-stacked into one image before running this program, which will subset a lower resolution image to match the boundaries of a high-resolution one. The inputs to this program are self-explanatory. Run the program for all the blocks of high-resolution images.

If the high-resolution imagery was selected by some other method, the Subset Lower Resolution Image program will not work, and another technique must be used. This occurred with the percent cover analysis of Zone 41, so the following method was used to subset the high-resolution and lower resolution images. The lower resolution and high-resolution image

were loaded into separate ERDAS Imagine viewers which were geolinked. In the viewer with the high-resolution image, an Inquire Box was positioned so that a pixel of the high-resolution image barely overlapped a pixel of the lower resolution one (see figure A16). Once the area to be subsetted has been captured by the Inquire Box, the ERDAS Imagine Data Prep -> Subset Image program subestted the high-resolution image. This procedure was repeated to classify all the high-resolution images.

Next, the lower resolution image and all the ancillary data layers were subsetted. The same technique was repeated to subset this imagery except the first pixel of lower resolution imagery that lay entirely within the high-resolution image was chosen as the boundary for the Inquire Box (see figure A16). The lower resolution imagery and ancillary data were then subsetted for every high-resolution classification.

Step 5: Convert Classifications to Percentages
This set of programs converts classified high-resolution images into lower resolution percent cover images (figure A18) .

Recode Classified Image to Binary
The model in figure A19 converts the high-resolution classified image to a binary one. A value of one represents the class of interest, and zero represents everything else.
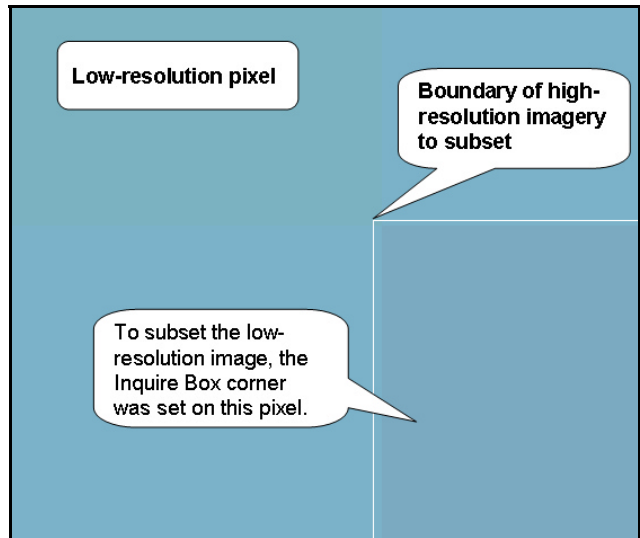

Figure A16—An illustration of the way to subset a high-resolution and low-resolution image so they match each other.
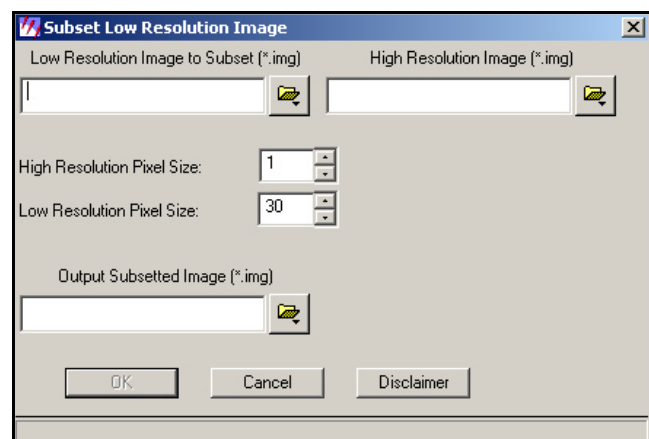

Figure A17—The ERDAS Imagine Subset Low-Resolution Image program, subsets a low-resolution image to match the boundaries of a high-resolution image.
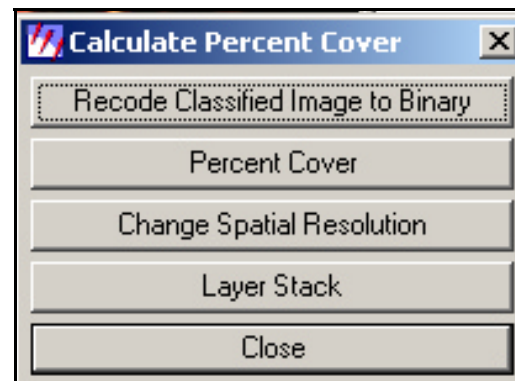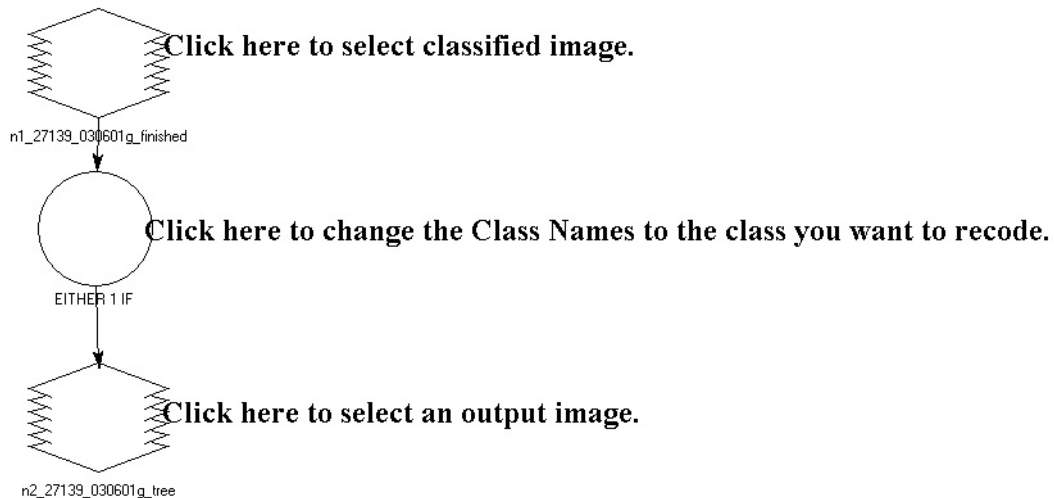

Figure A18—The instructions for ERDAS Image figure.

23

Recode Classified Image To Binary Image

Click here to select classified image.

n1_27139_030601g_finished

Click here to change the Class Names to the class you want to recode.

EITHER 1 IF

Click here to select an output image.

n2_27139_030601g_tree

Disclaimer:

Developed by the USDA Forest Service Remote Sensing Applications Center (RSAC). This program is freely distributable but is provided 'as is' and 'with all faults'. RSAC makes no warranties as to performance, merchantability, fitness for a particular purpose, or any other warranties whether expressed or implied.

Under no circumstances shall RSAC be liable for direct, indirect, special, incidental, or consequential damages resulting from the use, misuse, or inability to use this software, even if RSAC has been advised of the possibility of such damages.

Developed By:

Bonnie Ruefenacht, Ph.D.
USDA Forest Service Remote Sensing Applications Center
2222 West 2300 South
Salt Lake City, Utah 84119
bruefenacht@fs.fed.us 801-975-3828

A19—This ERDAS Imagine model converts a classified image to a binary one where the value of one represents the class of interest.

Percent Cover

The program in figure A20 calculates how many high-resolution pixels of the cover value are inside a single pixel of the lower resolution image. The Cover Value is the value of the class of interest. If the Recode Classified Image to Binary model was used, set this value to one. The output is a continuous image with a spatial resolution that matches the high-resolution image. The values represent percentages of cover.
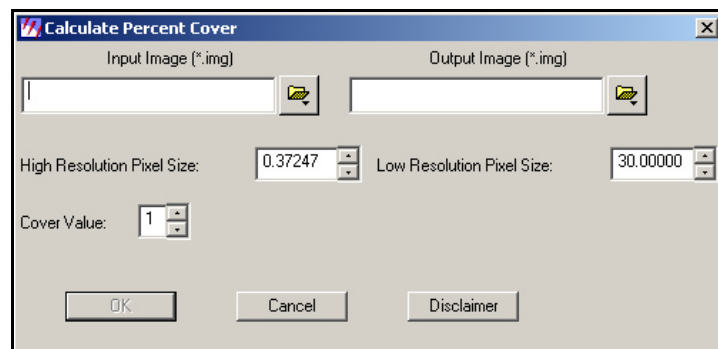
Figure A20—The ERDAS Imagine Calculate Percent Cover program, calculates how many high-resolution pixels of the cover value lie with in one low-resolution pixel and converts this number to a percentage.

24

Change Spatial Resolution

The spatial resolution of the output from the Percent Cover program now equals the high-resolution image. This spatial resolution, however, must be the same as the lower resolution image. Use the Change Spatial Resolution program (figure A21) to alter the pixel size of an image. The inputs to the program are self-explanatory.
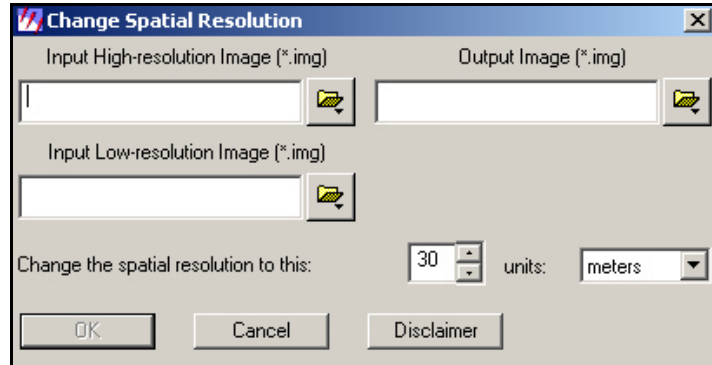

Figure A21—The ERDAS Imagine Change Spatial Resolution program, changes the cell size of the input image.

Layer Stack

The percent cover images, the lower resolution imagery, and the ancillary data need to be layer-stacked using ERDAS Imagine's Layer Stack program.

Step 6: Select Training/Testing Blocks

If desired, the imagery can be split into testing and training blocks. Cubist uses the testing blocks to test the accuracy of the training-data models. The Select Training/Testing Blocks program (figure A22) randomly chooses areas of the subsetted stacked imagery to use for training and testing. This program lets a user choose the Total Number of Blocks and the Number of Training Blocks. The Total Number of Blocks must be a perfect square root (e.g. 16, 25, 100, etc.). The Number of Training Blocks must be less than the Total Number of Blocks. The outputs are two images: one with the testing blocks, and the other with the training blocks (see figure A23).
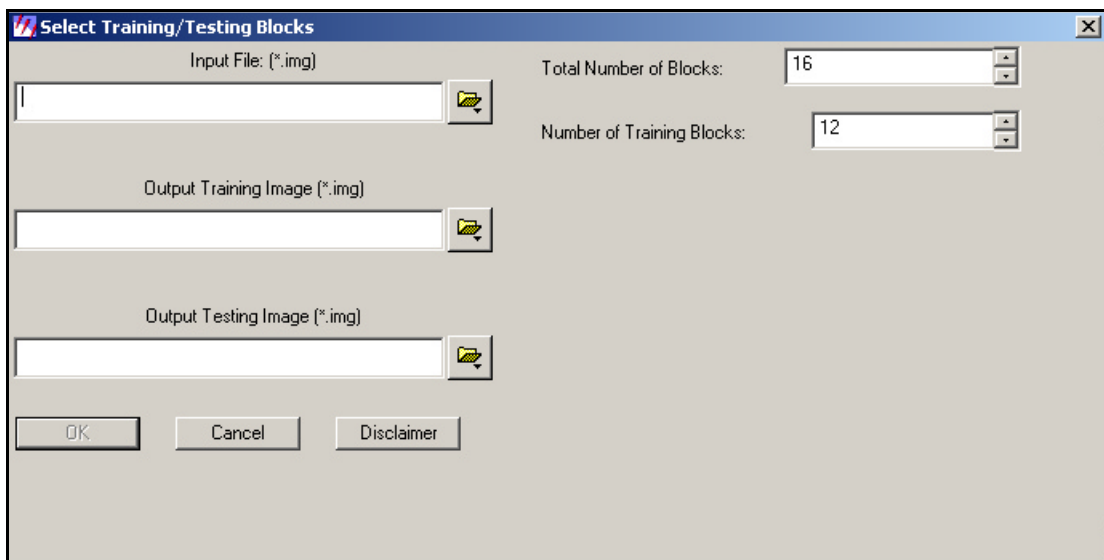

Figure A22—The ERDAS Imagine Select Training/Testing Blocks program, allows a user to divide an image into blocks so the program can randomly select them for testing and training. The Total Number of Blocks must be a perfect square root (e.g. 4, 9, 16, 25, etc.).
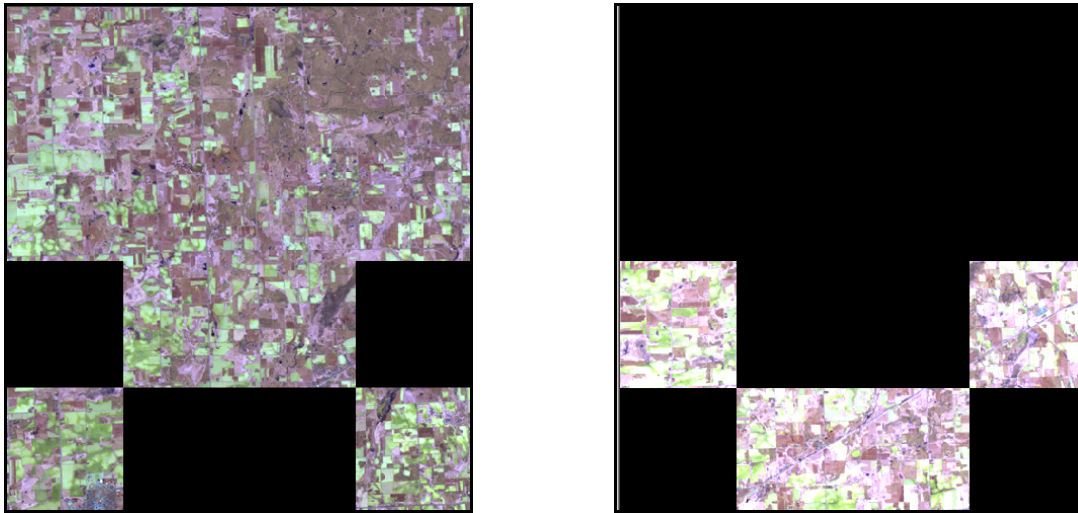
25

Figure A23—An example of the output from the ERDAS Imagine Select Training/Testing Blocks program. The image on the left shows 12 training blocks selected; the image on the right depicts 4 testing blocks selected.

Step 7: Prepare Data for Cubist

Cubist cannot read remote sensing data directly. The imagery must be converted to comma-delimited-text ASCII files. Use the Prepare Data for Cubist program (figure A24) to convert the imagery. The program creates two files, a .data file and a .names file. These two files are required Cubist inputs. If the user has testing data, a third file (.test) will be created as well.
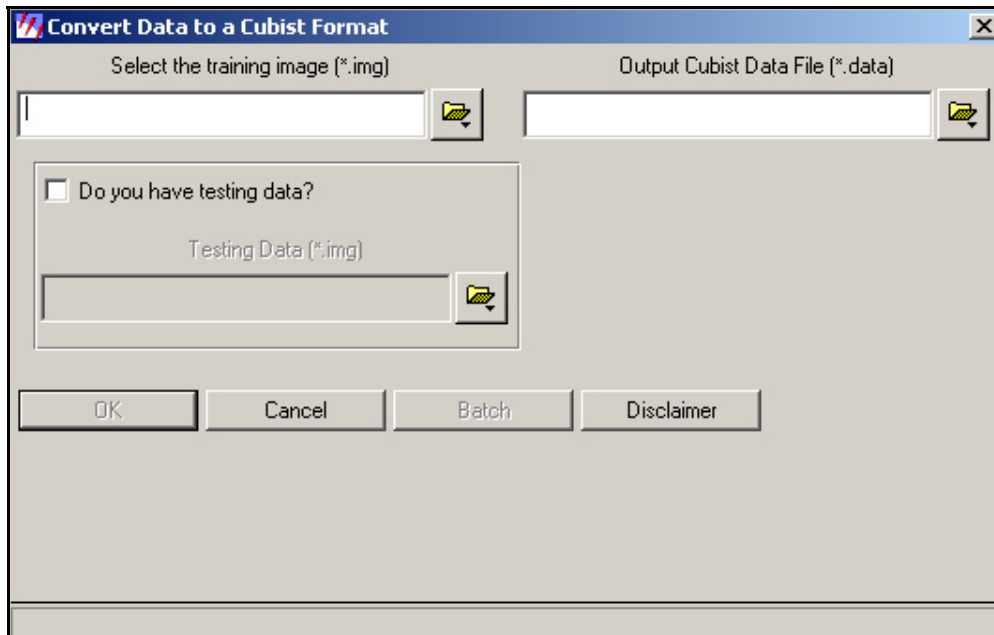

Figure A24—The ERDAS Imagine Prepare Data for Cubist program, converts an image into a comma-delimited-text ASCII file.

26

Step 8: Run Cubist

Cubist can be run outside of ERDAS Imagine, but to run it within Imagine, copy the LICENCE.DAT file to c:\users. For information on running Cubist, go to www.rulequest.com. Cubist can create models using three different methods: "rules alone," "instances and rules," and "let Cubist decide" (figure A25). If the Cubist models are going to be applied to a large data set, the "rules alone" option is recommended. The other options are time consuming. For example, the "instances and rules" option was originally selected to analyze zone 41. However, it would have taken more than 250 days to apply the Cubist models spatially. With the "rules alone" option, it took only 18 hours to apply them. Do not exclude attributes or ignore features. If you want to exclude or ignore items, edit the .data file in a spreadsheet program and delete the desired columns. Modify the .names file accordingly.

Figure A25—Cubist menu.

Step 9: Apply Cubist Results Spatially

The Apply Cubist Results Spatially program (figure A26) applies the Cubist models to the input image to produce a cover image. The input image must have the same number of bands as the number of variables used to build the Cubist model, excluding the dependent (or target) variable. The bands must be in the same order as the variables in the .names file. The names file should not contain any excluded or ignored attributes. If the target value is known, the dependent variable must be first in the .data file.
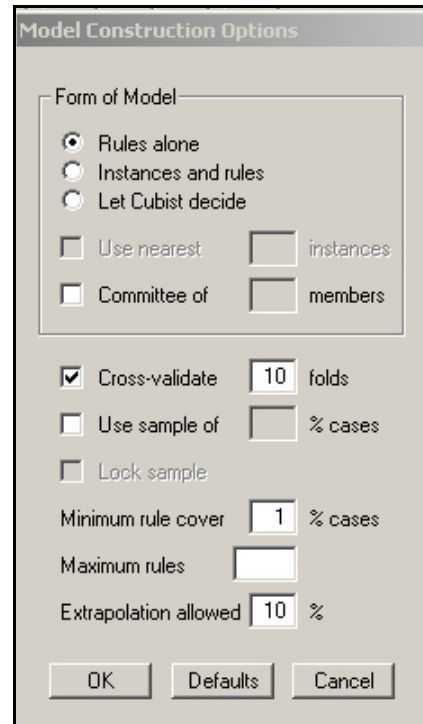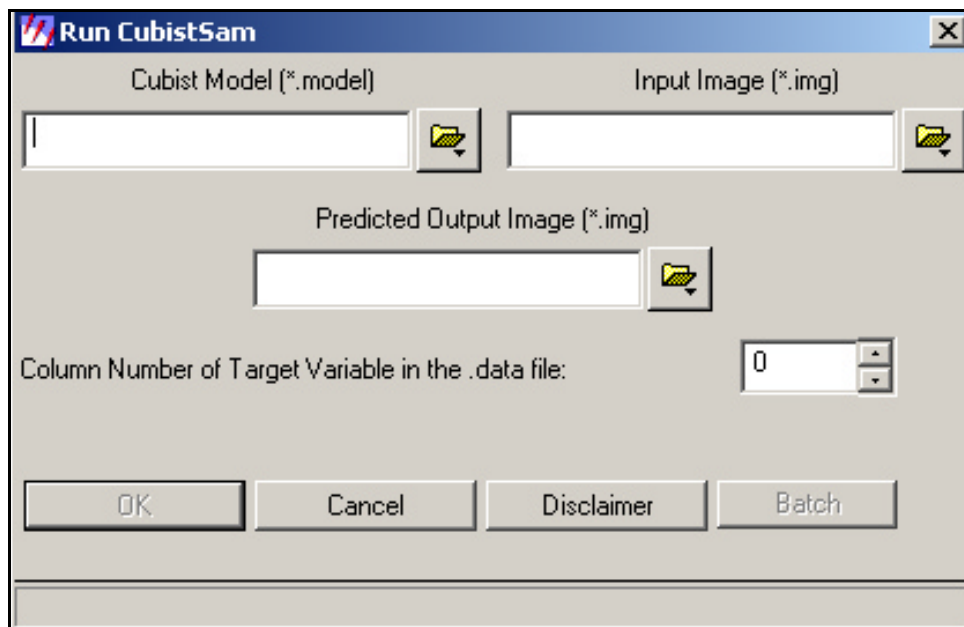

Figure A26—This program applies Cubist models to the input image and produces output image from the predictions.