

An Illustrated Methodology for Land Cover Mapping of Forests with Landsat-7 ETM+ Data

Methods in Support of EOSD Land Cover

**Version 3 Methods
March 2004**

Mike Wulder, Morgan Cranny, Jeff Dechka, and Joanne White
Canadian Forest Service
Pacific Forestry Centre, Victoria, BC, Canada

Please cite as:

Wulder, M., M. Cranny, J. Dechka and J. White, 2004; **An illustrated methodology for land cover mapping of Forests with Landsat-7 ETM+ data: Methods in Support of EOSD Land Cover, Version 3**, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada, March 2004, 35p.

Executive Summary

The following document describes the methodology used for classification of Landsat ETM+ data as part of Earth Observation for Sustainable Development of Forests (EOSD). Topics include pre-classification image preparation, detailed classification procedures (including aggregation/labelling), and post-classification processing.

The classification procedure involves pre-stratifying each image into four broad categories based on a normalized difference vegetation index (NDVI). Each category is then processed using an unsupervised K-means classifier. The input channels for the classification include the 6 optical Landsat-7 ETM+ channels as well as 1 texture measure. The texture is an intra-pixel variance derived from the 15m panchromatic data. The classification is labelled based on existing ground data and user knowledge. The EOSD classification legend consists of 22 land cover classes, plus one for NODATA. These classes have been adapted from those developed for the National Forest Inventory (NFI).

TABLE OF CONTENTS

1.0	INTRODUCTION.....	5
1.1	IMAGE PROCESSING	6
1.1.1	<i>Radiometric correction</i>	<i>6</i>
1.1.2	<i>Geometric correction</i>	<i>6</i>
1.1.3	<i>Image classification</i>	<i>6</i>
2.0	METHODOLOGY	9
2.1	IMAGE PREPARATION	9
2.1.1	<i>Automated image preparation</i>	<i>9</i>
2.1.2	<i>Manual image preparation</i>	<i>10</i>
2.1.2.1	Import source image files to .PIX files	10
2.1.2.2	Perform TOA correction	10
2.1.2.3	Create texture channel.....	11
2.1.2.4	Create NDVI channel.....	12
2.1.3	<i>Create image mask.....</i>	<i>15</i>
2.1.4	<i>Create cloud mask and remove clouds</i>	<i>17</i>
2.1.5	<i>Determine the hierarchical spectral regions from the NDVI</i>	<i>18</i>
2.2	IMAGE CLASSIFICATION.....	22
2.2.1	<i>Unsupervised K-means clustering</i>	<i>22</i>
2.2.2	<i>Merging and labelling of clusters</i>	<i>23</i>
2.2.3	<i>Combine classification results from the hierarchical spectral regions</i>	<i>29</i>
2.3	POST CLASSIFICATION PROCEDURES	31
2.3.1	<i>Slope from DEM</i>	<i>31</i>
2.3.1.1	Slope check of water / shadow.....	31
2.3.2	<i>Incorporation of other ancillary data sources.....</i>	<i>32</i>
3.0	QUALITY ASSURANCE	33
4.0	CREATION OF FINAL EOSD PRODUCT	33
	REFERENCES.....	34

List of Figures

Figure 1. General Classification Flow	8
---	---

List of Tables

Table 1. Stratification masks generated from the NDVI and associated classes.....	13
Table 2. EOSD land cover legend.....	25

1.0 Introduction

As part of the Earth Observation for Sustainable Development of Forests (EOSD) program, a land cover map of the forested area of Canada for year 2000 conditions is being developed. Implementation is on-going as of the date of this document's revision.

The strength of satellite remote sensing for land cover mapping is in the low cost per unit of land area associated with data capture and image analysis, as well as the large area coverage, and the ability to easily acquire data over geographically isolated areas. The classification of land cover over large geographic areas with remotely sensed data is increasingly common. An overview of the status and research priorities for large area mapping with satellites is found in Cihlar (2000). Nations (Loveland et al. 1991; Fuller et al. 1994; Cihlar and Beaubien 1998), continents (Stone et al. 1994), and the globe (Loveland and Belward 1997; Hansen et al. 2000; Loveland et al. 2000) have been mapped with a range of satellite data inputs and spatial resolutions. Unsupervised classification approaches are considered more suitable for large area mapping projects (Cihlar, 2000; Franklin and Wulder, 2002).

Landsat-7 Enhanced Thematic Mapper Plus (ETM+) imagery is the preferred data source for the estimation of NFI attributes based upon considerations such as the spatial resolution that will allow regional coverage with sufficient capturing of ground characteristics, the maturity of the pre- and post-processing techniques, data availability, relative low cost, thorough understanding of image characteristics by many technical analysts, and available software support (Franklin and Wulder, 2002). The project is also aided by the availability of a nationally consistent orthorectified image data set (Wulder et al., 2002).

Remotely sensed data has an implicit information content that is relative to the spatial resolution of the imagery. The image spatial resolution is often indicated through the pixel size. In the case of high spatial resolution imagery, with pixels of 1m on a side, individual trees may be isolated and subjected to information extraction. With the 30m spatial resolution of Landsat imagery, a number of trees are found within an individual pixel; as a result, the characteristics of individual trees and many within stand characteristics must be inferred. At the 30m spatial resolution, stand level characteristics are available from the digital imagery. The lower spatial resolution of the Landsat imagery is compensated for by the regional coverage offered. A single Landsat TM image covers 170km by 185km of the earth's surface (Lillesand and Kiefer, 1987). A thorough presentation of the historical and technical specifications of the Landsat sensors may be found in Lillesand and Kiefer (1987).

1.1 Image Processing

Prior to the analysis of optical remotely sensed data, a variety of considerations related to the integrity of the image data must be addressed to account for the geometric and radiometric characteristics of the imagery.

1.1.1 Radiometric correction

The concept of image radiometry is related to the spectral characteristics of the imagery. Radiometric corrections are applied to convert the sensor specific digital numbers to radiance values. Imagery received from the vendor is often in a format that has been converted to standard radiometric units or radiance values. The standard radiometric units are consistent in character for the entire image, however these units do not take into account atmospheric effects upon the imagery. Atmospheric effects (scattering, absorption, and emission) are a source of interference between the ground and the sensor. To compare imagery from different places or dates the atmospheric effects must be removed from the imagery. The magnitude of the atmospheric effects upon the signal received by the satellite varies by wavelength, resulting in a requirement for correction procedures appropriate to the wavelengths used (Avery and Berlin, 1992). An overview of radiometric correction procedures and algorithms may be found in Peddle et al. (2003).

1.1.2 Geometric correction

To locate ground features on imagery, or to compare a series of images, a geometric correction procedure is used to register each pixel to real world coordinates (Jensen 1996). A nearest neighbour geometric correction algorithm is suggested to best maintain the integrity of digital numbers through the transformation procedure (Jensen, 1996). The correction procedure used will be noted with the meta-data associated with each image processed. The majority of imagery used for the production of the EOSD classification is generated by the Centre for Topographic Information (CTI) of Natural Resources Canada. The CTI is producing a complete set of cloud free orthoimages covering the Canadian landmass (Wulder et al., 2002). The procedures outlined in the methodology section of this document assume that either geometrically corrected or orthorectified imagery is being used. If not, the user must complete this step prior to proceeding with the rest of the procedures outlined in this manual.

1.1.3 Image classification

The methodology that follows has been adapted from ‘A Guide to the Estimation of Canada’s National Forest Inventory Attributes from Landsat TM Data (Wulder *et al.*, 2001). Details on the classification system used for EOSD may be found in Wulder and Nelson (2003). Certain assumptions are associated with any procedure. For this classification we must assume that images are in the same projection and correctly orthorectified, that the images are radiometrically similar and have limited cloud cover (>5% area), and that the images are from similar seasons (ideally the height of the growing season).

Other issues to contend with for the project include heavy processing demands and high data storage capability. Canada requires approximately 600 images, after accounting for cloud cover, to represent the forested land base within Canada. Also note that each image will be in the neighbourhood of 1.5GB after classification.

This document outlines the general classification procedure starting with creation of the cloud mask, through NDVI thresholding, K-means classification under each mask, aggregation and labelling, to quality assurance of the final product.

A DEM can be an invaluable tool particularly for shadow removal (during a post-classification vetting phase). If using a DEM ensure that it has a resolution equal to or better than the resolution of the image. Acquiring a DEM for a particular area can be difficult; some elevation models are available for purchase over the web, for instance, DEM data for BC can be purchased the Land Data BC website <http://ldbcweb.landdata.gov.bc.ca:8001/LdbcSystem/index>.

We have tried to be as concise as possible in our description of classification steps. Additional elaboration on many of the steps followed may be found in a remote sensing text book or in the version 1 of the land cover mapping manual (Wulder *et al.*, 2001).

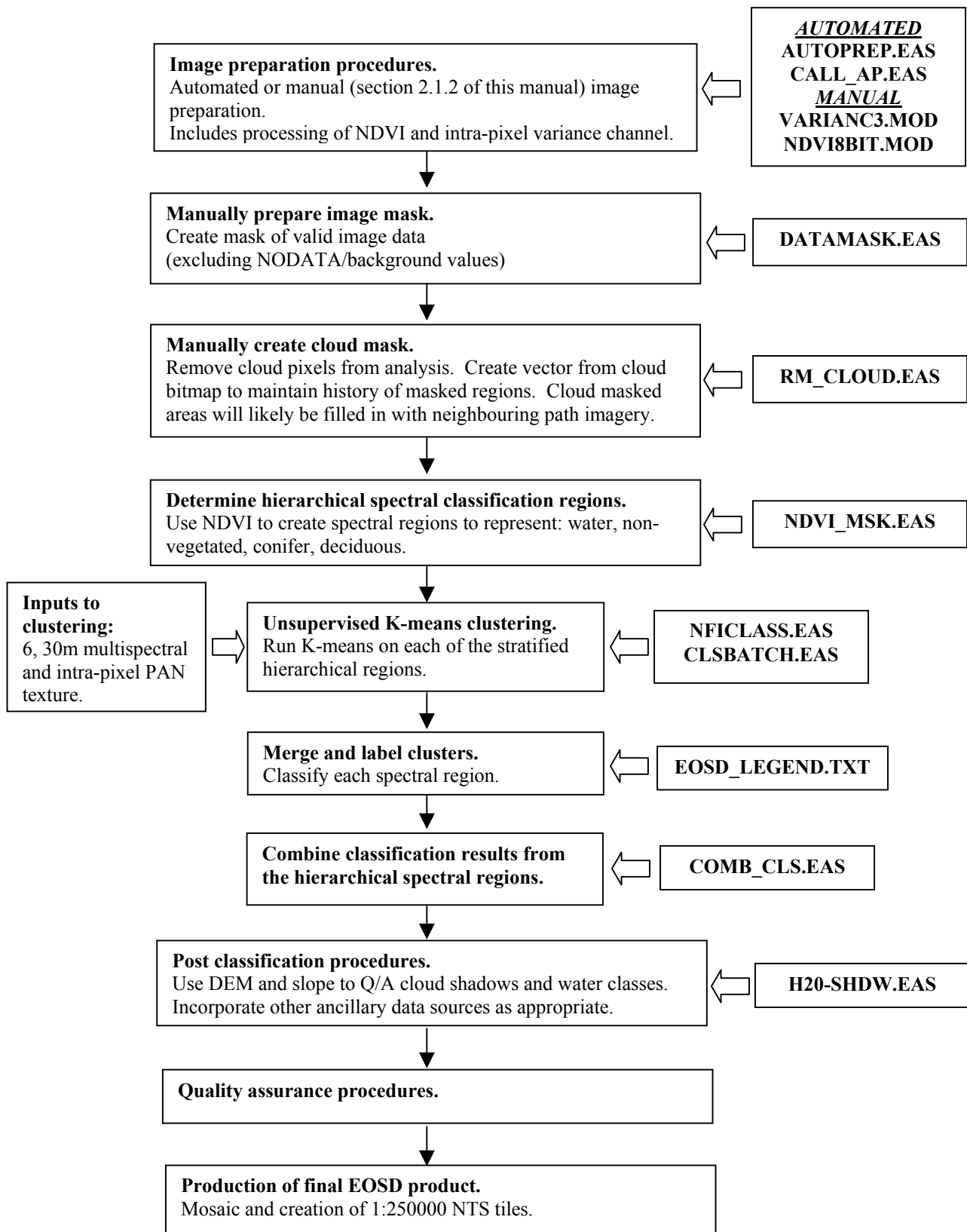


Figure 1. Procedural Flow of EO SD Methodology

2.0 **Methodology**

2.1 **Image preparation**

The following procedures may be used on images that have been radiometrically corrected or not. We suggest that you use radiometrically corrected images, particularly if you are creating a mosaic. You will require Landsat-7 ETM+ multispectral and panchromatic channels to follow the suggested methodology. There are two approaches: an automated approach and a manual approach. The automated approach involves the use of a script that automates the preparation of the image files for processing, performs the TOA correction and calculates the intra-pixel variance from the 15m panchromatic data. Alternatively, a manual approach may be used and the steps required to complete the preparation are outlined below. The EASI library files **NFILIB.EAS** and **EOSD_LIB.EAS** are required to run the scripts described in the procedures.

2.1.1 *Automated image preparation*

Many of the image preparation steps detailed below in the manual image preparation section may be completed using an automated script, **AUTOPREP.EAS**. This script does the following:

1. Imports CTI .TIFF files to a PCI .PIX file and backs up the .TIFF files;
2. Performs TOA correction on 30m multispectral Landsat channels;
3. Prepares 30m and 15m .PIX files (adds channels);
4. Calculates the intra-pixel variance for the 15m panchromatic Landsat channel and resets NODATA values;
5. Resamples 15m intra-pixel variance channel to 30m and add to 30m .PIX file;
6. Runs NDVI over the 30m .PIX file.

CALL_AP.EAS is a script which allows the **AUTOPREP.EAS** script to run on a master directory containing several image data directories. **CALL_AP.EAS** assumes that a certain naming convention has been used for the image files. File names must adhere to the EOSD modified naming convention, as follows:

CTI naming convention:	EOSD modified naming convention:
037014_0100_010727_17_01_utm15.tif	17_3714_20010727_utm15_b10.tif
PPPRRR_0100_YYMMDD_17_0C_utmNN.tif	PPPRRR_0100_YYMMDD_17_0C_utmNN.tif

Where C = channel #; PPP = Landsat path; RRR = Landsat row; NN = UTM zone

2.1.2 Manual image preparation

2.1.2.1 Import source image files to .PIX files

Add 16 8-bit unsigned and 1 16-bit signed channels to the 30m TM image using **PCIMOD**. Set the *PCIOP* to ADD and *PCIVAL* to 16, 1, 0, 0 and run **PCIMOD**.

```
PCIMOD  PCIDSK Database File Modification      V7.0 EASI/PACE  10:55 26Mar2004
FILE    - Database File Name                   :17_4922_utm10_30m_demo.pix
PCIOP   - Modification Option: ADD/DEL/COM:ADD
PCIVAL  - DBNC for ADD, or DBIC for DEL  }      16      1      0      0
```

Add 1 8-bit and 1 32-bit real channel to the PAN image using **PCIMOD**. Set the *PCIOP* to ADD and *PCIVAL* to 1, 0, 0, 1 and run **PCIMOD**.

```
PCIMOD  PCIDSK Database File Modification      V7.0 EASI/PACE  11:51 26Mar2004
FILE    - Database File Name                   :17_4922_utm10_15m_demo.pix
PCIOP   - Modification Option: ADD/DEL/COM:ADD
PCIVAL  - DBNC for ADD, or DBIC for DEL  }      1      0      0      1
```

2.1.2.2 Perform TOA correction

Top of atmosphere (TOA) reflectance is calculated using radiometry procedures developed at the Canadian Centre for Remote Sensing. For further information on radiometric image processing, refer to Peddle et al. (2003). Note that TOA scripts are automatically called and executed when the AUTOPREP.EAS script is run.

To obtain the most recent version of TOA executables, see:

<ftp.ccrs.nrcan.gc.ca/ad/landry>, or contact Robert Landry directly

(Robert.Landry@nrcan.gc.ca). The TOA executables must be placed in the appropriate executable directory (depending on your operating system).

The TOA executables may be run manually by executing the following commands in an EASI shell script:

```
TOARETM      V8.2 EASI/PACE  12:18 26Mar2004
FILE    - Database File Name                   :17_4922_utm10_30m_demo.pix
DBIC    - Database Input Channel List          >      1      2      3      4
          5      6
ETMBAND  - Landsat ETM+ Spectral Band          >      1      2      3      4
          5      7
DBOC    - Database Output Channel List         >      7      8      9     10
          11     12
MTLFILE  - Metadata Filename <xxx_MTL.L1G>    :L71049022_02219990912_MTL.L1G
TIME     - GMT Time of Acquisition <hh,mm>    >      19      5
TOARSCM  - Scaling Method <VARIABLE, FIXED>   :FIXED
PROTYPE  - Processing type: REPORT/FULL        :FULL
```

Note that TIME represents the hours and minutes at the scene centre time. This information may be found in the image metadata (or in the PGS report if CTI imagery is being used).

2.1.2.3 Create texture channel

The intra-pixel texture measure is computed from a 3x3 variance of the 15m panchromatic band. This generates the intra-pixel variability measure. The texture of the 15m panchromatic data is resampled to 30m for inclusion as supplemental data in the K-means clustering process. The result is a texture channel that represents the contents of a 30m pixel. Texture channels that incorporate the information from windows of several 30m pixels often result in artefacts in the clusters that are subsequently generated. The impact of these types of artefacts can be seen in Wulder *et al.* (2004), where the inclusion of elevation data as an input channel in the unsupervised clustering is illustrated.

The intra-pixel texture is created from the PAN image using the modelling script **VARIANC3.EAS**. The input will be the 8-bit channel that contains the Panchromatic data and the output will be the empty 32R.

```
VARIANC3 Computes 3X3 variance 12:05 26Mar2004
FILE - Database File Name :17_4922_utm10_15m_demo.pix
DBIC - Database Input Channel List > 1
DBOC - Database Output Channel List > 3
EASI>r varianc3
```

The resultant 32R channel needs to be scaled to 8-bit using **SCALE**. Run **HIS** first to determine the minimum and maximum values of the 32R channel to use as *INRANGE* in **SCALE**.

```
HIS Histogramming Database Image V9.1 EASI/PACE 12:09 26Mar2004
FILE - Database File Name :17_4922_utm10_15m_demo.pix
DBIC - Database Input Channel List } 3
GMOD - Graphic Mode: ON/OFF :OFF
CMOD - Cumulative Mode: ON/OFF :OFF
PCMO - Percentage Mode: ON/OFF :OFF
NSAM - Number of Pixel Values to Sample>
TRIM - Tail Trimming % (Left, Right) >
HISW - Histogram Window >
MASK - Area Mask (Window or Bitmap) >
REPORT - Report Mode: TERM/OFF/filename :TmpNam()
```

The *FILI* and *FILO* are the PAN image, *OUTRANGE* is 0 to 255, *DBIC* is the 32R channel and *DBOC* an empty 8U channel, *SFUNCT* is NQ (Normalised Quantization). The *INRANGE* is stored in internal IMSTAT variables (so the *INRANGE* is specified as IMSTAT[11], IMSTAT[12]).

SCALE		Scale Image Data		V9.1 EASI/PACE 12:16 26Mar2004	
FILI	- Database Input File Name	:	17_4922_utm10_15m_demo.pix		
FILO	- Database Output File Name	:	17_4922_utm10_15m_demo.pix		
DBIC	- Database Input Channel List	}	3		
DBOC	- Database Output Channel List	}	2		
DBIW	- Database Input Window	>			
DBOW	- Database Output Window	>			
INRANGE	- Input Range (Min,Max)	>	0	7371	
TRIM	- Tail Trimming % (Left, Right)	>			
OUTRANGE	- Output Range (Min,Max)	>	0	255	
SFUNCT	- Scaling Function or NQ/EQ	:	NQ		

Background values in the output channel need to be reset to 0, this can be done with a simple MODEL step.

MODEL		Modelling Program		V9.1 EASI/PACE 12:18 26Mar2004	
FILE	- Database File Name	:	17_4922_utm10_15m_demo.pix		
SOURCE	- Source of Model	:	if %3 = 0 then %1 = 0; endif		
UNDEFVAL	- Value for Undefined Operations	>	0		
REPORT	- Report Mode: TERM/OFF/filename	:			

Once the data has been scaled it needs to be transferred to the TM image. This is best accomplished using **REGPRO**. *FILI* is the PAN image and *FILO* is the TM image. Both *INGEO* and *OUTGEO* should be set to 1. *DBIC* is the scaled intra-pixel data while *DBOC* is an empty 8U channel. We suggest setting *RESAMPLE* to BILINEAR. To make **REGPRO** work more quickly, set the *MEMSIZE* to the highest available RAM memory (i.e. 512 MB).

REGPRO		Image Registration and Projection		V9.1 EASI/PACE 12:22 26Mar2004	
FILI	- Database Input File Name	:	17_4922_utm10_15m_demo.pix		
INGEO	- Input Georeference Segment	>	1		
DBIC	- Database Input Channel List	}	2		
DBGC	- Database Ground Control Segment	>			
FILO	- Database Output File Name	:	17_4922_utm10_30m_demo.pix		
OUTGEO	- Output Georeference Segment	>	1		
DBOC	- Database Output Channel List	}	7		
RESAMPLE	- Resample Mode: NEAR/BILIN/CUBIC	:	BILI		
ORDER	- Order of Polynomial Function	>			
SAMPLING	- Sampling Interval	>	1		
MEMSIZE	- Working Memory Size (Mbytes)	>	512		

2.1.2.4 Create NDVI channel

Stratification of cluster regions is based on a normalized difference vegetation index (NDVI). Four strata are derived from this product based on “biomass” and are listed in the following table.

Table 1. Stratification masks generated from the NDVI and associated classes.

Mask	Approximations of Associated Classes
Non-vegetated (non-land)	Water
Non-vegetated (land)	Exposed Land
Low Reflectance Vegetation	Coniferous
High Reflectance Vegetation	Deciduous / Herb

NDVI is based on the on the following equation

$$\text{NDVI} = [(bnd4 - bnd3) / (bnd4 + bnd3)] \quad (1)$$

$$\text{Scaled NDVI} = [\text{NDVI} + 1] * 127 \quad (2)$$

The NDVI equation results in a range of values between -1 to $+1$. Scaling allows for storage economisation and the use of an 8-bit image display channel.

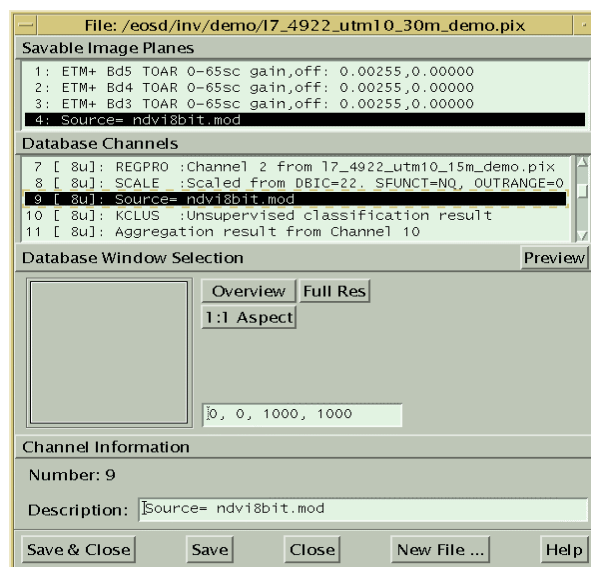
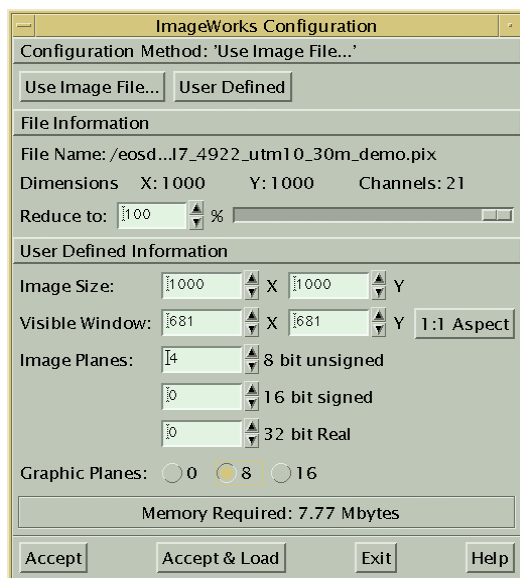
The **EASI** modelling script **NDVI8BIT.EAS** is used to create the NDVI channel. The input channels are the red and infrared channels of the TM image and the output is an empty 8U channel.

```

NDVI8BIT Scaled NDVI                                     12:26 26Mar2004
FILE      - Database File Name                           :17_4922_utm10_30m_demo.pix
DBIC      - Database Input Channel List                   >      3      4
DBOC      - Database Output Channel List                  >      8
BACKVAL   - Background Grey-level Value                  >      0

```

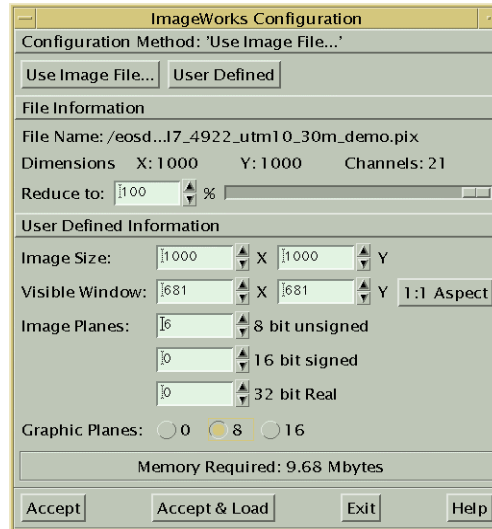
Once the NDVI has been created, the 0 data underneath the image mask should be converted to 1, this will help in the vegetation and cloud mask step. Start **Imageworks** with 4 image planes and 8 graphic planes.



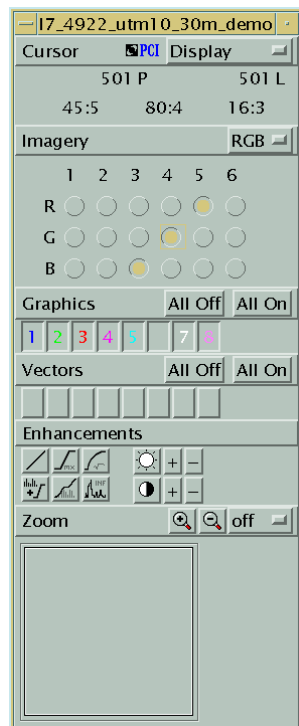
Regardless of whether the manual or automated approach to image preparation have been used, the following processing steps must be completed manually: creation of an image mask that identifies valid data areas in the image (as opposed to background NODATA values).

2.1.3 Create image mask

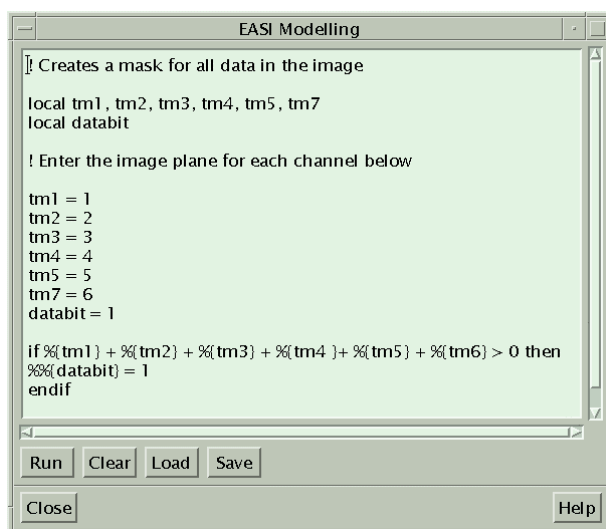
An image mask is required for a couple of steps. To create the mask, start **Imageworks** and select the TM image (*Use Image File...*) at full resolution with 6 channels and 8-bitplanes. *Accept & Load*, to fill the image planes with the default values.



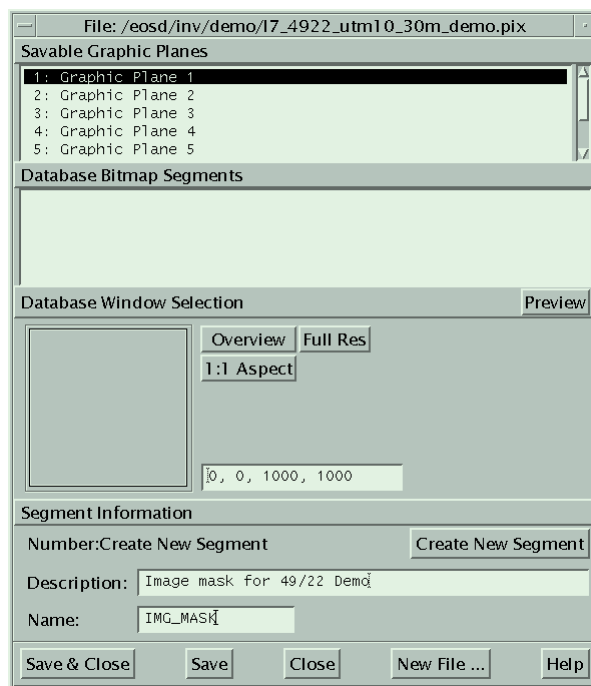
Set RGB to 5,4,3 and apply an equalization enhancement.



Select *Tools|Modelling...* from the main image window toolbar. In the new **EASI** modelling window Load the **DATAMASK.EAS** script and *Run*.

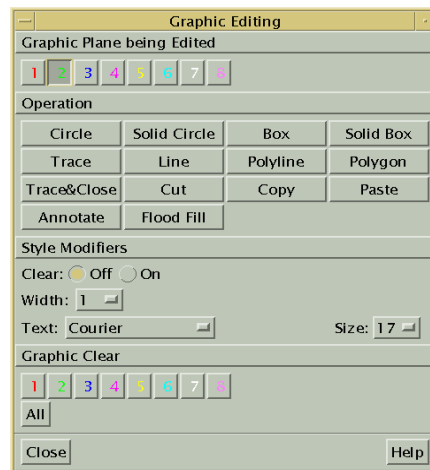


This will create a bitplane that is activated where all image channels sum to more than 0. Use *File|Save Graphics...* to write the new image mask bitmap to an empty graphic plane in the image. Make a note of which segment is saved to by choosing *Save* and then *Close* the window when complete.

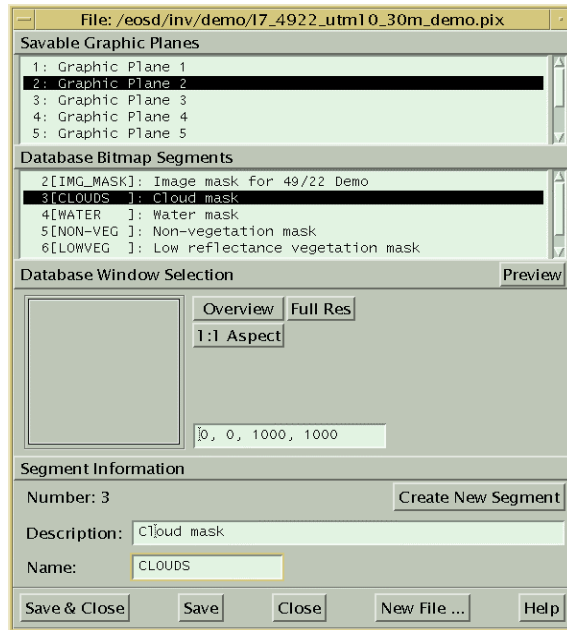


2.1.4 Create cloud mask and remove clouds

The cloud mask is created manually as the transition from cloud to less cloud, and haze, is problematic. Manual masking incorporates shadow, cloud, and haze. *File|Load Image...* the NDVI channel and TM 5,4,3 into the available image planes. Once loaded, select only the NDVI image plane in B&W, apply an equalization enhancement and create a zoom window. Then set RGB to be 5,4,3 and apply an equalization enhancement. Open the graphic edit window, *Edit|Graphic...*, and choose an empty graphic plane to work with.



Find cloud, haze and associated shadow in the image. It is better to take too much than too little, particularly if there is considerable overlap between images. Digitize around the cloud area using *Trace&Close* and then *Flood Fill* each polygon. Take care doing this step as there is no undo. It is also advised that you save the graphic channel, *File|Save Graphic...* from the main image window often to avoid frustration.



Once the cloud mask has been created and the TM image has been backed up, it is time to remove the data underneath the clouds. Again this can be accomplished using a simple script file, **RM CLOUD.EAS**. This file will convert to 0 values, any data underneath the cloud mask for all 8 existing channels in the TM image. DBIB should equal the segment that contains the cloud mask bitmap.

```

RM CLOUD Remove cloud from image under bitmap          12:30 26Mar2004
FILE  - Database File Name                          :17_4922_utm10_30m_demo.pix
DBIC  - Database Input Channel List                   >          1          -8
DBIB  - Database Input Bitmap                         >

```

2.1.5 Determine the hierarchical spectral regions from the NDVI

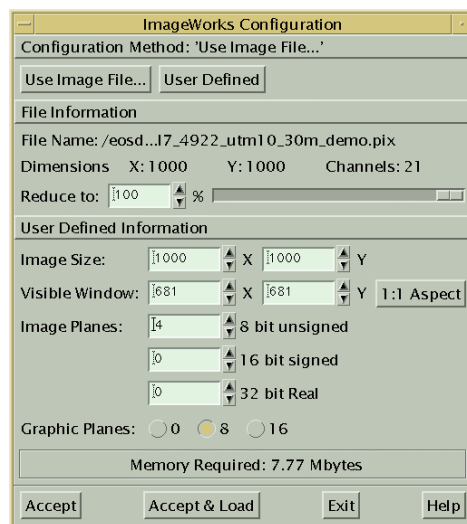
Once the NDVI is complete, the next phase of the process involves manual assessment of the image to create mask regions. The break for non-vegetated was determined after testing various thresholds using a manual approach that involved density slicing to decide what “cut-off” should be used to divide the image into the four masks. The values may vary from region to region.

A valuable tool in the assessment of the image is to use the NDVI histogram. Overlapping land cover types can, and will, occur in each mask however are labelled according to their class regardless of the parent mask. For example, if some fringe “exposed land” classes are clustered in the “non-vegetated non-land mask” they are still labelled “exposed land”. Therefore, the definitive edge of the mask is not as critical as separating general groups of vegetation and non-vegetation communities with the intended purpose of increasing the number of clusters generated using K-means in order to improve separability between classes.

Masks are created using a bitmap modelling approach, identifying a suitable range and applying it to a graphic channel. To deal with outside 0 values only those pixels under the image mask are included in the NDVI thresholding. The NDVI channel was further processed by converting all 0 values that fell under the image mask but not under the cloud mask to 1.

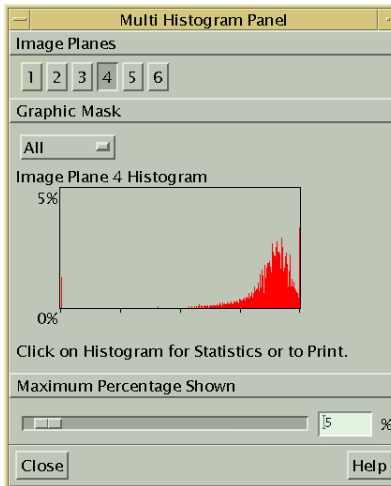
Once the masks have been generated, a histogram sample for each mask should be obtained so that the total pixels in each mask can be summed. This summation should equal the total image pixels and exclude the bounding black fill area.

In **Imageworks**, select the TM image at 100% resolution with 4 8U channels and 8 graphic planes and choose *Accept*.

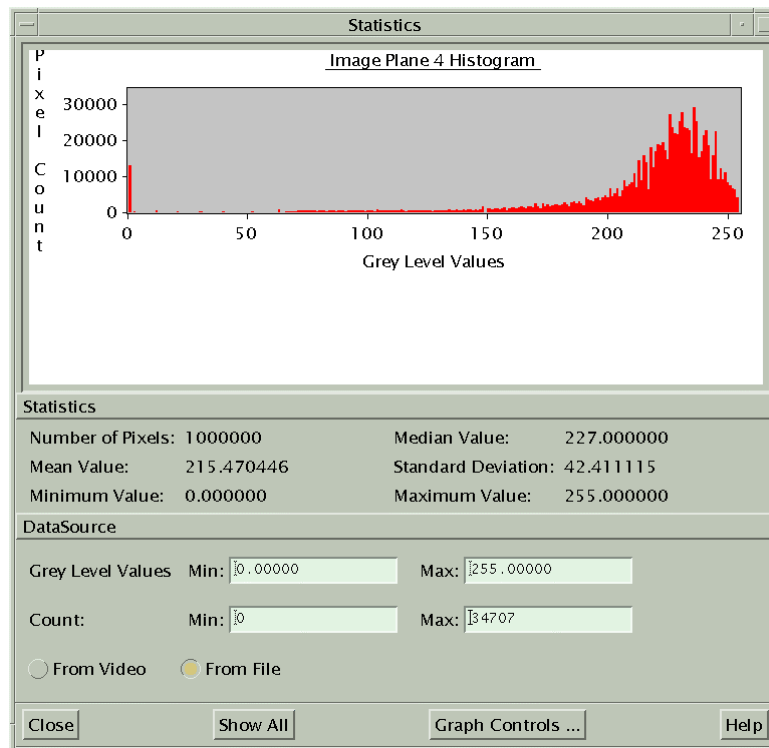


Once the image window and control panel appear, *File|Load Image...* , to load the NDVI channel into image plane 1, and 5,4,3 into the remaining image planes. *Load&Close*, select 2,3,4 for RGB and apply an equalization enhancement. Create a zoom window.

The NDVI histogram is a very useful tool, choose *View|Histograms...* to open the **Multi Histogram Panel**.

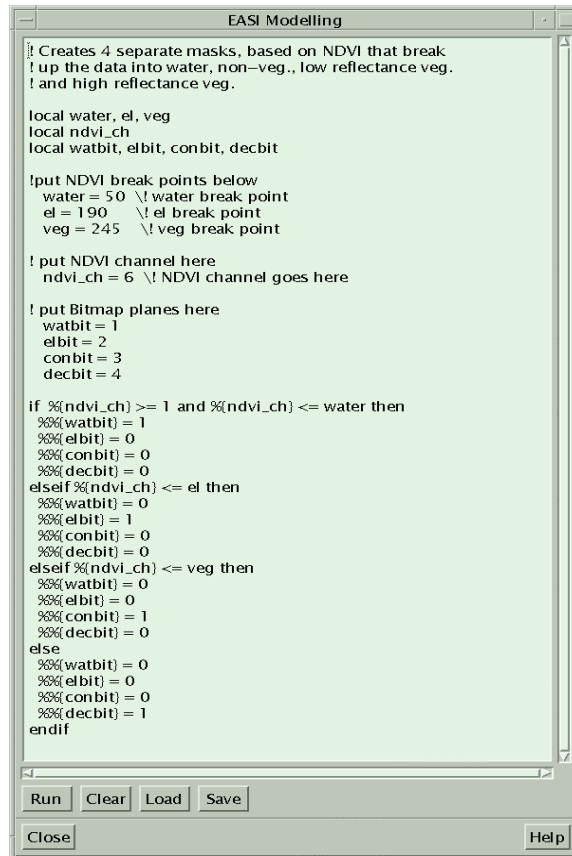


Double click the histogram to open the **Statistics** window.

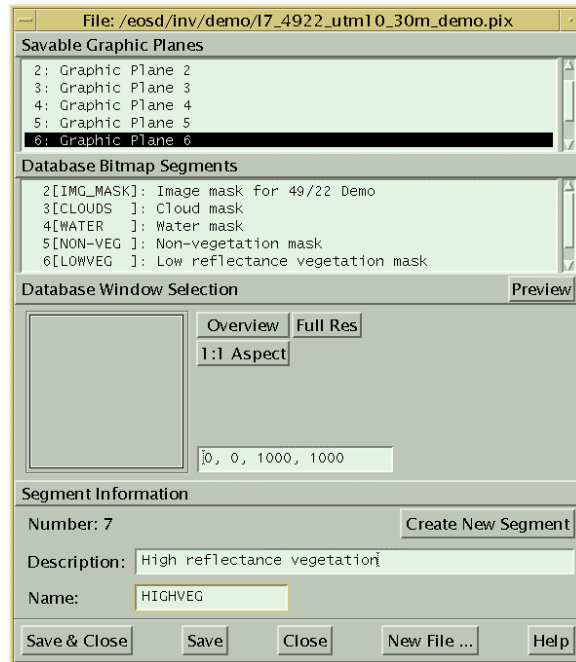


The statistics window gives a more detailed view of the histogram which can be zoomed into and out of. It also gives *Number of Pixels*, *Median Value*, *Mean Value*, *Standard Deviation*, *Minimum* and *Maximum Value*. Be sure to select the *From File* button. To zoom in, press and hold the left mouse button and create a box around the desired area. To zoom out, press the right mouse button and choose the desired option.

Switch to B&W and load the NDVI channel and apply the same stretch. Open *Tools|Modelling...* and load the **NDVI_MSK.EAS** script.



This script should be adjusted to match the channels you loaded. The script essentially activates an empty bitplane based on a specified DN range. As previously mentioned, the ranges will be for water, non-vegetation, low reflectance vegetation, and high reflectance vegetation. Once the bitmaps have been created, *File|Save Graphics...* to save them to the image.



2.2 Image Classification

2.2.1 Unsupervised K-means clustering

In this step, the **NFICLASS.EAS** script will be run for each of the vegetation masks in a batched process. The batch script, **CLSBATCH.EAS**, runs the NFICLASS.EAS script on a file using several masks. Provided the image input files are in the required standard format (see section 2.1.1), the user should not have to modify this script. Processing of the NFICLASS.EAS script can be lengthy. For a large image this process can take quite some time to process, exceeding 5 hours in some cases (depending on local computing environments).

Essentially this is K-means clustering underneath a mask. It also creates LUTs and a PCT for each classification which aid in labelling. The input parameters for the K-means classifier include:

NUMCLUS = 241;
MOVETHRS = 0.1;
MAXITER = 12;
NSAM = 50% of pixels under mask
DBIC include TM 1, TM 2, TM 3, TM 4, TM 5, TM 7, and tex1.

2.2.2 Merging and labelling of clusters

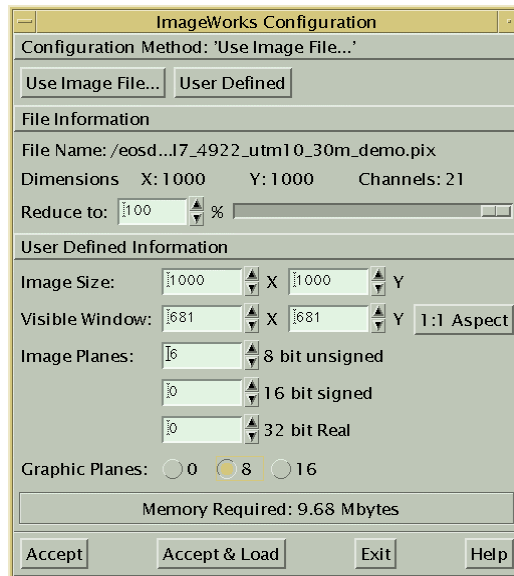
Labelling is based on the analysts assessment of clusters with the aid of training data (image visualization, maps, local knowledge, etc.).

The auto-aggregation procedure from earlier methods has been removed. Instead, each cluster is evaluated individually. The automated procedure often grouped clusters that should not have been grouped. Grouping of all clusters does not require much additional effort as many clusters that are appropriate to group together, are obvious. This also aids the analyst in getting used to the clusters present and the classes represented. This is usually the most time consuming and user intensive step. There will be up to 241 classes for each of the classification masks but usually there are between 150 and 200 classes to label for each of the masks.

If the area to be classified will be a mosaic of several scenes, choose a master image. This image should represent an area of particular interest and have as much ground truth information as possible. If using forest cover data, care should be taken as it can be out of date and too general for some classes. Successive images should be labelled using the previous classification as a guideline. All classification labelling should be targeted to match the master image. This will help to create a seamless mosaic. Of course phenological differences between scenes, based on time of capture, have to be taken into account and may affect the labelling procedure.
























Class labelling is based on the EOSD legend (Table 2). It must be recognized that not all classes listed in the EOSD legend may be contained in a single image. Furthermore some classes may be too similar to be labelled individually and may be combined. In the pilot study for instance, rock/rubble was combined with the exposed land class and many of the wetland classes did not get stratified consistently in the clustering procedure. To enable integration of the EOSD land cover of the forested area of Canada with classifications produced to represent other sectors (such as agricultural areas and Canada's north), specific digital numbers for each class are required. Therefore classes must be assigned according to the class numbering laid out in the [EOSD legend file](#).

For labelling, start **Imageworks** and select the image to be classified (*Use Image File...*) at 100% resolution with 6 channels and 8 graphic planes and *Accept*.

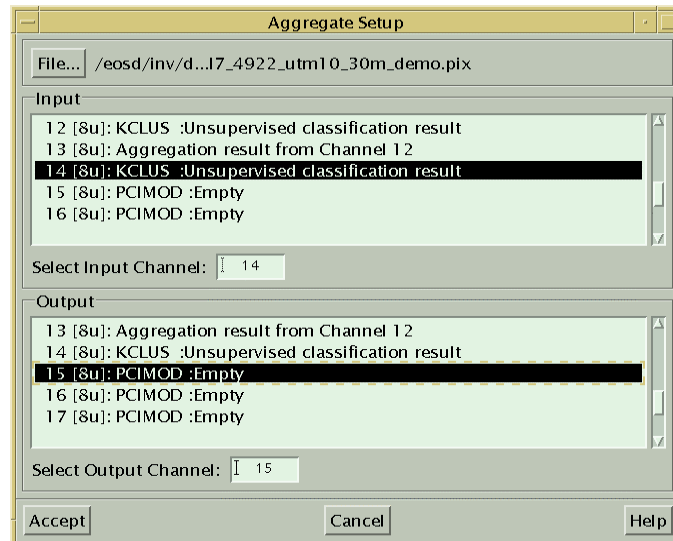


File|Load Image..., load bands 5,4,3,2,1 into image planes 2,3,4,5,6, *Load & Close*. If there are any vectors to be loaded use the *File|Load Vector...* to load those needed. Select 2,3,4 for RGB and apply adaptive enhancement, and create a zoom window, *View|Zoom Window...* Then select 4,5,6 for RGB and apply the same enhancement and create a second zoom window. These will be used to help you in the aggregation and labelling process. Position them on screen so that they can be seen easily.

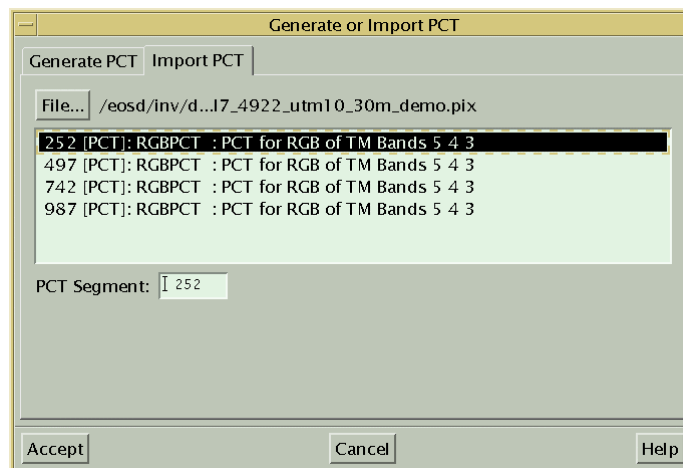
Table 2. EOSD land cover legend.

	CLASS	DESCRIPTION
	No Data	
	Shadow	
	Cloud	
	Snow/Ice	Glacier/snow
	Rock/Rubble	Bedrock, rubble, talus, blockfield, rubbly mine spoils, or lava beds.
	Exposed Land	River sediments, exposed soils, pond or lake sediments, reservoir margins, beaches, landings, burned areas, road surfaces, mudflat sediments, cutbanks, moraines, gravel pits, tailings, railway surfaces, buildings and parking, or other non-vegetated surfaces.
	Water	Lakes, reservoirs, rivers, streams, or salt water.
	Shrub - Tall	At least 20% ground cover which is at least one-third shrub; average shrub height greater than or equal to 2 m.
	Shrub - Low	At least 20% ground cover which is at least one-third shrub; average shrub height less than 2 m.
	Herb	Vascular plant without woody stem (grasses, crops, forbs, gramminoids); minimum of 20% ground cover or one-third of total vegetation must be herb.
	Bryoids	Bryophytes (mosses, liverworts, and hornworts) and lichen (foliose or fruticose; not crustose); minimum of 20% ground cover or one-third of total vegetation must be a bryophyte or lichen
	Wetland - Treed	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is coniferous, broadleaf, or mixed wood.
	Wetland - Shrub	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is tall, low, or a mixture of tall and low shrub.
	Wetland - Herb	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is herb.
	Coniferous - Dense	Greater than 60% crown closure; coniferous trees are 75% or more of total basal area.
	Coniferous - Open	26-60% crown closure; coniferous trees are 75% or more of total basal area.
	Coniferous - Sparse	10-25% crown closure; coniferous trees are 75% or more of total basal area.
	Broadleaf - Dense	Greater than 60% crown closure; broadleaf trees are 75% or more of total basal area.
	Broadleaf - Open	26-60% crown closure; broadleaf trees are 75% or more of total basal area.
	Broadleaf - Sparse	10-25% crown closure; broadleaf trees are 75% or more of total basal area.
	Mixed Wood - Dense	Greater than 60% crown closure; neither coniferous nor broadleaf tree account for 75% or more of total basal area.
	Mixed Wood - Open	26-60% crown closure; neither coniferous nor broadleaf tree account for 75% or more of total basal area.
	Mixed Wood - Sparse	10-25% crown closure; neither coniferous nor broadleaf tree account for 75% or more of total basal area.

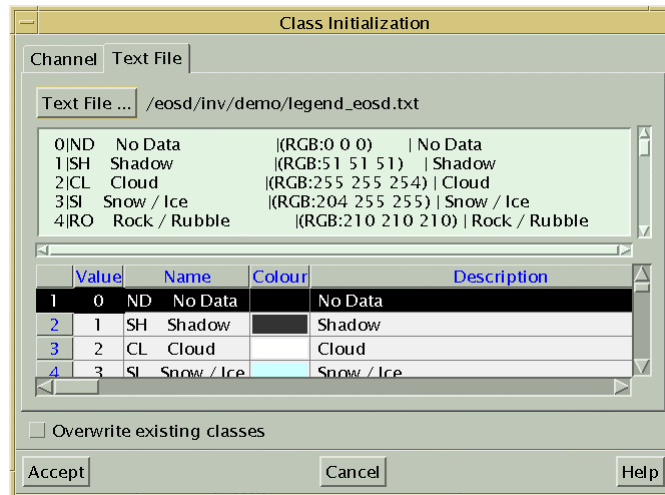
Initiate the aggregation process, *Classify|Aggregate...*, and choose the first classified channel for *Input* and an empty 8-bit channel for *Output*, *Accept*.



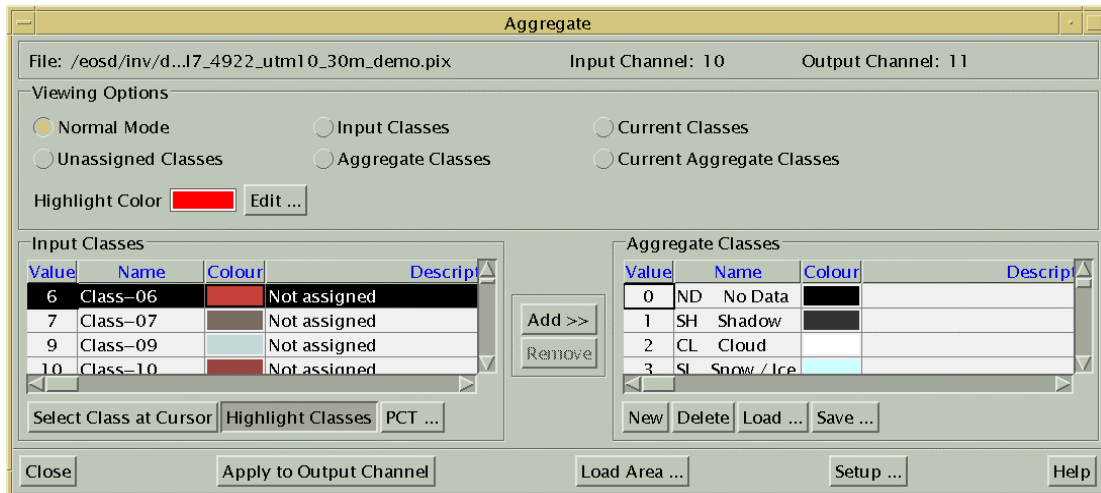
When the Aggregate window pops up select *PCT...* and switch to the *Import PCT* tab, load the current file, *File...* Select the appropriate PCT segment and *Accept*.



Under *Aggregate Classes*, select *Load...*, switch to the *Text File* tab and select *Text File...* choosing the **EOSD legend file** from the list and *Accept*.



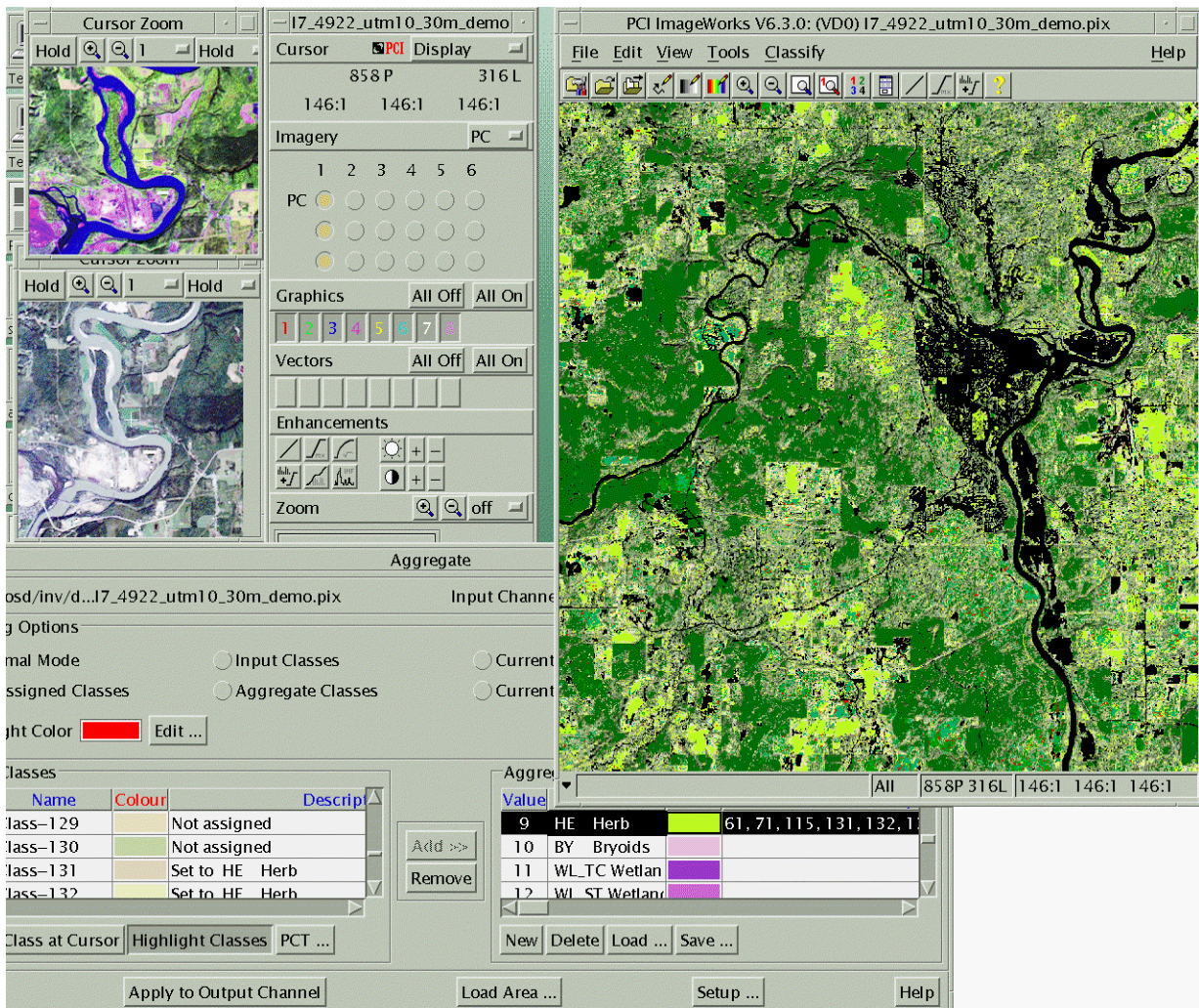
The aggregation window should look like this



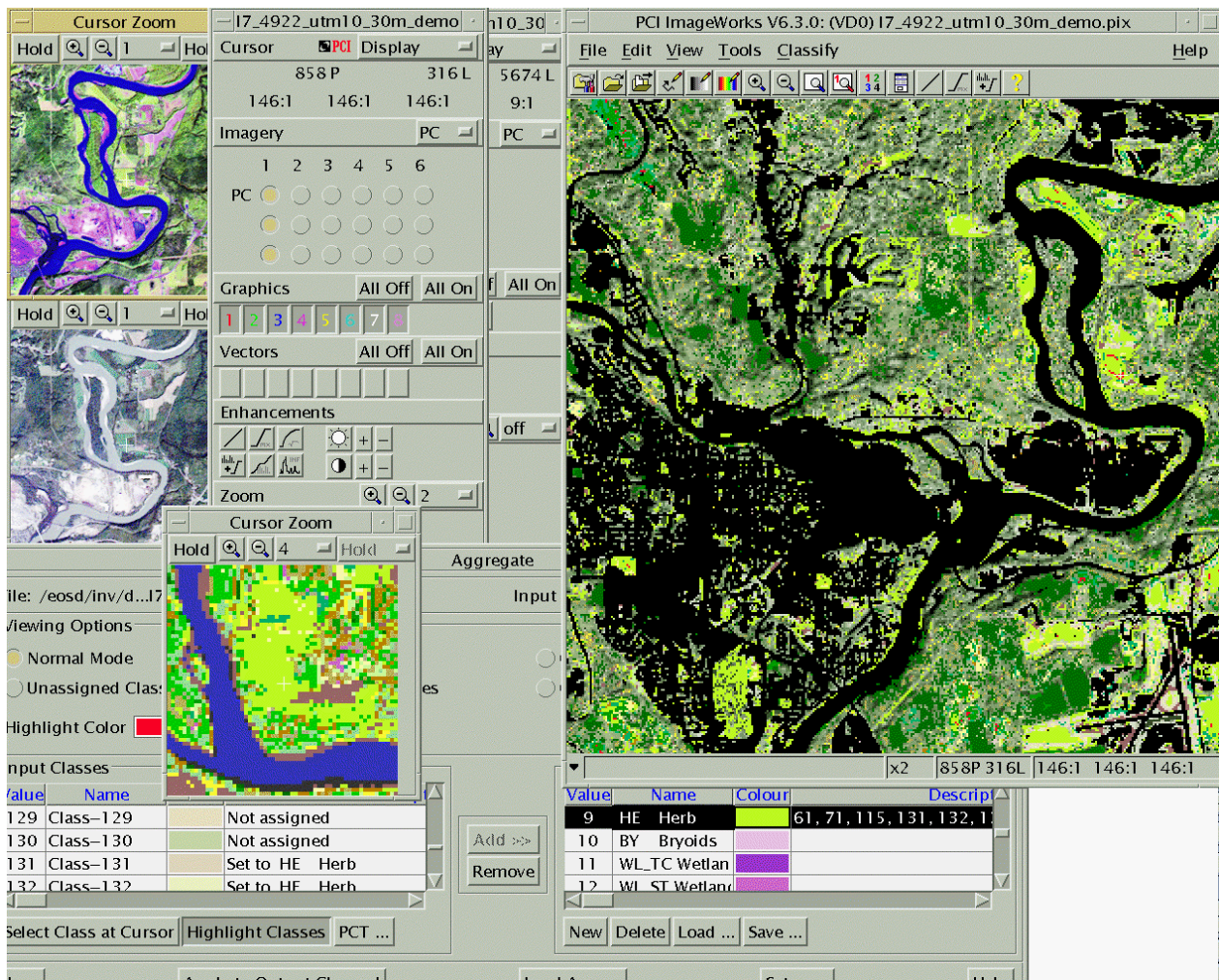
Select a class to label from the *Input Classes* list and assign it to the *Aggregate Classes* list with *Add>>*. It is usually easier to have the *Highlight Classes* button activated. Proceed through the list assigning input classes to aggregate classes until all have been assigned. It is a good idea to *Save...* the aggregate classes list frequently in case of a crash. Be sure to save to a new file. Starting with obvious classes is good for building confidence and getting a rhythm. In cases where a cluster may have captured more than one class, choose the dominant class for that cluster (comparing values for clusters at more than one location on the image).

Note that no one class can have any more than 150 input classes assigned to it or the program will fault when saving the legend file and all work will be lost. If you are approaching the 150 limit, create a new temporary class to take the excess, this can be recalculated to the appropriate class after labelling.

Once satisfied with the labelling, use the *Apply to Output Channel* button to write the data to the image channel and *Close* the *Aggregate* window.

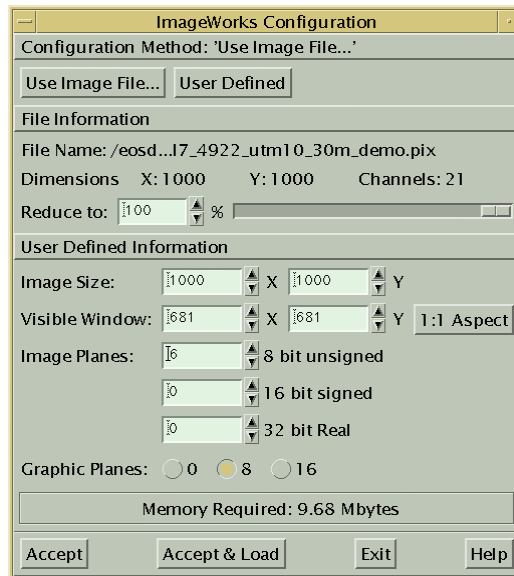


Repeat this process for each of the remaining 3 classified channels (masks). Once the master scene, or subsequent scene, has been classified it can be used in subsequent labelling as an aid. To do this, start a second **Imageworks** session. Choose *Use Image File...* and select the classified image. Load 4 image planes with 0 graphic planes at 100% resolution. Choose 5,4,3 for RGB, apply enhancement and create a zoom window. Then choose the classified image plane, load the appropriate PCT and create a zoom window. *View|Link Windows...* on both images to link them together. Minimize the previously classified window and arrange the new zoom windows and the control panel on screen so that all are visible. See following figure for an example.

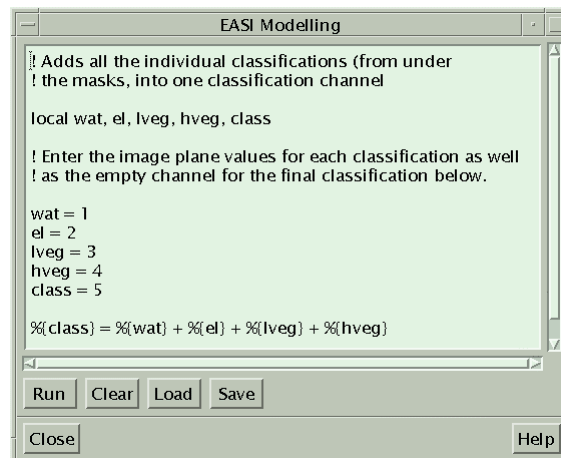


2.2.3 Combine classification results from the hierarchical spectral regions

Once all classes are labelled they are written to an output channel. There will be four channels created for every scene, one for each of the four masks used. In **Imageworks**, select the TM image at 100% resolution with 5 8U channels and 8 graphic planes and choose *Accept*.



Load the 4 classified channels and choose *Tools|Modelling...* to load the **COMB_CLS.EAS** model.



This adds the values of the first 4 channels together to a 5th channel. Since classified pixels are unique to each mask and areas without classes have values of zero, adding the four labelled classifications to a single channel can be achieved through the modelling process. *File|Save Image...* to save this newly built image plane to an empty 8-bit channel in the TM image.

2.3 Post classification Procedures

2.3.1 Slope from DEM

A slope channel is required for an optional, but recommended, post classification step to ensure lakes are not misclassified as shadow (or vice versa). To create this channel involves two steps. In the first step use **SLP** to generate an 8-bit slope channel from the 16-bit 25m DEM. The *DBIC* is the 16S slope channel, *DBOC* is an empty 8U channel, *PXSZ* is 25, 25 and *ELSZ* is 1.

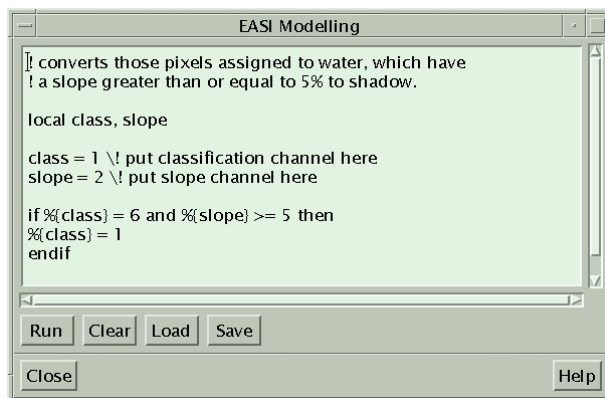
SLP	Slope of Elevation Data	V7.0 EASI/PACE	10:10 25Mar2002
FILE	- Database File Name	:/eosd/inv/trim-control/dem/nts93.pix	
DBIC	- Database Input Channel List	>	1
DBOC	- Database Output Channel List	>	2
PXSZ	- Pixel Ground-Size in Metres	>	25 25
ELSZ	- Elevation Step Size in Metres	>	1

The second step involves using **REGPRO** to transfer the slope channel to the classification image. *FILI* is the DEM file, *FILO* is the classification image, *DBIC* is the 8U slope channel and *DBOC* is an empty 8U channel. Set *RESAMPLE* to BILIN and *SAMPLING* to 1.

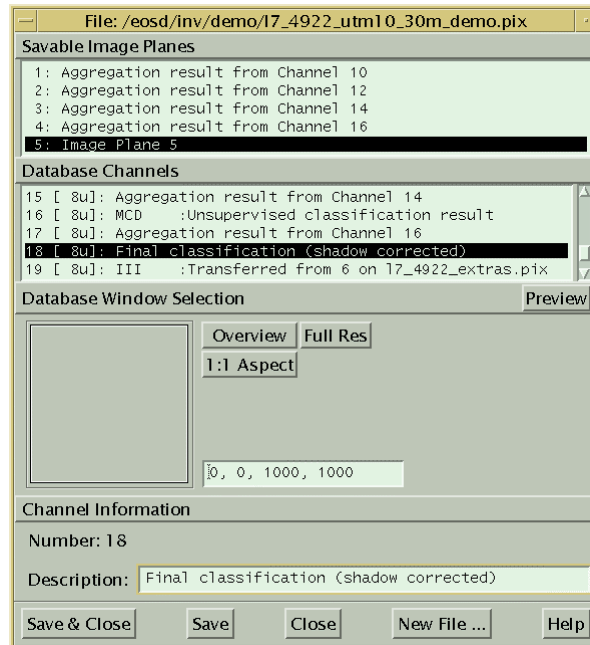
REGPRO	Image Registration and Projection	V7.0 EASI/PACE	10:13 25Mar2002
FILI	- Database Input File Name	:/eosd/inv/trim-control/dem/nts93.pix	
INGEO	- Input Georeference Segment	>	1
DBIC	- Database Input Channel List	>	2
DBGC	- Database Ground Control Segment	>	
FILO	- Database Output File Name	:17_4922_utm10_30m_demo.pix	
OUTGEO	- Output Georeference Segment	>	1
DBOC	- Database Output Channel List	>	19
RESAMPLE	- Resample Mode: NEAR/BILIN/CUBIC	:BILI	
ORDER	- Order of Polynomial Function	>	
SAMPLING	- Sampling Interval	>	1
MEMSIZE	- Working Memory Size (Mbytes)	>	512

2.3.1.1 Slope check of water / shadow

In many instances shadow is misclassified as water, particularly in mountainous terrain. To correct this, load the combined classification channel and the slope channel. *Tools|Modelling...* to load the **H20-SHDW.EAS** script.



This will convert any pixels assigned to the water class with a slope greater than 5% to Shadow. Be sure to save this image plane and note that the shadow correction has been performed.



2.3.2 Incorporation of other ancillary data sources

In conjunction with DEM data, other ancillary data sources such as the National Topographic Database (NTDB) or Baseline Thematic Mapping (BTM) data may also be used for revising or enhancing the classification. While elevation and slope data from the DEM are particularly useful, water body and wetland data from NTDB and agriculture data from BTM may be assessed to uncover labeling biases or issues with the final classification (to enable revision, or in some circumstances, conversion of the label using logical rules). Data from these sources must be rasterized and added to channels within the scene to be used in a modeling script. **TOTAL_MOD.EAS** is an example script that can be altered to suit the need of each individual classified scene.

3.0 Quality Assurance

A means for characterizing the accuracy of the EOSD land cover classification product is required. Procedures for the determination and characterization of EOSD classification accuracy at a national scale are currently under development (contact M. Wulder). This national accuracy assessment protocol will involve the sampling of a subset of EOSD products across the country, and the application of robust statistical methods to determine overall product accuracy.

Specific regions may wish to assess the accuracy of their classifications as the product is in development. To this end, we provide the following guidelines that should be followed when conducting an accuracy assessment of an EOSD product:

1. The validation set should consist of independent, purpose acquired field data (collected by ground or airborne means) that relates directly to the EOSD classes, and must not include data that was used to calibrate the classification during production.
2. If the validation data is culled from forest inventory data or inferred from airphotos, care must be taken to ensure that the locating of the points is appropriate and that the classification used in the reference data is representative (and that the categories represent a reasonable cross-comparison).
3. At this time, we recommend following the methods outlined by Czaplewski (2003). This involves the development of a contingency table relating known (field data) cover information to the classes assigned through the clustering and labelling. The approach requires collection of field data to represent the cross section of classes in the EOSD legend. These field measured classes may then be compared to the classes found at the same location in the image classification. In so doing, a comparison, on a class by class basis, to examine the success of the classification is enabled. Confidence intervals are also reported to provide a more robust indication of true accuracy.
4. The recommended target classification accuracy is 85% overall and across all classes.

4.0 Creation of Final EOSD Product

The final specifications for the EOSD product are laid out in the [***EOSD Land Cover Classification Report, Version 4.2***](#) The final EOSD product is delivered as a NTS 1:250,000 mapsheet and classified products overlapping with any given tile must be mosaicked together prior to clipping the imagery to the appropriate NTS mapsheet. An automated process has been developed for mosaicking and tiling the classified EOSD images to create the final EOSD product. The following document details this automated process: [***Documentation for automatically generating EOSD_lc final product***](#)

References

- Avery, T.E. and Berlin, G.L. 1992. *Fundamentals of Remote Sensing and Airphoto Interpretation*. 5th ed. Toronto: Maxwell Macmillan; 472 p.
- Cihlar, J. 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, 21(6-7): 1093-1114.
- Cihlar, J. and Beaubien, J. 1998. *Land cover of Canada Version 1.1. Special Publication, NBIOME Project*. Produced by the Canada Centre for Remote Sensing and the Canadian Forest Service, Natural Resources Canada. Ottawa, Ontario, Canada.
- Czaplewski, R.L. 2003. Statistical design and methodological considerations for the accuracy assessment of maps of forest condition. *Remote Sensing of Forest Environments: Concepts and Case Studies*, edited by M.A. Wulder and S.E. Franklin, Boston: Kluwer Academic Publishers: 115-141.
- Franklin, S.E. and Wulder, M.A. 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, 26(2): 173-205.
- Fuller, R., Groom, G. and Jones, A. 1994. The land cover map of Great Britain: An automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, 60(5): 553-562.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., and Sohlberg, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21(6&7): 1331-1364.
- Jensen, R. 1996. *Introductory Digital Image Processing. A Remote Sensing Perspective*. Second Edition. New Jersey: Prentice Hall.
- Lillesand, T. and Kiefer, R. 1987. *Remote sensing and image interpretation*. Toronto: John Wiley & Sons; 721 p.
- Loveland, T. R. and Belward, A.S. 1997. The IGBP-DIS global 1-km land cover data set, DISCover: first results. *International Journal of Remote Sensing*, 18(15): 3289-3295.
- Loveland, T.R., Merchant, J.W., Ohlen, D.O. and Brown, J.F. 1991. Development of a land-cover characteristics database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, 57(11): 1453-1463.
- Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W. 2000. Development of a global land cover characteristic database

- and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*. 21(6&7): 1303-1330.
- Peddle, D.R., Teillet, P.M and Wulder, M.A. 2003. Radiometric Image Processing, *Remote Sensing of Forest Environments: Concepts and Case Studies*, edited by M.A. Wulder and S.E. Franklin, Boston: Kluwer Academic Publishers: 181-208.
- Stone, T., Schlesinger, P., Houghton, T., Woodwell, G. 1994. A map of the vegetation of South America based on satellite imagery. *Photogrammetric Engineering and Remote Sensing*, 60(5): 541-551.
- Wulder, M., Gillis, M., Luther, J., and Dyk, A. 2001. A Guide to the Estimation of Canada's National Forest Inventory Attributes from Landsat TM Data, Version 2. Natural Resources Canada, Canadian Forest Service, Working paper. Victoria, British Columbia, Canada: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre; 84 p.
- Wulder, M., E. Loubier and D. Richardson. 2002. A Landsat-7 ETM+ orthoimage coverage of Canada. *Canadian Journal of Remote Sensing*, 28(5): 667-671.
- Wulder, M. and T. Nelson. 2003. *EOSD Legend: Characteristics, Suitability, and Compatibility, Version 2*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada.
- Wulder, M.A., Franklin, S.E., White, J.C., Cranny, M.M. and Dechka, J.A. 2004. Inclusion of topographic variables in an unsupervised classification of satellite imagery. *Canadian Journal of Remote Sensing*. 30(2): 137-149.