

SP2Learn Manual

Version 1.0 Release

SP2Learn was created as a joint collaboration between the Illinois State Water Survey (ISWS) and the National Center for Supercomputing Applications (NCSA) at UIUC. Funding support was provided by NCSA Faculty Fellow Program, and leveraged by our past funding provided by the National Aeronautics and Space Administration (NASA), and National Archive and Record Administration (NARA). The project was led by Yu-Feng-Lin (ISWS) and Peter Bajcsy (NCSA UIUC). The main programming contribution to SP2Learn came from Chulyun Kim. This document represents a preliminary user guide for an on-going research and development project and hence the code is viewed as a research effort rather than as a commercial package.

Revision History

Revision	Date	Author	Notes
1.0	08/21/2007	PB, CK	Initial version of the document
1.1	09/14/2007	CK	Update screen captures
1.1	10/18/2007	PB, CK	Update text

Table Of Contents

Introduction.....	4
Chapter 1 Installation Instructions	6
1.1 Hardware requirements.....	6
1.2 Software requirements	6
1.3 Installation steps.....	6
1.4 Notes about Execution	6
1.5 Notes about Data.....	7
Chapter 2 Description of Test Data Sets	8
2.1 Raster Files.....	8
2.1.1 HDF EOS	8
2.1.2 TIFF	8
2.1.3 SRTM.....	8
2.1.4 Other supported file formats	8
2.2 Vector Files.....	9
2.2.1 ESRI Shapefiles	9
2.2.2 Other supported file formats	9
Chapter 3 Description of Processing Steps	10
3.1 Load Raster	10
3.1.1 Adding and Removing files	10
3.1.2 Mosaic.....	10
3.1.3 Functionality	11
3.1.4 Utility	13
3.2 Registration.....	13
3.2.1 Parameters.....	14
3.3 MaskCreate	15
3.4 Attribute Selection	23
3.4.1 Options.....	23
3.5 Rules	24
3.5.1 Columns	25
3.5.2 Options.....	26
3.6 Apply Rules	27
Chapter 4 Additional Notes	29
4.1 Bug Reports and Bug Fixes	29
4.2 Known Problems.....	30
4.2.1 Out of Memory Errors	30
4.2.2 Other Problems	30

4.2.3	Additional Notes	30
4.3	Acknowledgments.....	31
4.4	License Agreement	31
4.4.1	Illinois Open Source License	31

Introduction

The motivation for developing SP2Learn comes from studying ground water models. The research and development in this area addresses the scientific questions about accurate modeling of ground water recharge and discharge (R/D) rates using physics-based and data-driven modeling techniques.

SP2Learn can be viewed as an encapsulated workflow for (1) loading multiple raster files (images), (2) integrating and mosaicking all raster data sets to form a stack with consistent spatial resolution as well as geographic projection, (3) loading other files (boundaries, points or images) to create a mask for pixel selection purposes, (4) integrating the existing stack of raster images with other masking information, (5) selecting boundaries or image regions of interest and extracting variables from the stack of images, (6) performing data-driven modeling of selected input and output variables, (7) analyzing data-driven model to assign a relevance coefficient to input variables, and (8) mapping the data-driven model at a pixel level to spatial domain. All aforementioned steps are supported by visualizations (colorful, gray-scale or pseudo-color) of input, intermediate, and output data sets, as well as the data models. An overview of the functionality is provided in Figure 1a.

In order to bring together so much functionality, the architecture of SP2Learn leverages several technologies as illustrated in Figure 1b. The majority of the SP2Learn code is based on Image to Learn (Im2Learn) developed at NCSA with additional calls to (a) HDF5 library developed by NCSA and (b) MODIS Reprojection Tool (MRT) to perform geographic re-projection developed by NASA to perform. These software components are described in more detail in Chapter 4.

The structure of this document follows the processing sequence as described above. The software installation steps are provided in Chapter 1. The software release comes with test data sets that were used in our research and are briefly described in Chapter 2. The processing steps are outlined in Chapter 3 with pictorial examples. Additional notes regarding bug reports, known problems, code maintenance, code dissemination, and licensing are presented in Chapter 4.

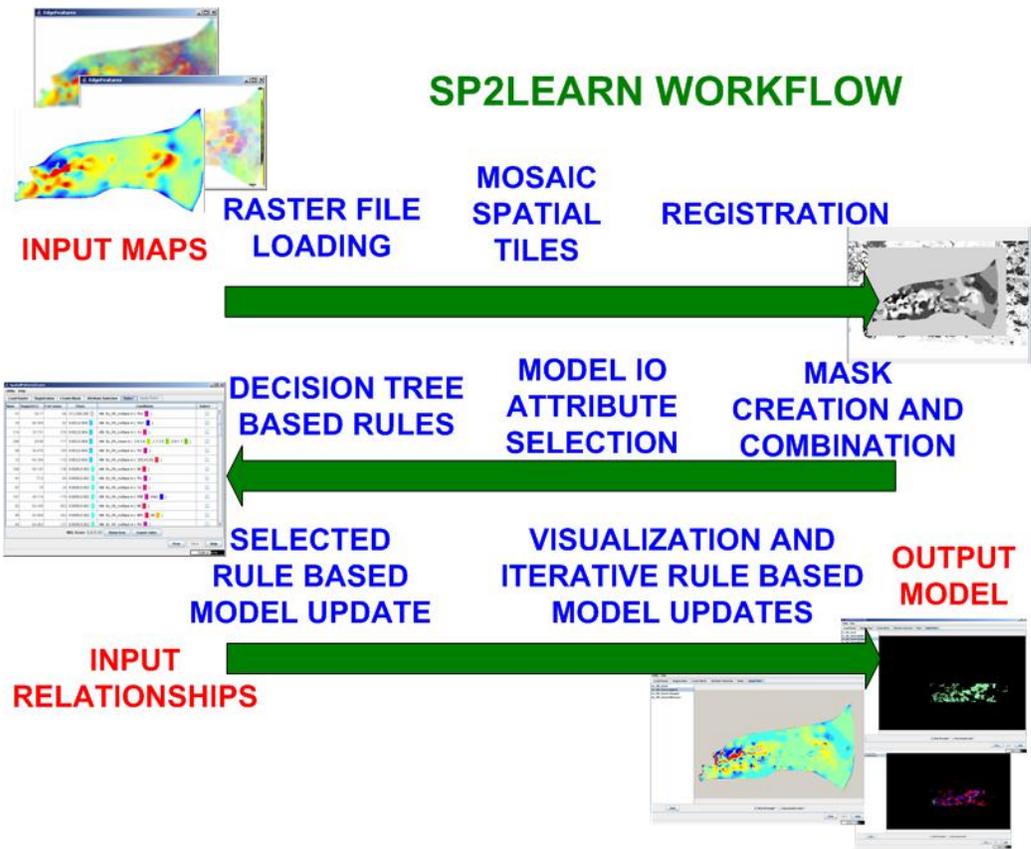


Figure 1a: An overview of SP2Learn functionality.

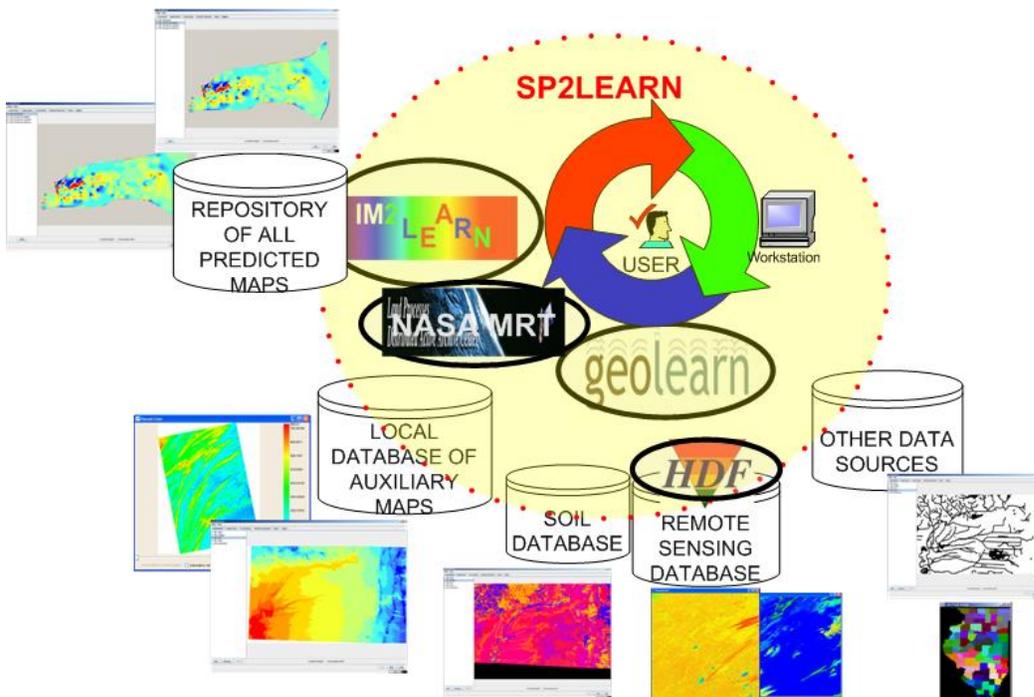


Figure 1b: SP2Learn technology overview

Chapter 1 Installation Instructions

1.1 Hardware requirements

For processing large raster data sets, at least 512MB of RAM is recommended. Some processing steps might require significant computational resources. However, for the test data sets, a regular desktop computer should be able to handle the processing. The execution duration depends on the hardware specification and the algorithmic computational needs.

1.2 Software requirements

- An Operating System capable of running Java
- Java 1.5

1.3 Installation steps

To install the SP2Learn application, go to the Image Spatial Data Analysis (ISDA) NCSA website <http://isda.ncsa.uiuc.edu/download> and register to download. Your password will be e-mailed to you. After entering your password, select SP2Learn from the list of available tools.

Once you have downloaded SP2Learn, double click on the zip file and extract the files to an appropriate location. Navigate to the directory where you extracted the files, and double click on SP2Learn.bat.

1.4 Notes about Execution

The code is design to handle large size raster data. The word large refers to data sizes that would not fit into the allocated heap memory space during the SP2Learn execution. We use the terms ‘out-core-representation’ and ‘out-core-processing’ to indicate that the data access is enabled via partitioning data into chunks (e.g., image tiles), saving temporary data chunks on a disk and then loading a processing chunks of data sequentially. This approach is different from the ‘in-core’ representation and processing when the entire data set is loaded into RAM.

In the current implementation, the raster files that are larger than 2048x2048 (rows x cols) are represented as ‘out-of-core’. If the program fails then it might occur that the temporary chunks of data are not removed from the temporary directory. A user might check his/her default directory (e.g., C:\Documents and Settings\‘user name’\Local Settings\Temp) for .img2 files and remove them manually.

The SP2Learn code is written in java which might not release memory immediately. In order to force garbage collection of unused RAM memory, one can right mouse click on the memory status bar in the lower right corner of the SP2Learn application. The memory status bar shows how much RAM is being currently occupied by SP2Learn.

1.5 Notes about Data

The distributed zip file contains a directory called data. One can run the main step of SP2Learn with the files in the data directory. The directory contains the following files:

- (1) DEM_Illinois – sub-sampled USGS digital elevation map of Illinois in geographic projection
- (2) fpar_h10v05_2003_185.tif - fraction of Photosynthetic active radiation extracted from a MODIS HDF EOS file. The file was re-projected from sinusoidal projection to albers equal area conic projection
- (3) IL_County – county boundaries in the state of Illinois stored in ESRI Shapefile format
- (4) testGeoPoints.csv – tabular data with three columns (latitude, longitude and temperature over the state of Illinois) in a comma separated value file format.

For testing purposes, one can load the data sets (1) and (2) in the first step (load raster step) and then select subsets of pixels for data-driven modeling by using data sets (3) or (4) in the third step (Create Mask step).

Chapter 2 Description of Test Data Sets

2.1 Raster Files

2.1.1 HDF EOS

Remote sensing datasets acquired by the MODIS satellite and distributed by NASA are stored in HDF EOS format. HDF binary files can be directly loaded using Im2Learn. We are using Land Surface Temperature (LST), Emissivity, Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), Fraction of Photosynthetically Active Radiation (FPAR), Snow Cover, Albedo, Land Cover variables. All above variables are processed using SP2Learn. Test datasets contain data for only two tiles (h10v05 and h11v05) and only for the year 2003 starting on the 177th Julian day. The HDF files that can be processed by SP2Learn have to contain georeferencing information.

HDF-EOS data can be downloaded from the URL:

<http://edcimswww.cr.usgs.gov/pub/imswelcome/>

2.1.2 TIFF

SP2Learn software has also the capability to load raster files stored in a tiff format. All preprocessed datasets are saved on disk in a tiff format. They can be loaded by Im2Learn loaders if the pre-processing step should be skipped and the data are ready for integration and data modeling as explained in this document. Arc Grid, SRTM, and HDF EOS data can be saved as tiff files after processing. Several processed datasets are available with the code release (HDF preprocessed data, climate data for Appalachian region, and soil data).

2.1.3 SRTM

Elevation datasets from SRTM are used to extract different features such as Slope, Aspect, Flow Direction and Flow Accumulation. SRTM tiles can be mosaicked together using Im2Learn. Each of the extracted features is saved in tiff format (available with the test datasets).

SRTM data can be downloaded from the URL:

<http://edc.usgs.gov/products/elevation/srtmtdted.html>

2.1.4 Other supported file formats

In order to perform raster file integration, one needs to know the georeferencing information. In the current software, the georeferencing information can be extracted only from the file formats supported by Im2Learn, such as USGS DEM (Float), ESRI IMG format, HDF EOS, GeoTiff, and SRTM. Other image formats, such as jpg, gif, tif (Ver. 6.0 with no compression), pgm, ppm, bmp, png, pnm, img, are supported by Im2Learn but cannot be processed.

2.2 Vector Files

2.2.1 ESRI Shapefiles

One of the test files is an eco-region shapefile at level III. This vector file provides a mask for processing subsets of pixels defined by a user-driven selection of eco-region of interest.

Eco-region shapefiles can be downloaded from the URL:

http://www.epa.gov/wed/pages/ecoregions/level_iii.htm

From the web site one obtains a coverage file that was processed by us. The processing involves the change of geographic projection to Albers Equal Area and saving out the file in the ESRI shapefile format.

2.2.2 Other supported file formats

There is a support for STATSGO and SSURGO DLG vector file format in Im2Learn software. However, these vector file formats are not currently enabled.

Chapter 3 Description of Processing Steps

3.1 Load Raster

3.1.1 Adding and Removing files

The inputs of the analyses are raster images with data and georeference information. Typical raster images will be stored in HDF EOS or GeoTIFF format. This step allows you to select multiple raster images that will be used in later steps. To add raster images to the list press the Add button. This will bring up a file selection box. It is possible to add files individually, or several at a time. To select multiple images from the same folder, control-click on each file you wish to add. If you want to add an entire list of files, you can shift-click on the first and last items of the list. After you have selected the appropriate file(s), click on Open. An example of adding raster images is shown in Figure 2. In the left list box, the names of added images are listed and the preview image is shown at the right panel. To remove raster images from the list, select the raster image(s) and press the Remove button. MODIS HDF files without QA/QC masks can be added in the Load Raster panel. If you add a MODIS HDF file, you need to select the appropriate data field from the MODIS Grid.

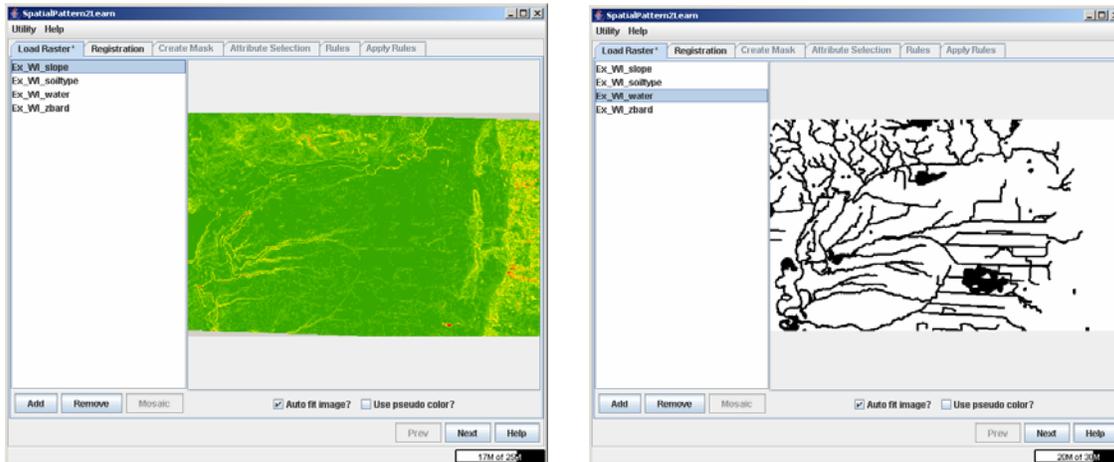


Figure 2: Added raster images.

3.1.2 Mosaic

The purpose of mosaicking is to stitch together spatial tiles (e.g. SRTM tiles) to form a mosaic. To combine multiple raster images into a single raster image, select the appropriate images using control-click, then click Mosaic. This will produce a new single raster image that replaces the raster images that were used to create it. In Figure 3, three raster images are mosaicked as a single image. The first three images are vertically continuous and mosaicked seamlessly in the last image.

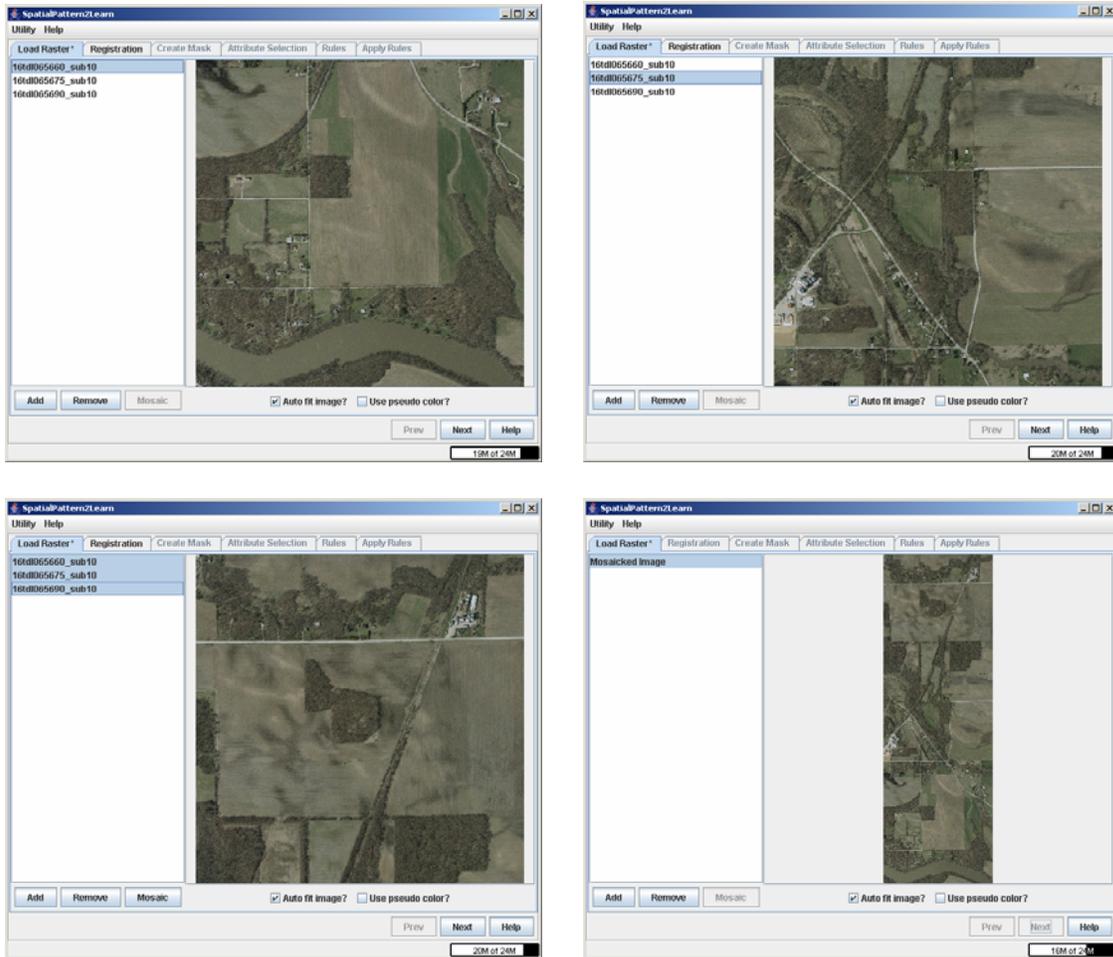


Figure 3: Mosaicked raster images.

3.1.3 Functionality

Selecting a raster image will show a preview of the raster image on the right. Right clicking on the image preview will display information that the raster contains as shown in Figure 4. The raster image will be scaled to fit the window by default. The window itself can be resized, consequently causing the image to refit. If **auto fit image** is unchecked, the raster image will be shown at 100%. The scroll bars can be used to navigate the image when it is at 100%. There is a zoom menu when right clicking on an image preview. This option changes digital resolution of an image. The option fit will automatically rescale an image to fit the image panel size. The option custom lets a user define the scaling factor either as a multiplicative factor or as a percentage (by appending % to the number) of the original image size. For example using the number 2, or 200% will result in an image that is twice the size of the original. Figure 5 shows the pop-up menu when right-clicking.

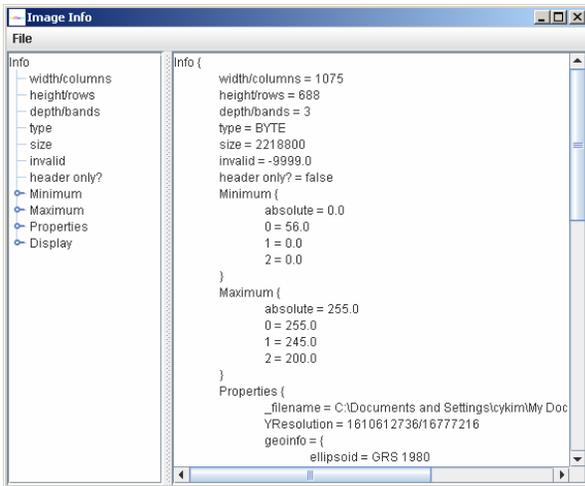


Figure 4: Image information dialog box.

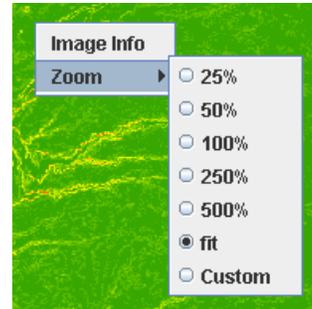


Figure 5: Pop-up menu when right-clicking.

In all processing steps, there is a possibility to visualize the data in the Image Panel on the right side of the processing window. If pseudo-color is used, the absolute values are color coded from minimum (blue) to medium (green) and maximum (red). If the input image contains multiple bands then the slider bar on the bottom can be used to pseudo-color each band separately. The check box “**Automatically set min and max value?**” should be unchecked when the assignment of min and maximum values need to be controlled by a user by entering the values in the edit boxes in the lower right corner. The result of such assignment might provide better understanding of data. The check box labeled as **Use min/max of whole image** is relevant when a visualized image contains multiple bands and the min and max values should be set per band or for the whole image. In Figure 6, a typical pseudo colored image is given. The image has three bands and the captured screen shows the first band information. It uses min/max of whole image and automatically sets min/max value.

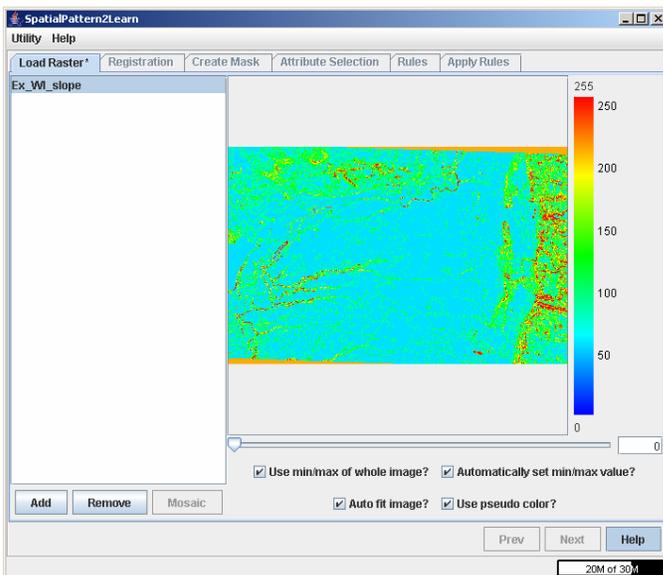


Figure 6: Pseudo colored image.

After you have loaded the appropriate raster images, you can proceed to the next step by clicking on the Integration tab or by clicking **Next**. If you have loaded at most one image, you cannot go next.

3.1.4 Utility

You can assign a label to each color in a raster image by Label Editor. If you assign labels, it makes your future results more readable. You can execute Label Editor at Utility menu. When you select 'Utility->Label Editor' as shown in Figure 7a, a file-open dialog box occurs and you can select an image where you want to assign labels. Then, Label Editor shows all distinct color values and their labels. Figure 7b and 7c show these steps. You can start to edit labels by double-clicking a label cell. If you have another image which already has the label information, you can import the label information by pushing the Import button. In order to keep your new label information, you should save by clicking the Save button.

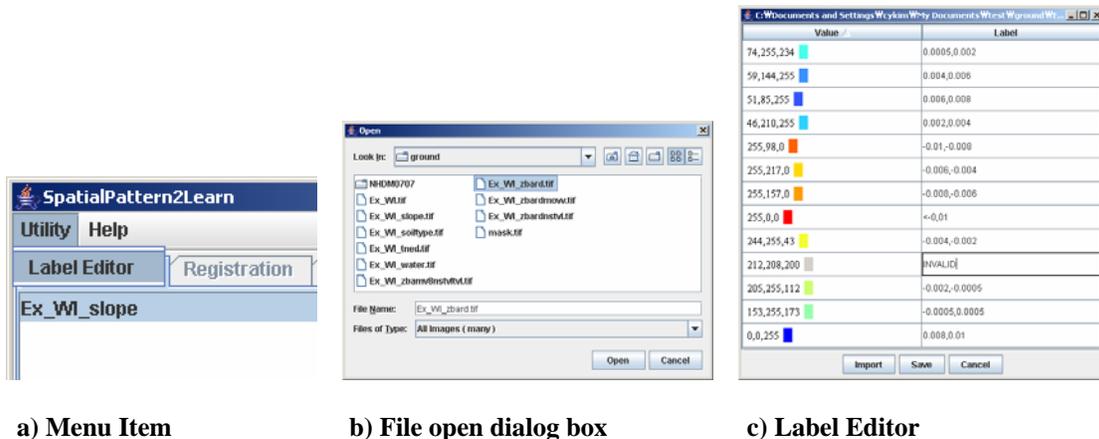


Figure 7: Steps of using Label Editor.

3.2 Registration

This step registers all raster data sets to form a stack with consistent spatial resolution as well as geographic projection. That is, registration takes multiple raster images and forms a single image stack of variables.

The registration step has the same preview functionality as Load Raster step. For Registration and Create Mask panels, the small triangles can be clicked on to expand a specific frame in the window. The bars separating frames can be dragged to resize the frames.

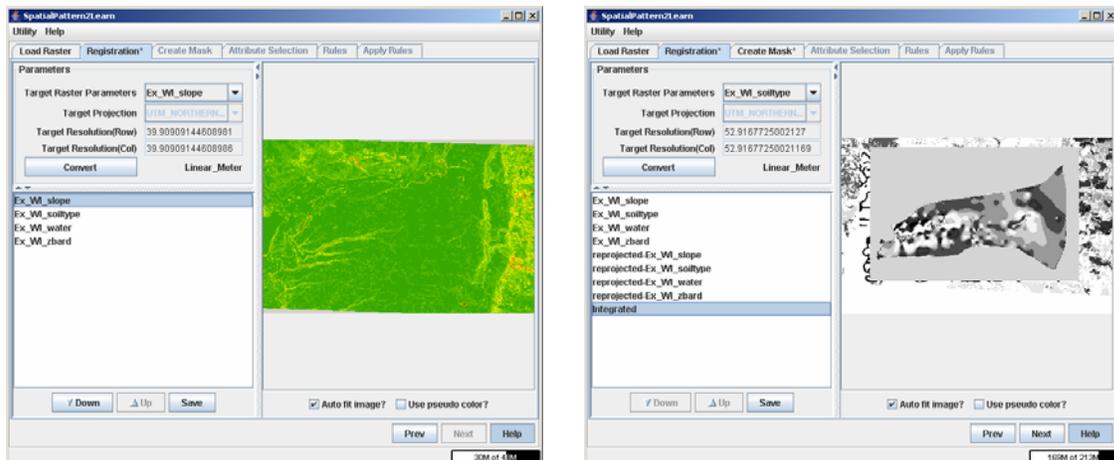
3.2.1 Parameters

There are two major parameters, target projection and resolution. These parameters are selected as those of an image among the loaded images. Target Raster Parameters is the list of the raster images that were loaded in the previous step. Selecting a raster of the list will accordingly change the Target Projection and Target Resolution boxes to the projection and resolution of the selected raster. Projection and Resolution are not allowed to be changed individually. When you have selected the appropriate projection and resolution, click Convert. This will first reproject each image to the selected projection and resolution. Then, SP2Learn will register the reprojected images into a single raster image named **Integrated**. Initial status before converting and result status after converting are illustrated in Figure 8.

The difference between Mosaic in the Load Raster step and Registration in this step is as followings. If two images are mosaicked, two images are converted to a single image. However, if they are registered, the integrated image may look like a single image, but the reprojected images are still two separated images in the same coordinate system. In addition, in order to mosaic images, the images should have the same number of bands, but images which have the different number of bands can be registered.

The first image in the ListBox is the bottom image in the integrated image and the last image is the top image. In order to change the order, select an image in the ListBox and click UP or DOWN button. Then, the integrated image is immediately updated. Note that you can reorder only the original images but not the reprojected images. The order of the reprojected images will follow that of the original images.

Once you have completed registration, you can proceed to the Create Mask step by clicking on its tab or by clicking Next.



a) Initial status before converting

b) Result status after converting

Figure 8: Steps of registration

3.3 MaskCreate

As implied by the name, a mask hides part of an image so that only certain pixels will be used for Attribute Selection and Modeling. There are several reasons to use the Create Mask step. One is that Modeling and Visualization of large size data sets will take too long with an unmasked image on a desktop computer. Another is that you may wish to include/exclude specific data for the modeling and visualization steps.

This step will allow you to create masks, do operations on them, and apply one of them on the rasters loaded in the previous step and at the same time get prepared to provide a table for modeling. For creating a mask, the integrated image is used as a baseline.

There are three sections in this window: panel "Mask Creation"(Figure 9a) and panel "Mask Operation"(Figure 9b) on the left and main panel(Figure 9c) on the right.

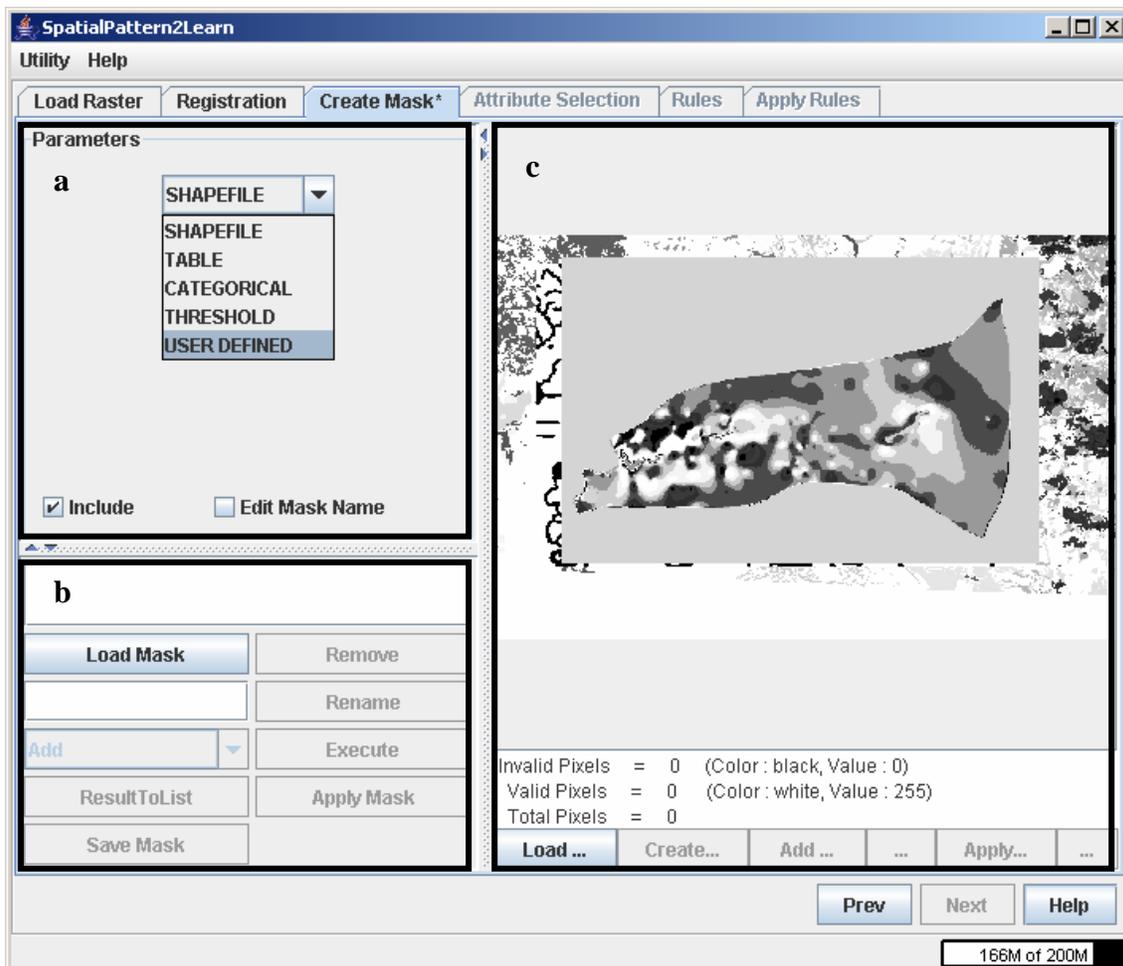


Figure 9: Create mask step

1. "Mask Creation" panel consists of four parts:

1.1 A combo box on the top: providing to you up to five methods to create a mask: "SHAPEFILE", "TABLE", "CATEGORICAL", "THRESHOLD", "USER DEFINED"(please see Note 1);

1.2 A parameter portion right below the combo box: allowing you to specify any parameters required for mask creation;

- Check box "include": this will take effect every time you create a mask. When it is selected, pixels of the mask in the chosen area will be white; otherwise, they will be black. Black area will be masked.

- Check box "Edit Mask Name": this will take effect every time you click "Add To List" button. When it is selected, you are allowed to set a name to the mask of interest by typing into the text field in "Mask Operation" panel.

2. "Mask Operation" panel contains a list and functional buttons. The list, together with a text area on the top of this panel will allow you to store and display masks you are interested in. An image frame will be popped up displaying the selected mask, if you click a mask in the list (for details about functions of the other components in "MaskOperation" panel, please see Note 2).

3. The main panel consists three parts:

3.1 An image panel on the top to display the integrated raster or created masks;

3.2 A text area to show number of valid and invalid pixels of a mask, and in the case of creating mask by "Threshold", to show threshold values.

3.3 Buttons (for details about functions of each button in this part, please see Note 4).

The first thing to do in this step is to select a method from the five methods available in the combo box in "Mask Creation" panel. By selecting one of the five options in the combo box, two parts of the screen will be updated: 1, a parameter panel will appear right below the combo box if necessary (note: methods "SHAPEFILE" and "CATEGORICAL" do not require any parameters, thus there is no parameter panels for them); 2, buttons at bottom of the main panel will be updated, according to the method you have selected.

A typical working flow in this step can be described as: create masks using one or more of the five mask creating methods; add the mask of interest into list; if necessary, generate new masks by carrying out operation on existing masks; apply the mask of interest to the original raster, and at the same time get ready to prepare a table containing values of all valid pixels (pixels with white color in the applied mask), which will serve as input data for the following steps ("Attribute Selection", "Rules" and "Apply Rules").

Note 1:

Following is five methods to create a mask:

1, Method "SHAPEFILE": to create a mask from the original raster, based on a shape file. After loading a shape file with proper geo-referencing information for the raster, a new image with boundaries will be displayed in image panel, on which you can select any boundaries of interest by mouse clicking. A new mask can then be generated based on your selections. You can return to the boundary image using button "Reset", if you are not satisfied with resulting mask.

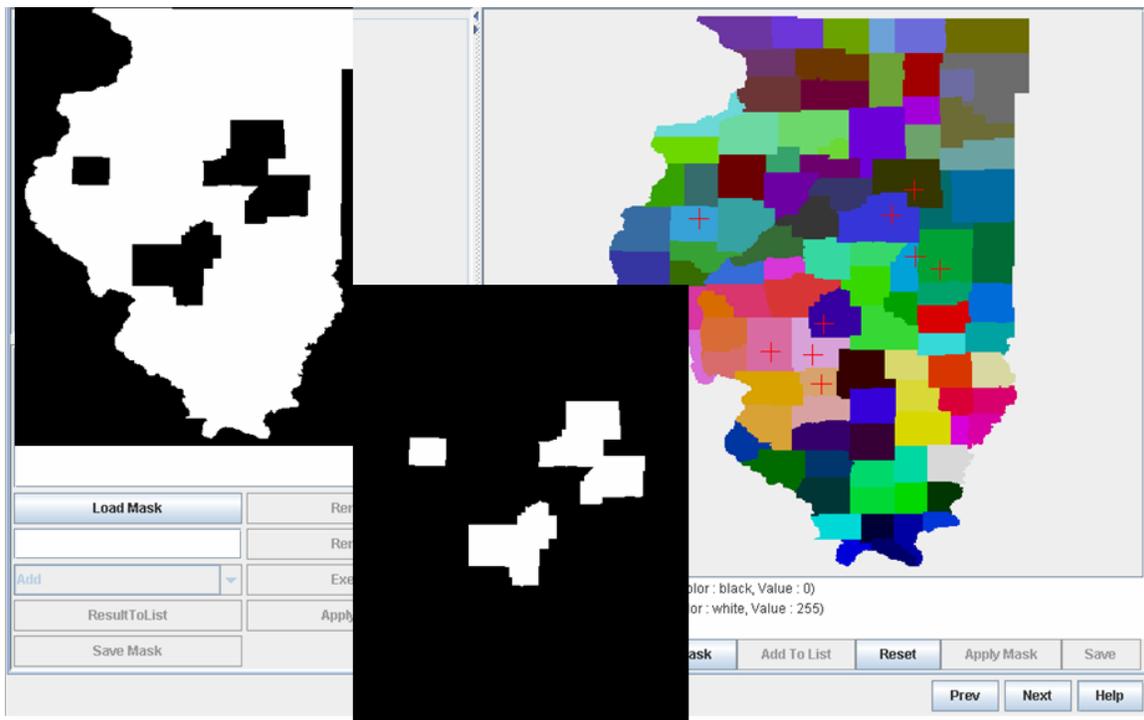


Figure 10: Method "SHAPEFILE"

2, Method "TABLE": In this method, you should first load a table as either a CSV file or MS.Excel file, from which points with geo-referencing information can be extracted. You can view the table in a new frame using button 'Show Table'. Before creating a mask, you are responsible to specify the following the parameters: number of columns that store latitude and longitude information; radius, by which you specify how near a pixel should be away from a certain points, in order to be selected; and in case of loading an MS.Excel file, the number of worksheet you want to use;

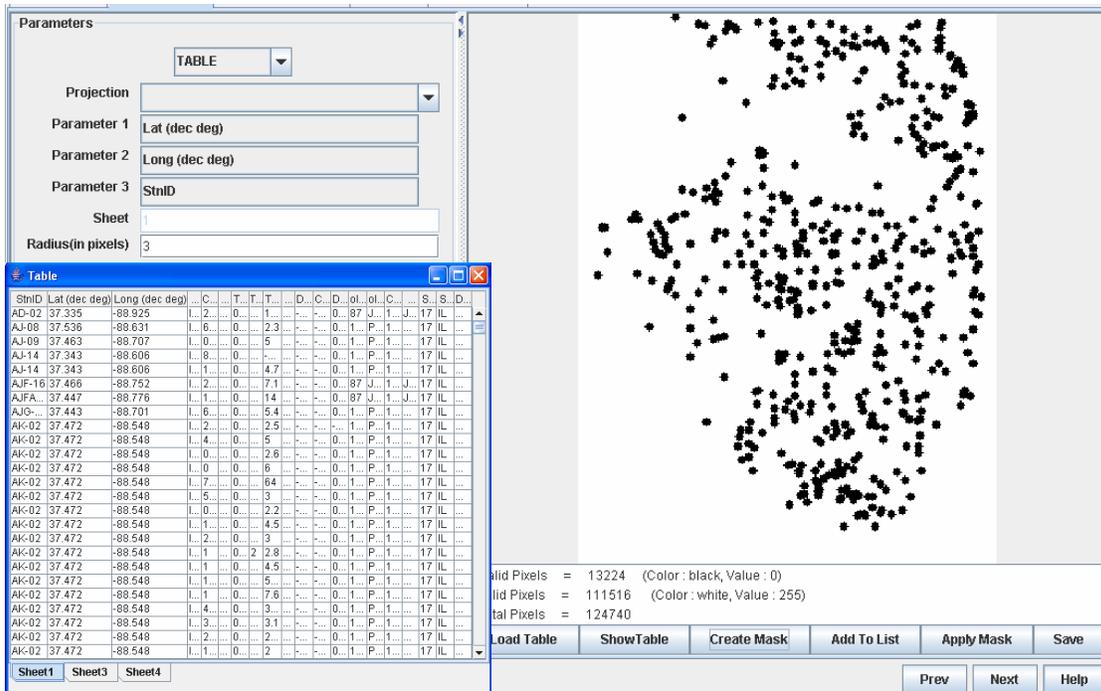


Figure 11: Method “TABLE”

3. Method "CATEGORICAL": to create a mask from the original raster, based on a raster with categorical labels. With such a raster loaded, you will be able to select any label of interest by mouse clicking in the image panel. In the resulting mask, pixels with the same color as any of the pixels you selected will become white and others will become black, if checkbox "include" is selected. Again, you can return to categorical raster by clicking button "Reset";

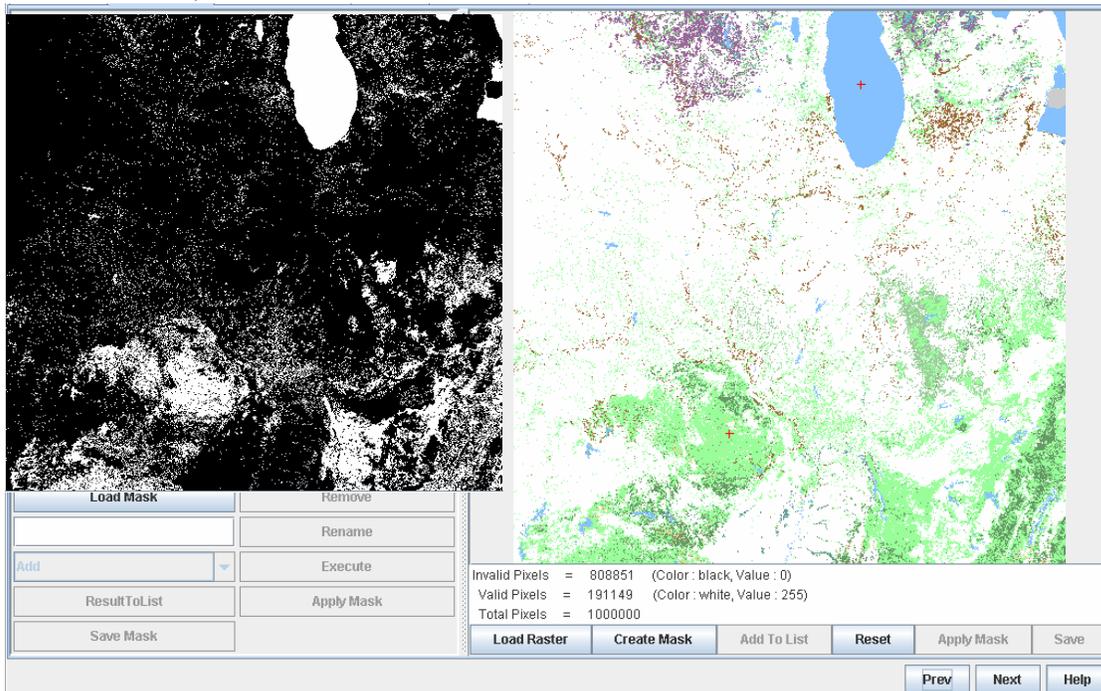


Figure 22: Method “CATEGORICAL”

4, Method "THRESHOLD": to create a mask by applying one or two threshold values to a raster, which can either be the original raster loaded in previous step or the new raster you loaded with button "Load Threshold" at the bottom of main panel. There are three modes to set threshold values: "Scalar", "Box (vector)" and "Plane (2 vectors)", depending how many bands the raster has. For a raster with a single band, only "Scalar" mode is applicable. Otherwise, all three modes are active. For each mode, there are four ways to threshold image, as is showed by a combo box on the right of parameter section of "Mask Creation" panel. A new mask will be created every time you move the either of the two slider bars or enter a number and press "Enter" in either of the two text fields.

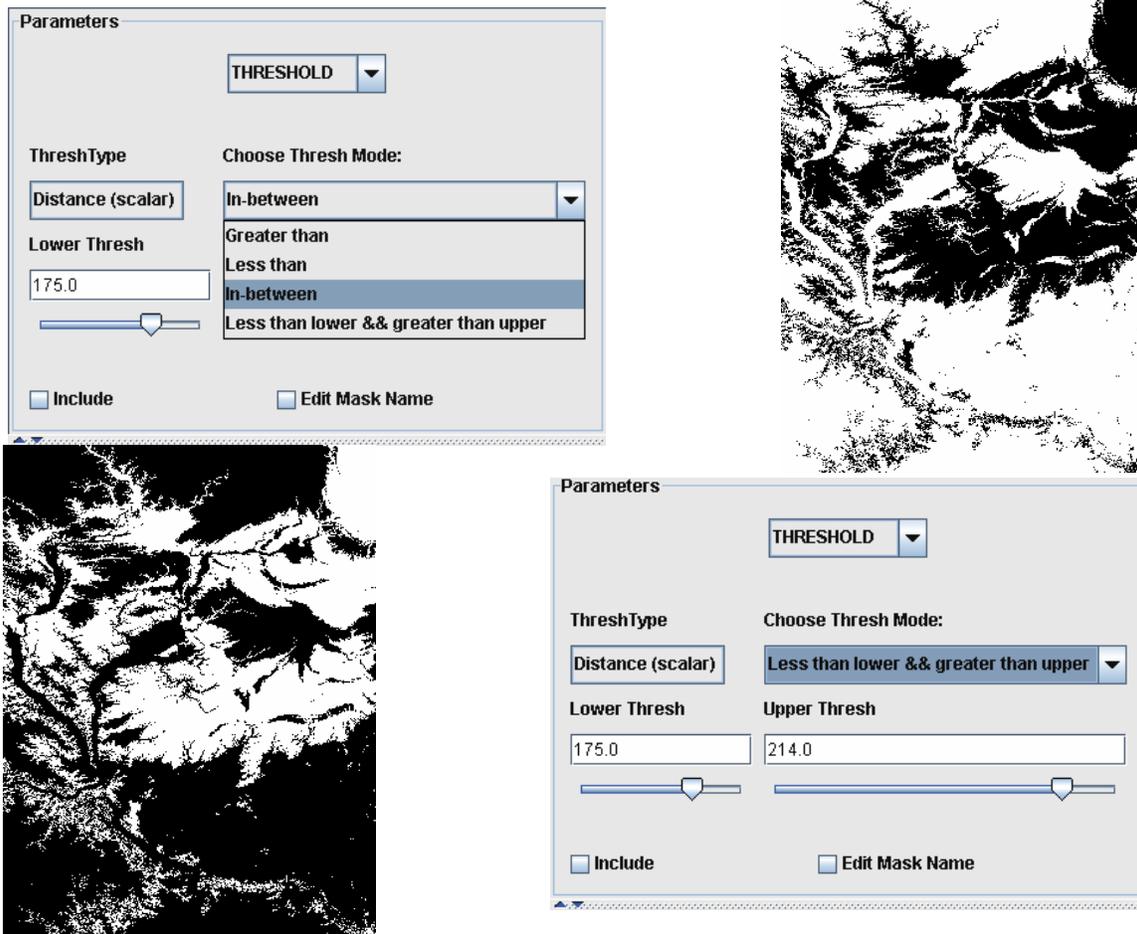


Figure 33: Method "THRESHOLD"

5, Method "USER DEFINED": In this method, you will be able to create a mask by drawing rectangles and ovals at any desirable location on the original raster by following three steps. First, specify a rectangle on the raster by mouse dragging; second, if the rectangle is on a desirable location, click button "Paint" to generate a modified raster with a white area of either rectangular or oval at the specified location, depending on which shape you choose; finally, click "Create Mask" to create a new mask with all pixels in white rectangles and ovals displayed as white, if check box "include" is selected.

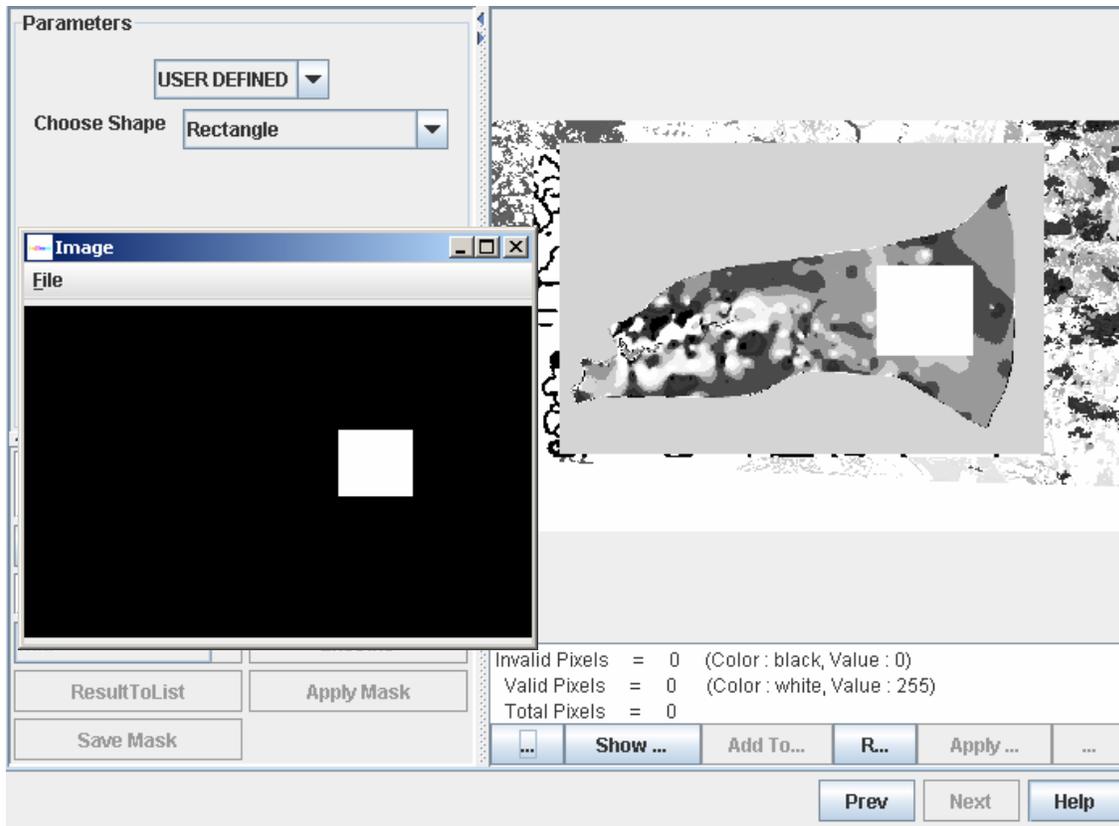


Figure 44: Method “USER DEFINED”

Note 2:

Following is functions of the other components in "Mask Operation" panel:

Button "Load Mask": to load masks into list from hard drive;

Button "Remove": to remove selected masks from the list;

Button "Rename": to replace the name of a selected mask with the string in a text field on left of it (note: only one mask can be renamed at once);

Button "Execute": to execute selected operation on selected masks;

Button "ResultToList": to add into list the mask produced by executing a certain operation;

Button "Apply Mask": to apply the mask selected to the original raster loaded in previous step in such a way that the resulting image will display any pixels in their original color if their corresponding pixels in the mask are white, while other pixels in the raster will be displayed in black.

Button "Save Mask": to save selected masks into hard drive in ".tif" format;

Text Field: place where to enter an desired name, before clicking "rename" button or "Add To List" button if check box "Edit Mask Name" is selected;

Combo Box: containing nine mask operations (for details about each operation in this Combo Box, please see Note 3).

Note 3:

There are totally nine operations that can be executed on masks, which is divided into three categories:

1, operation that support multiple operands:

Add: to create a new image by adding the values of corresponding pixels in all operands;

Multiply: to create a new image by multiplying the values of corresponding pixels in all operands;

Average: to create a new image by adding the values of corresponding pixels in all operands and divided by number of operands;

And: to create a new image by performing Boolean "AND" operation on all operands;

Or: to create a new image by performing Boolean "OR" operation on all operands;

2, operation that need a target image and an action image:

Subtract: to create a new image by subtracting the values of pixels in action image from those of pixels in target image;

Divide: to create a new image by dividing the values of pixels in target image by those of pixels in action image;

Xor: to create a new image by performing Boolean "XOR" operation on target image by action image;

3, operation that need only one image:

Inverse: to create a new image by inverting values of all pixels in the operand

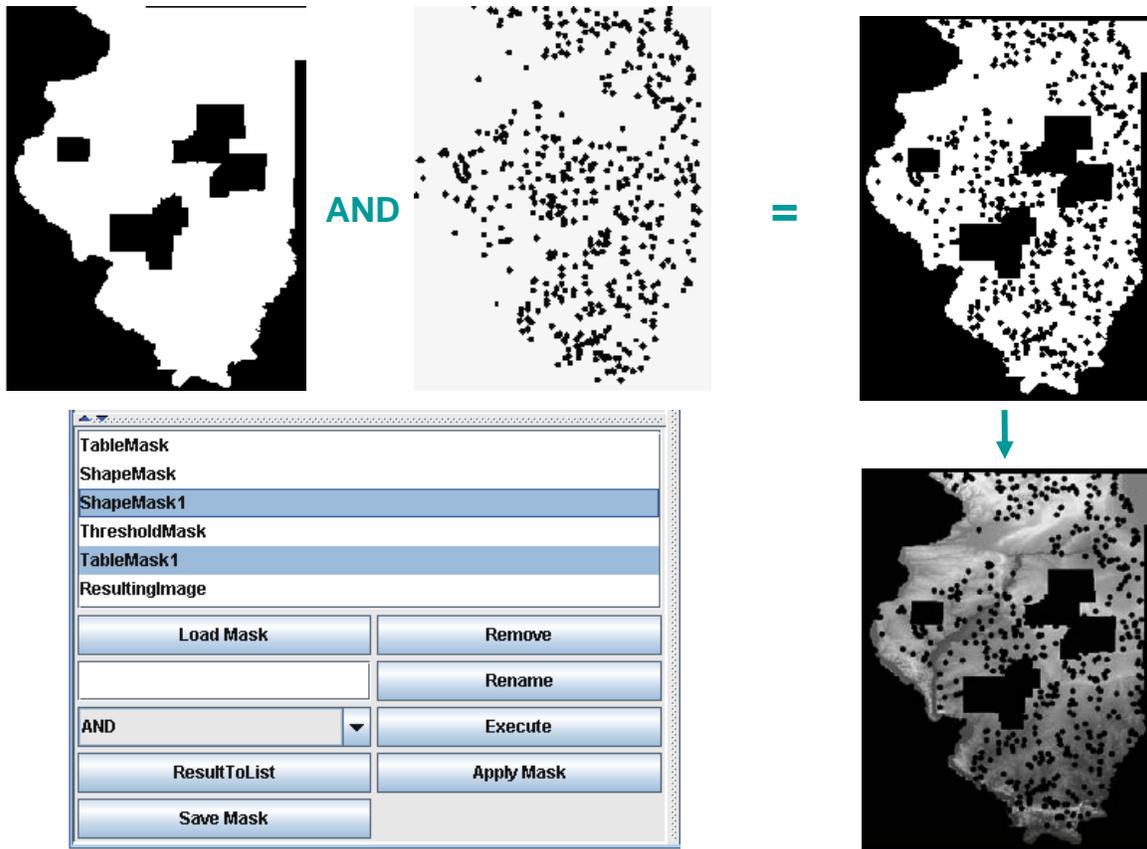


Figure 55: Mask operations.

Note 4:

1, Button "Load Table", "Load Shape", "Load Categorical" and "Load Threshold":

To load corresponding files from hard drive;

2, Button "Create Mask":

To create a new mask using method you select, taking into account whether checkbox "include" is selected;

3, Button "Add To List":

To add the mask being displayed in the image panel into list;

4, Button "Reset":

To go back to the initial stage of the process of mask creation, according to the mask creating method selected;

5, Button "Apply Mask":

To apply the mask selected to the original raster loaded in previous step in such a way

that the resulting image will display any pixels in their original color if their corresponding pixels in the mask are white, while other pixels in the raster will be displayed in black:

6, Button "Save":

To save any image being displayed in image panel onto hard drive;

3.4 Attribute Selection

This step allows you to select input and output attributes to be used for the decision tree algorithm. At least one attribute must be selected for both input and output selections. Multiple rows can be selected as input attributes but only a single row can be selected as the output attribute. Click on the attribute name to select it, or click on it again to deselect. After selecting attributes, you can go next. It is not recommended to select a row as not only input but also output attributes.

3.4.1 Options

If you want to see all data values in the masked area, click a check-box titled 'Show table'. In order to save the data values, click 'Save table' button. You can choose the file format between Excel and CVS formats. When a decision tree is constructed, the decision tree is pruned to make an optimal decision tree as default. But if you do not want the pruning, unselect a check box titled 'Prune tree?'

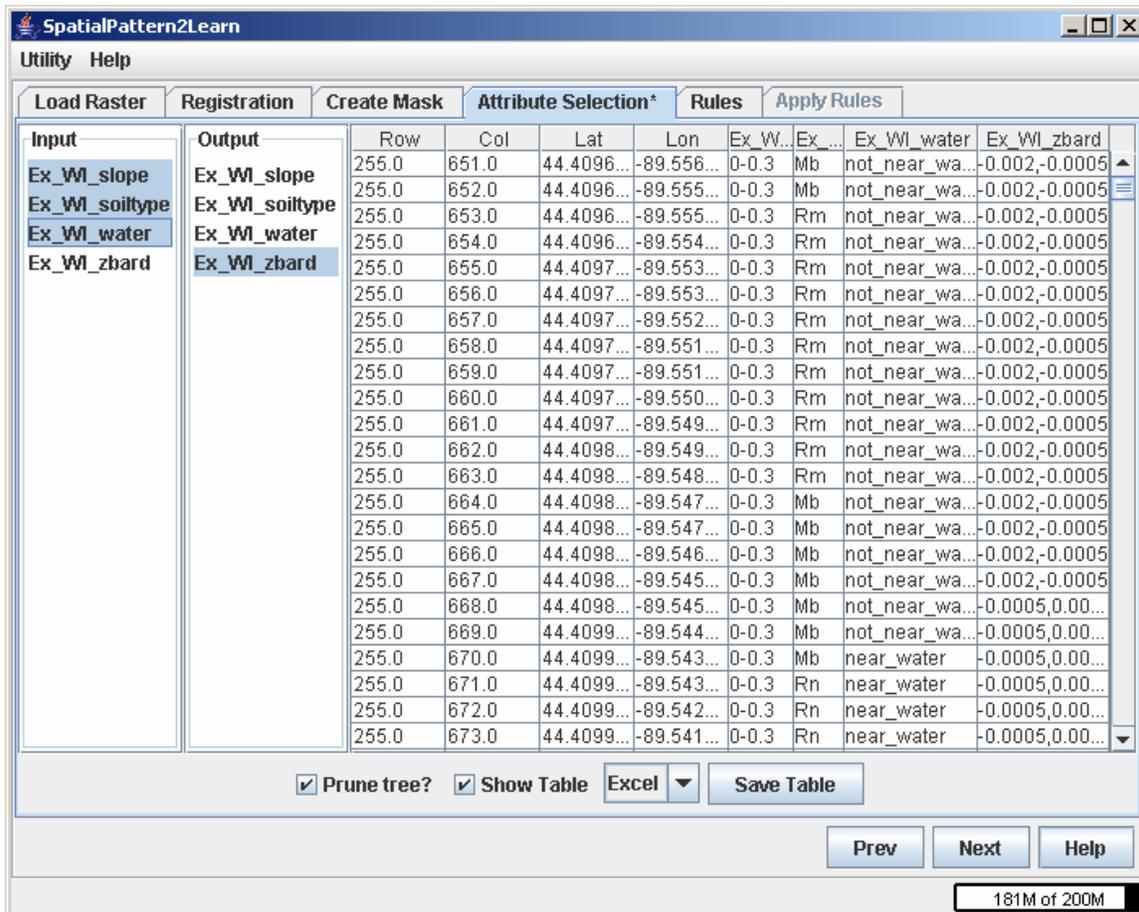


Figure 66: Step of attribute selection

3.5 Rules

This step shows all rules found by a decision tree based on attributes you select in the previous step. A rule consists of conditions and a class (conclusion). If the conditions are satisfied, the class is concluded. There are two types of conditions, positive and negative. Conditions with 'in' are positive conditions and those with 'not in' are negative conditions. For positive conditions, if there are multiple values, these have 'or' relationship. But, for negative conditions, these have 'nor' relationship.

If you click the header of a column, the column is sorted. An upside triangle means decreasing order and a downside triangle means increasing order.

Num	Support(%)	# of cases	Class	Conditions	Select
51	36.17	94	212,208,200	Attr. Ex_WI_soiltype in { Rm }	<input type="checkbox"/>
34	66.304	92	0.002,0.004	Attr. Ex_WI_soiltype in { WyC }	<input type="checkbox"/>
214	33.731	335	0.002,0.004	Attr. Ex_WI_soiltype in { Ov }	<input type="checkbox"/>
368	29.06	117	0.002,0.004	Attr. Ex_WI_slope in { 2.6-3.6 ,1.7-2.6 ,0.9-1.7 }	<input type="checkbox"/>
90	50.476	105	0.002,0.004	Attr. Ex_WI_soiltype in { Rn }	<input type="checkbox"/>
32	64.394	132	0.002,0.004	Attr. Ex_WI_soiltype in { 255,43,50 }	<input type="checkbox"/>
164	55.147	136	0.0005,0.002	Attr. Ex_WI_soiltype in { Mf }	<input type="checkbox"/>
91	72.5	40	0.0005,0.002	Attr. Ex_WI_soiltype in { Rn }	<input type="checkbox"/>
67	75	24	0.0005,0.002	Attr. Ex_WI_soiltype in { Ov }	<input type="checkbox"/>
141	44.118	170	0.0005,0.002	Attr. Ex_WI_soiltype in { RfB ,WyD }	<input type="checkbox"/>
92	64.345	962	0.0005,0.002	Attr. Ex_WI_soiltype in { Mf }	<input type="checkbox"/>
94	83.884	242	0.0005,0.002	Attr. Ex_WI_soiltype in { MfC ,Bt }	<input type="checkbox"/>
44	64.463	121	0.0005,0.002	Attr. Ex_WI_soiltype in { Rn }	<input type="checkbox"/>

MDL Score 42435.88 Show tree Export rules

Prev Next Help

163M of 267M

Figure 77: Step of rules

3.5.1 Columns

Num: Node number in a decision tree. Since rules are only leaf nodes in the tree, these numbers are not continuous. You can match a rule and a node using this number.

Support(%): Among all cases satisfying conditions, the ratio of cases having the same class(conclusion).

of cases: The number of cases satisfying conditions

Class: Conclusion of a rule

Conditions: Conditions of a rule

Select: This checkbox is for Applying rules. Only selected rules are applied at next step.

3.5.2 Options

MDL Score: MDL score of a decision tree. The less the score is, the better the tree is.

Show tree: Shows the decision tree in another dialog box as given Figure 18.

Export rules: Export all rules into a text file with XML format as shown Figure 19.

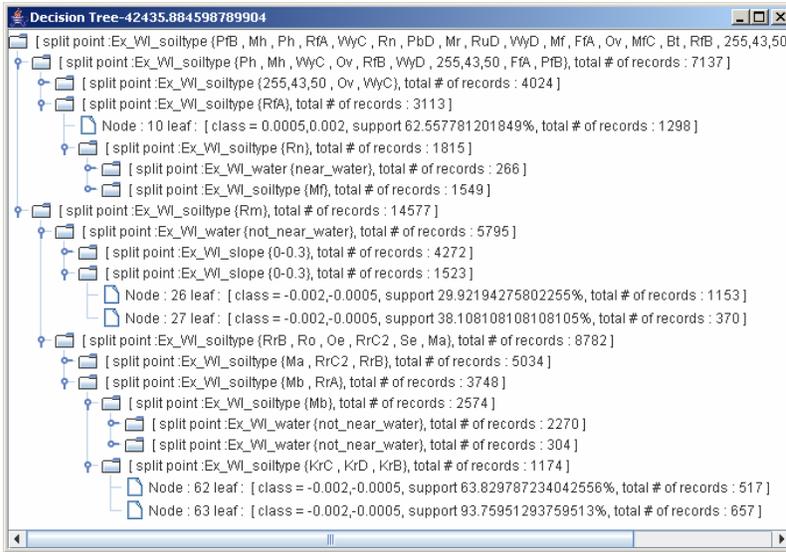


Figure 88: Decision tree

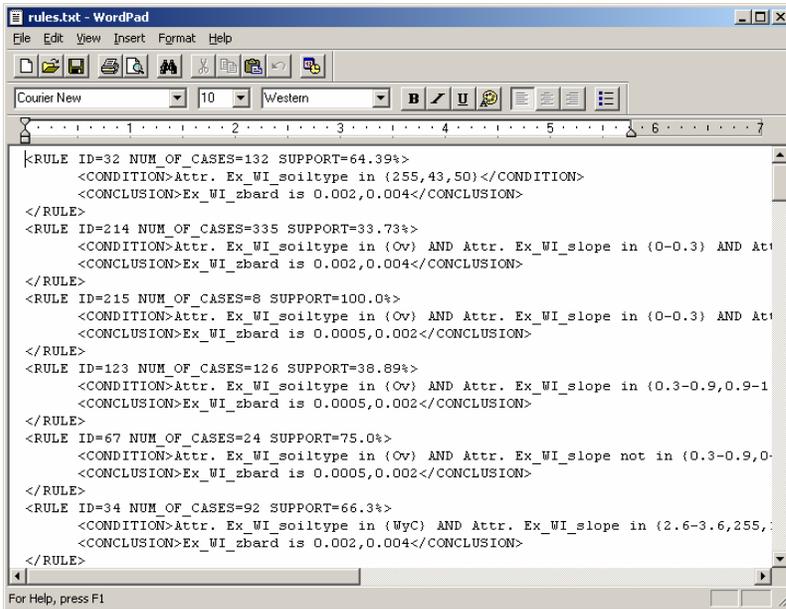


Figure 99: Exported text file of found rules

3.6 Apply Rules

This step shows images where rules you select in the previous step are applied. The image is the output attribute you select in the Attribute-Selection step. There are four images in the list box. The first image is the original image and the second is the applied image. The third image shows only changed pixels and the fourth shows the difference between original and applied values. Note that the selected rules are applied to only the masked area. You can save an image using the Save button.

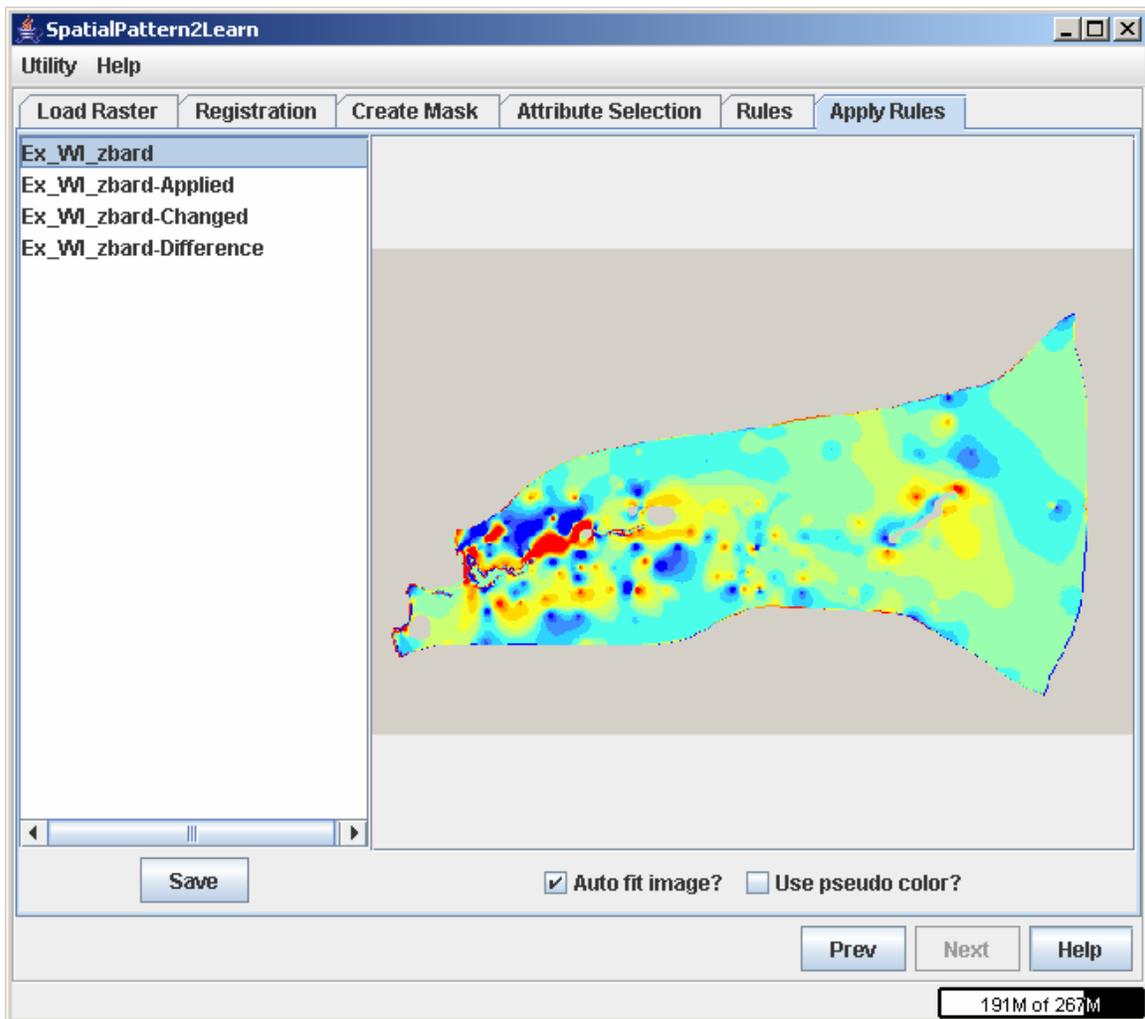


Figure 100. Original vales of the R/D model.

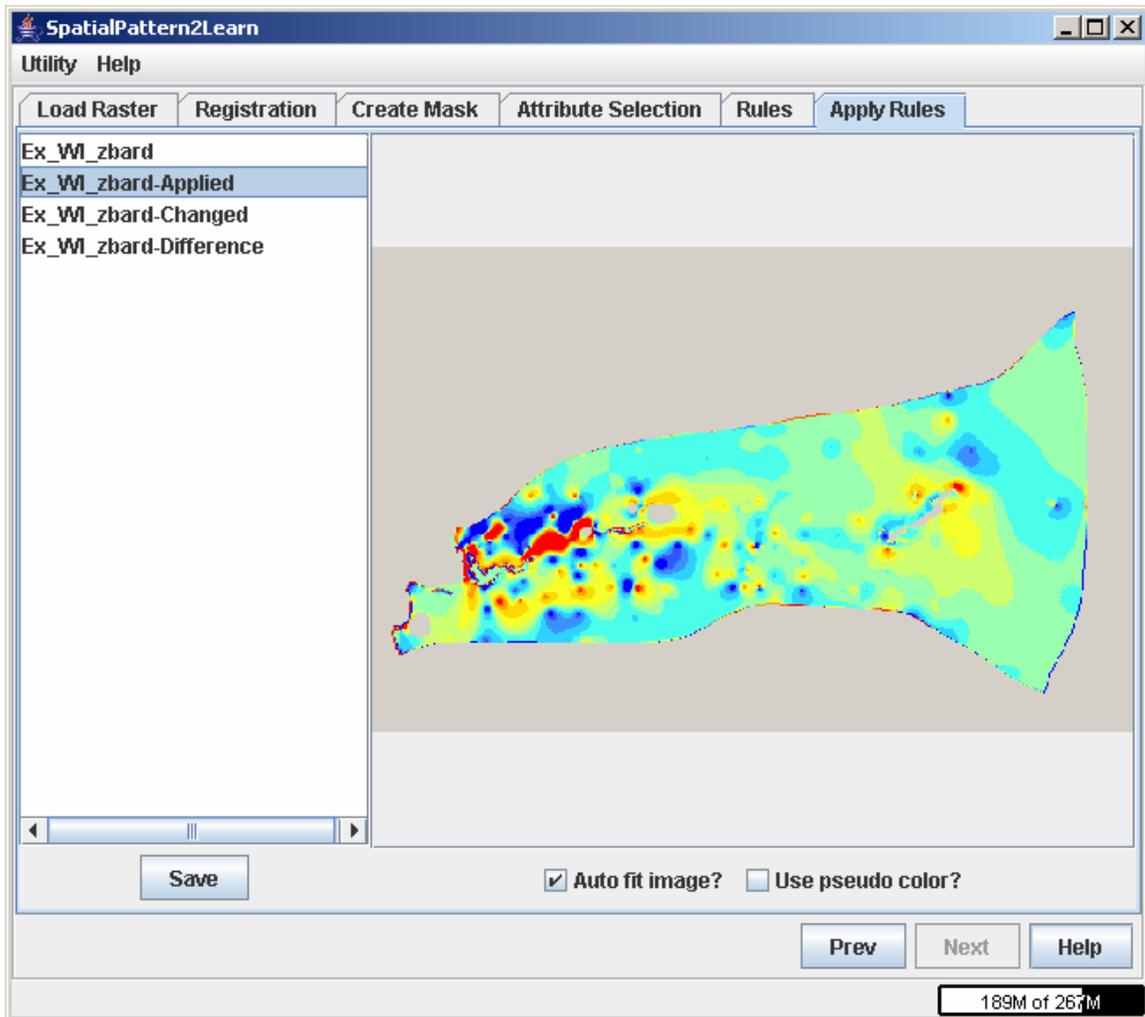


Figure 111. Modified vales of the R/D model using the rule:

```

<RULE ID=112 NUM_OF_CASES=1734 SUPPORT=70.24%>
  <CONDITION>Attr. Ex_WI_soiltype in {Mb} AND Attr. Ex_WI_water in
  {not_near_water}</CONDITION>
  <CONCLUSION>Ex_WI_zbard is -0.002,-0.0005</CONCLUSION>
</RULE>

```

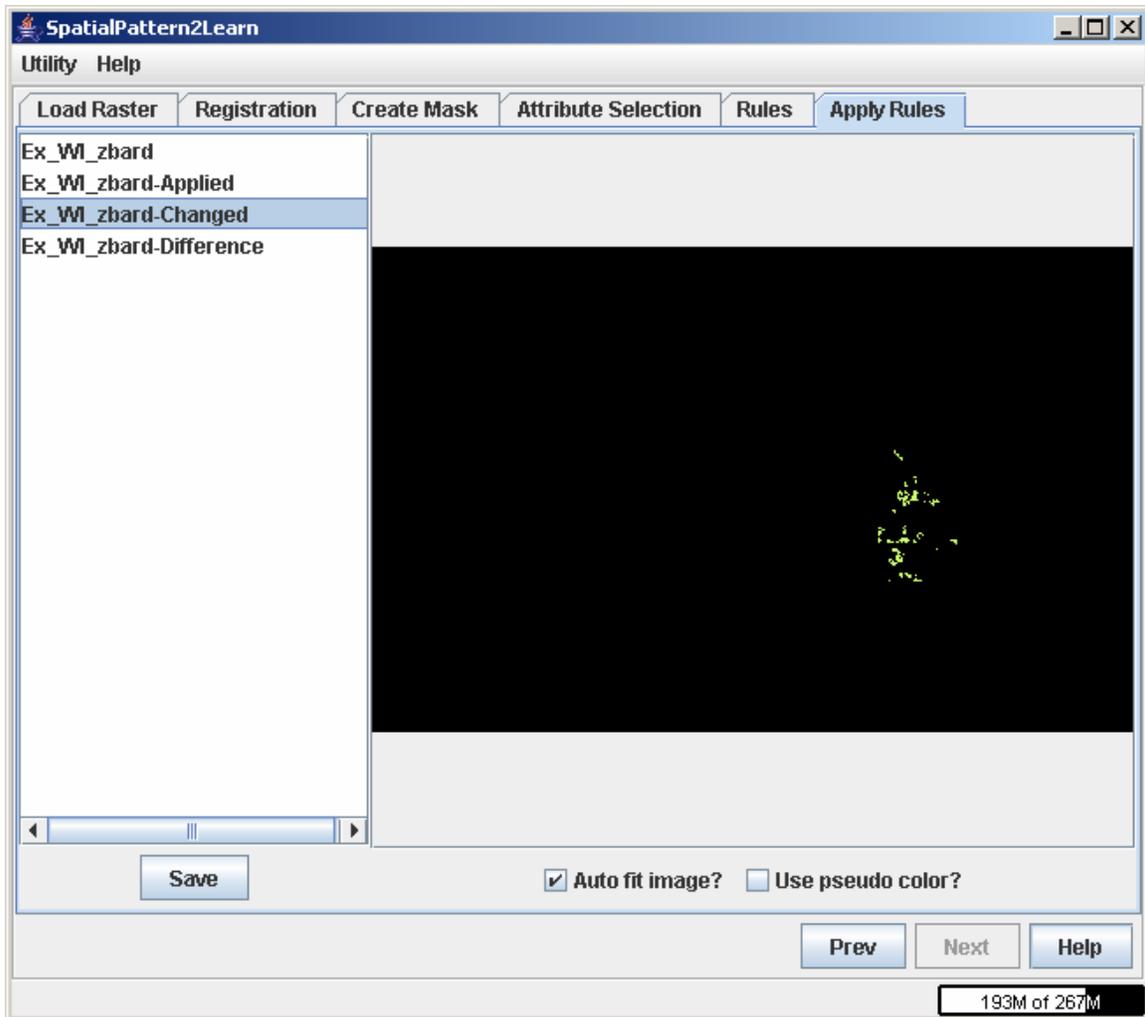


Figure 22: Part changed by selected rules

The following sections describe the user interfaces after launching SP2Learn. The first panel after launching the application is shown in **Error! Reference source not found..** An alternative (but less detailed) form of documentation is available in Help->Show Help or by clicking the **Help** Button.

Chapter 4 Additional Notes

4.1 Bug Reports and Bug Fixes

SP2Learn is research code. We appreciate your feedback on functionality and usability. If you find bugs, please, report them to isda@ncsa.uiuc.edu or record them at <http://isda.ncsa.uiuc.edu/bugs/>. However, we do not have resources to provide commercial quality software and code maintenance. Thus, the bugs will be fixed depending on our availability.

4.2 Known Problems

4.2.1 Out of Memory Errors

You can tell if out of memory errors occur through three different methods. An error called “Java heap space” will pop up in SP2Learn but will not display additional information.

If out of memory errors occur, please modify the `-Xmx512M` parameter in `SP2Learn.bat` to a higher number. The `-Xmx` parameter dictates the maximum amount of memory SP2Learn can use. To do this, open `SP2Learn.bat` in a text editor (e.g. NotePad) and change the 512 to a higher number. However, amounts above `~1400M` are generally too much for the Java Virtual Machine.

Note: If you use other versions of SP2Learn such as the unix version, please change the values in those files accordingly. See the figure below for an example of how to edit `SP2Learn.bat`.



Figure 212: Editing SP2Learn.bat

*Additionally, right clicking on the meter at the lower right hand corner of the SP2Learn window will force garbage collection. This means that Java will try to reclaim memory that is not in use. Please try to do this periodically.

4.2.2 Other Problems

*Save dialogs will display the file name of the last loaded or saved file. However, they will append the correct file extension when saving.

*The software assumes a screen of at least 1280 pixels wide. We are aware of this issue and are currently looking at redesigning the interface to allow for smaller screen resolutions.

4.2.3 Additional Notes

*If SP2Learn needs to be forced to quit, closing the command prompt window (labeled `C:\WINDOWS\system32\cmd.exe`) will quickly close the application.

*If you are experiencing installation problems, please check the registry values in `HKEY_CURRENT_USER->Software->JavaSoft->Prefs`. These values can be manually edited or deleted. To use the Registry Editor, click on the Start Menu->Run and type in “regedit”

4.3 Acknowledgments

This code called SP2Learn consists of the following packages: HDF, JAI, DOM4J, Apache LOG4J, Apache Commons Logging.

4.4 License Agreement

4.4.1 Illinois Open Source License

University of Illinois/NCSA

Open Source License according to

<http://www.otm.uiuc.edu/faculty/forms/opensource.asp>

Copyright © 2007, NCSA/UIUC. All rights reserved.

Developed by:

Name of Development Groups:

Image Spatial Data Analysis Group (ISDAG)

<http://isda.ncsa.uiuc.edu/>

Center for Ground Water Science (CGS)

<http://www.sws.uiuc.edu/gws/>

Name of Institutions:

National Center for Supercomputing Applications (NCSA)

<http://www.ncsa.uiuc.edu/>

Illinois State Water Survey (IWS)

<http://www.sws.uiuc.edu/>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal with the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimers.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimers in the documentation and/or other materials provided with the distribution. Neither the names of University of Illinois/NCSA/ISWS, nor the names of its contributors may be used to endorse or promote products derived from this Software without specific prior written permission.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE CONTRIBUTORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS WITH THE SOFTWARE.