GEOtest, a.s. tel.: 548 125 111

Šmahova 1244/112, 627 00 Brno fax: 545 217 979

ID No.: 46344942 Tax ID No.: CZ46344942 e-mail: info@geotest.cz

Geological and remediation work, geotechnical and hydrogeological investigation

Number and title of contract: **16 0327 Ethiopia – holistic management Arba Minch**

Client: Czech Development Agency, Nerudova 3, 118 50 Praha 1

Mendel University in Brno, Zemědělská 1665/1, 613 00 Brno

**ANNEX 3**

**CLASSIFICATION PROCEDURE GUIDEBOOK**

Project manager: **Mgr. Jan Oprchal**,Production Manager, GIS Specialist

Prepared by: **Mgr. Jiří Hladík**, GIS Specialist

**Bc**. **Cristina Medina Solano**,GIS Specialist

Approved by: **Ing. Jaromír Novák**, Production Manager

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**RNDr. Lubomír Klímek, MBA**

Managing Director

Brno, February 2018

Table of contents

[1. Introduction 3](#_Toc506300031)

[2. Data and software 3](#_Toc506300032)

[3. Procedure of classification 3](#_Toc506300033)

[3.1 NDVI 3](#_Toc506300034)

[3.2 Training phase 8](#_Toc506300035)

[3.3 Verification phase 13](#_Toc506300036)

# 1. Introduction

The main objective of the image classification is to automatically categorize all the pixels from the image into different classes that represent the topic or uses of soil. Usually, multispectral data are used to perform the classification; in this case, it is used a spectral pattern present inside the data of each pixel as a numeric base for the categorization. The remote sensing has the aim to identify and classify the materials of the terrestrial surface and all the processes that occur in it through the electromagnetic radiation that comes from itself (M.A. Gilabert et all. 1997).

So that in the present guide, we are going to follow the necessary steps to complete a total vegetation classification map on one study area.

# 2. Data and software

The data we need to complete the classification procedures is a satellite image of the whole study area and a reference image - that can be another satellite image or an orthophoto. If the satellite image has a good quality of resolution, no other picture is necessary. In this example, as a reference image, we used the Worldview-2 satellite image from Digital Globe 2016 with the PS (PanSharpening) image, offering a resolution of 0,6 m/pixel helping to identify some kind of vegetation via a visual analysis.

The software we are going to use in this guide is:

* ArcGIS 10.3
* ENVI 5.3

# 3. Procedure of classification

The multispectral data contain radiation information that each object on the surface has. Regarding the vegetation coverage, its reflectivity is determined by some external factors (illumination, type of observation and atmospheric perturbance) and by the optical characteristics and spatial distribution of its constituents. In this way, the vegetation indexes can help to complete the classification tasks.

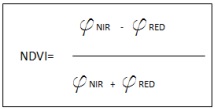
The vegetations indexes were designed to distinguish the vegetation from soil and help in this way the classification procedures. We can define the Vegetation Index as a parameter calculated from the values of the reflectivity of different wavelengths and that brings information in relation with the vegetation, minimizing the perturbance of the atmospheric conditions and soil (M.A. Gilabert et all. 1997).

## 3.1 NDVI

In this study, we applied the NDVI index to have a general view of the state of the vegetation in the study area. This index was introduced with the goal to separate the vegetation from the radiation that the soil produces (Rouse et al., 1974) but can be saturated in conditions where the vegetation is very dense (Chen *et al.,* 2013). This index is based on the radiometric behaviour of the vegetation in relation to the photosynthetic activity and the foliar structure of the plants, allowing to determinate how vigorous is the plant.

The NDVI works with the Red and NIR (Near Infrared) bands of the spectrum. The healthy vegetation shows a difference in the reflectance between these bands.

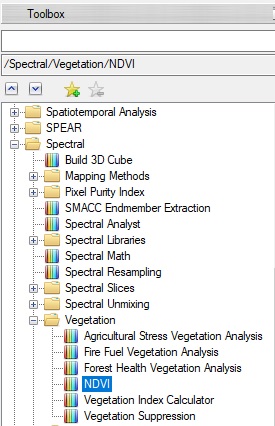
The following Eq.1 is the NDVI equation, where NIR is the reflectance in the near infrared band and RED is the reflectance in the red band of the visible spectrum.



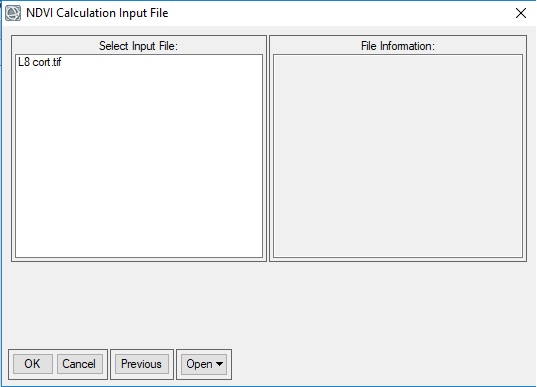
Eq.1: NDVI equation.

To calculate NDVI in the study area, we launch ENVI 5.3 software and follow the steps:

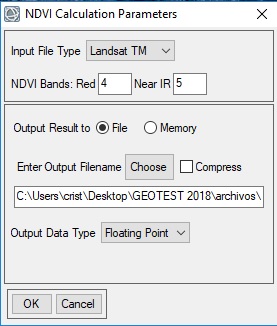
First, we open the satellite image with natural colours band combination (the combination depends of the satellite image is being worked, in this case, the base satellite image in a Landsat 8 satellite image of the Arba Minch-Chamo Mile area in Ethiopia), secondly, we go to the Toolbox and find the following path:



Once selected the “NDVI” option, the following windows appear:



Select the input file and go to the following window:



Here we have to select the Red and Near IR bands of the satellite image we are using, we have to check this information from the characteristics of the satellite image, it changes between one and another satellite although they belong to the same chain of satellites, Landsat 8 has different bands numeration than Landsat 7 for example. After that, we enter an output filename and continue. Select “Floating Point” Output Data type, and by pressing “OK” we finally have our NDVI file:



Now we can observe the differences in the tones of the image in greyscale colour. The green leaves reflect in the wavelength close to the NIR, while in the Red spectrum it absorbs because of the chlorophyll pigment. When the vegetation is unhealthy, chlorophyll pigments lose and start to reflect more in the Red band, therefore, the values that the NDVI shows go from -1 to 1, depending on the type of coverage.

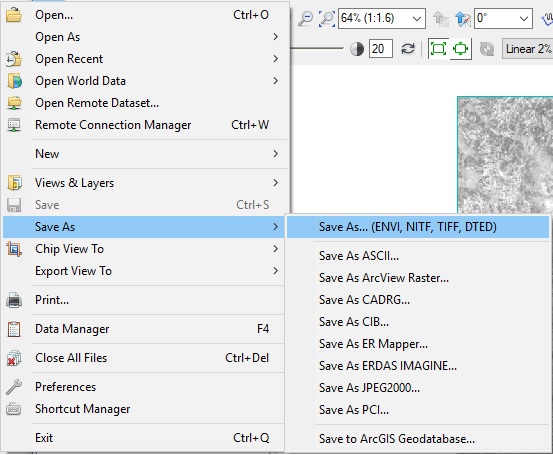
The water has a value of reflectance R>NIR, so values low to 0 mean water, clouds have same values in Red and NIR so values near 0 are clouds. Naked soil and ill/poor vegetation has values over 0 but near to it and finally the dense vegetation and healthy one has the high values of NDVI, the more near to 1, better is the state of the vegetation. The following table 1 shows values of NDVI and its correspondence.

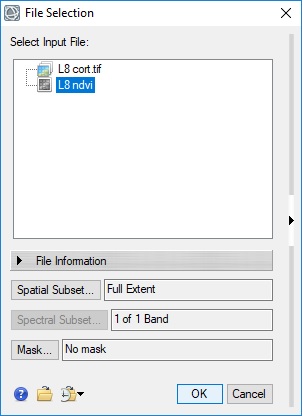
Tab. 1: Distribution of the NDVI values

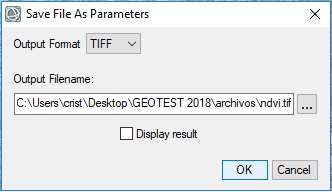
|  |  |
| --- | --- |
| **COVERAGE** | **NDVI VALUE** |
| Water | From -1 to 0 |
| Clouds | Near 0 |
| Naked soil and poor vegetation | From 0 to 0,20 |
| Health/Dense vegetation | Up to 0,20 |

NDVI has a high ecologic value due to the high power of the estimation of the fraction of the photosynthetic active radiation intercepted by the vegetation (Montheit, 1981), the primary production (Tucker et al, 1985) and the operation of the ecosystems (Virginia and Wall, 2001).

Now, to continue working with this image we export the NDVI image into a raster file in order to work with it in ArcGIS 10.3 software. To export the NDVI ENVI file we follow the next steps:







Now, we have our NDVI file in raster (.tiff) format and we can open it in ArcGIS 10.3 environment to generate a map and check the NDVI values in more detail.

We go to ArcGIS and open the NDVI raster file, now, we can observe the values of the NDVI in the study area and generate a complete map. The following picture 1 shows the screen of ArcGIS 10.3 where on the left side we can observe the values that in this case go from -0.25 to 0.51. We can compare now the values with those given in the previous table 1.

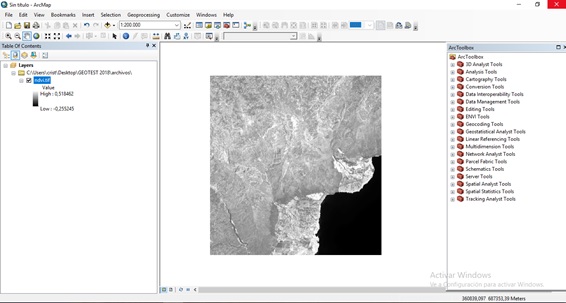


Fig. 1: Final screen shot in ArcGIS when NDVI is completed.

After we know the quality of the vegetation in the area, we can proceed to elaborate the classification vegetation study.

## 3.2 Training phase

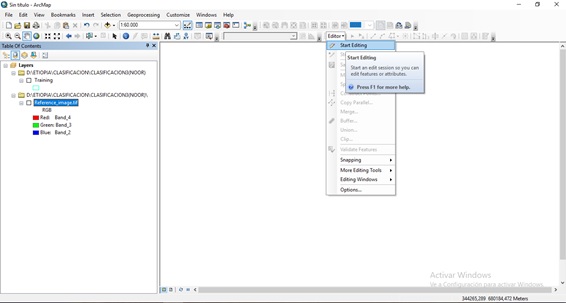
The classification procedure consists of two different phases: The training phase first and the second one, the verification phase.

At first, a training phase was made of the different types of vegetation observed in the study area, taking the satellite image Worldview-2 from Digital Globe 2016 with the PS image as a reference, offering a resolution of 0,6 m/pixel contributing to identifying some kinds of vegetation through a visual analysis. With this resolution, we can identify the different kind of trees, crops, urban areas, water, etc.

This first training phase is about drawing different polygons for all the surface of the image of the different uses observed. The training phase is performed in ArcGIS 10.3 software.

To perform the training phase, we use the PS satellite image from Worldview (from the PS image) and open it in ArcGIS.

Now we are going to draw different polygons over the different surfaces and save it with it correspondent name into the attribute table. Thus, we create a new shapefile and activate the Editor in ArcGIS and, going over the reference image, we draw as many polygons as possible from different uses of soil identified.



In the following Fig. 1 we can observe an example of the training phase, we can differentiate particular uses of soil signed in different colours each. The polygons should be drawn in homogeneous areas, not touching one with another and the size is free is that to say, the size can be flexible but it is important to have a big number of polygons from each use.

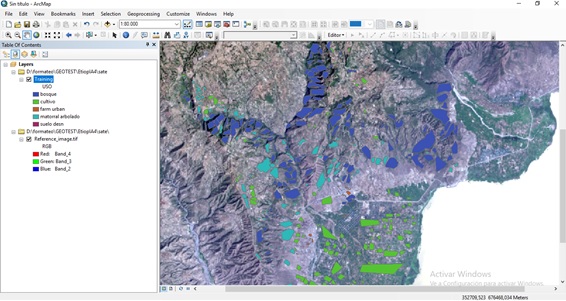
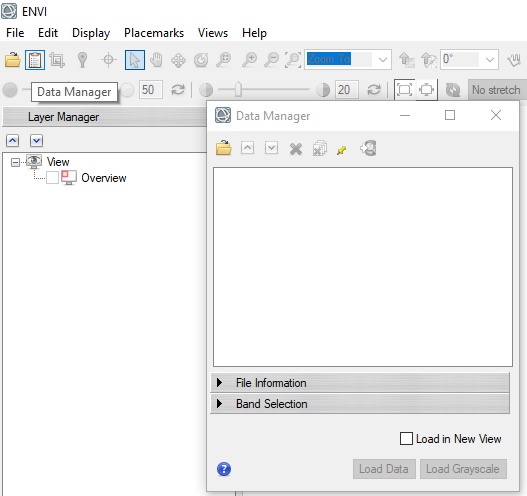


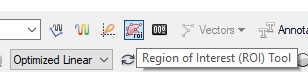
Fig. 2: Training phase process example

After completing the training phase, we check if the process is going on well and we will make a separability spectral analysis in ENVI 5.3 software. This test will be the similar procedure that we have to perform to complete the classification.

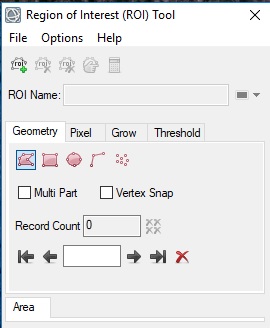
First, open ENVI 5.3 software and go to the following path to change the satellite image:



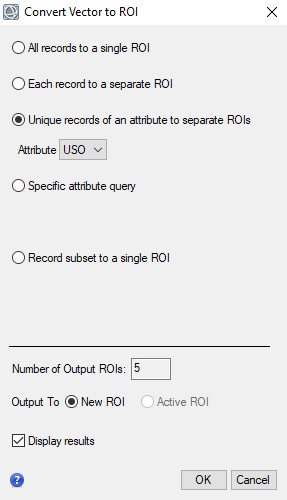
Once opened, we import the Training shapefile we have generated in ArcGIS previously. The polygons we have drawn are ROI (Regions of Interest), so, to import them to ENVI, go to the Menu bar:



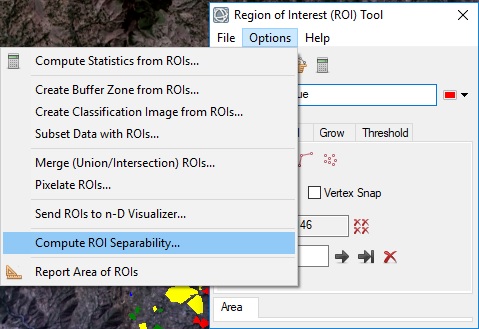
And the follow window will appear:



Press **File>Import vector**,and select the Training shapefile. The window, shown below, will appear. Here, we select “Unique records of an attribute to separate ROIs”, in “Attribute” we select the name of the column, where we specified the land use. We check if the number of ROIs is the same as the number of uses of soil that we have in our Training shapefile in the box under the line (in this example the number of ROIs is 5, so it is correct), and finally press “OK”.



From this moment, ENVI will start the process of importing, this process will take some time depending on the quantity of ROIs (number of uses differentiated) and the hardware used. Once the importation is finalized, we can compute the “ROI separability” to check if our Training phase is being completed correctly.



ROI separability is the tool that computes spectral separability between selected ROI pairs for a given input file. Both the Jeffries-Matusita and Transformed Divergence separability measures are reported. These values range from 0 to 2.0 and indicate how well the selected ROI pairs are statistically separated. Values greater than 1.9 indicate that ROI pairs have good separability. For ROI pairs with lower separability values, we should attempt to improve the separability by editing ROIs or by selecting new ROIs. For ROI pairs with very low separability values (less than 1), we might want to combine them into a single ROI (Harrys geospatial, <http://www.harrisgeospatial.com/docs/RegionOfInterestTool.html#ROISeparability>).

In Fig. 3, there is an example of how the ROI separability file looks like. In the picture, there only is an example of about two of the uses of soil from the total, but the process is completed with all the uses and compares each one with all the rest.

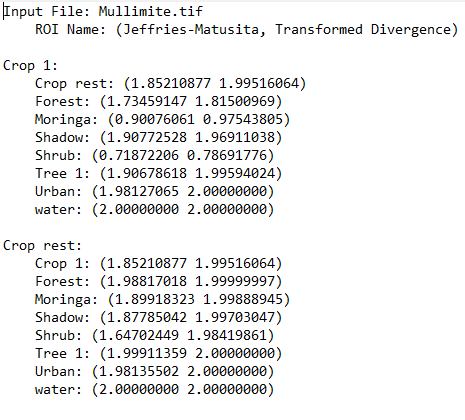


Fig. 3: ROI separability file example.

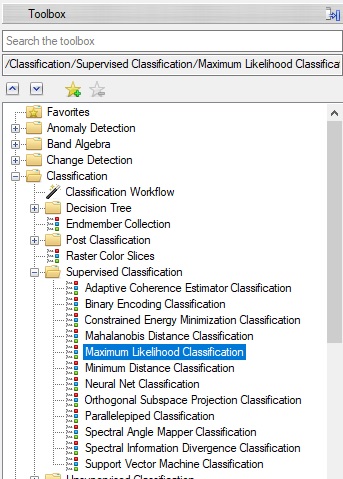
At the end of the file, there is a sum with all the uses ordered from the best separability to the worst one. As commented before, if the separability has values lower than 1, the uses should be unified, between 1 to 1.5 we should improve the separability drawing more polygons and over 1.9 the separability is perfect. We should repeat this process until the spectral separability of the ROIs are in the correct values and then we can continue to the following phase, the Verification phase.

## 3.3 Verification phase

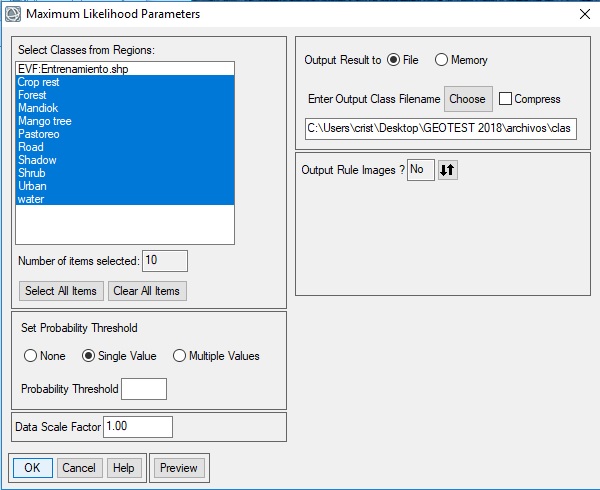
So, after completing the training phase, we go to the verification phase. This verification phase consists in improving the spectral separability by making more polygons of each use. For this, we will use the same shapefile that we used in the training phase, but with the difference that we will draw more polygons of each use in it, in this way the verification phase is more complex because the quantity of the polygons will increase considerably.

After editing the new polygons, we will perform the spectral separability analysis in ENVI, following the same steps specified previously, if the values are correct we can proceed to make the classification process.

With the ROIs changed and the satellite image opened, we go to Toolbox>Classification>Supervised classification>Maximum Likelihood Classification:



Once opened the classification function, it will pop up the following window:



On the left top, we can see all the ROIs that we have imported, click on “Select All Items”. On the top right, we click on “File” and below that we select the destiny and name of the output file. In “Output Rule Images” select “No”. Click “OK” and the process will start. We wait until the process ends and finally we will have the full classification image. On Fig. 4, there is an example:

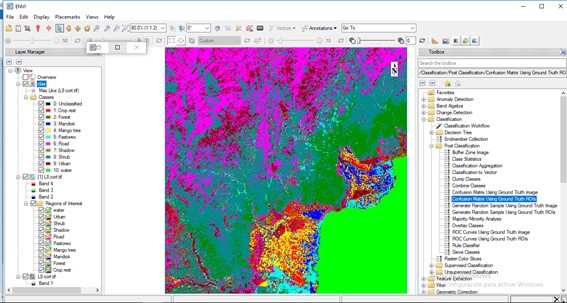


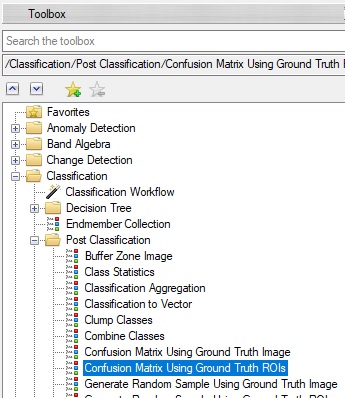
Fig. 4: Screenshot with the full classification image completed

Once we have the classification image completed, we are going to check the quality of the classification. To perform this task, we will make an analysis post classification, we will make the Confusion Matrix analysis. This analysis will give us the “overall accuracy” that it is calculated dividing the total number of pixels correctly classified by a total number of reference pixels being the result the total percentage. In Fig. 5, there is one example of how is the data that the Confusion Matrix Analysis gives.

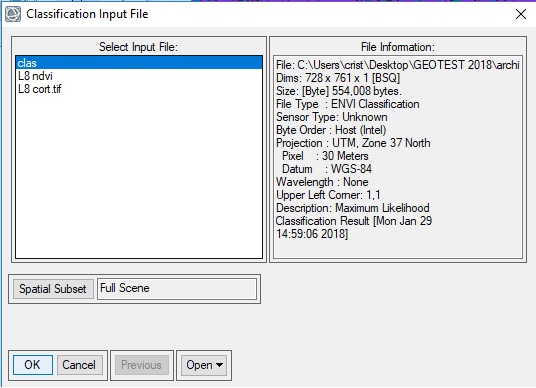


Fig. 5: Example of Confusion Matrix Analysis result

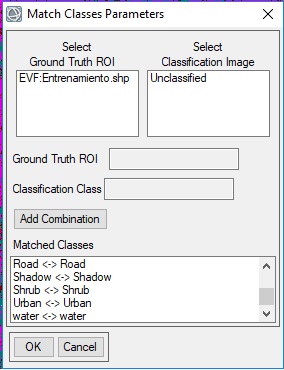
To make the Confusion Matrix Analysis we go to: Toolbox>Classification>Post Classification>Confusion Matrix Using Ground Truth ROIs



The following window will appear, in the box, we select the satellite image in which we want to perform the Confusion Matrix, click “Ok”.

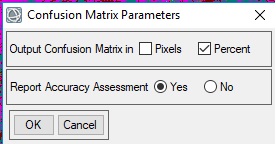


After selecting the satellite image, we have to select what parameters we want to analyse and we have to select them in the following window:



In this window, we set the combination of the ROIs, so we have to select all of them and they must appear in the box called “Matched classes” as shown in the example. Once made this selection, click “OK”.

The following window will be the last one we have to complete. Here we select that the “Output Confusion Matrix” is in percent and mark “Yes” in the “Report Accuracy Assessment” as shown in the example, in this way we will obtain the Report accuracy in percentage.



The Kappa Coefficient is the total accuracy in values from 0 to 1. The following picture shows an example of the result in a Confusion Matrix Analysis. The appropriate result for a vegetation classification analysis is minimum 80% of Overall Accuracy, if we don’t get this result we should return to the previous steps in the elaboration of “Training” and “Verification” polygons shapes and rebuild the process to obtain the maximum efficacy as possible.

