Exercise 5-7
Vegetation Analysis in Arid Environments

In this exercise, we will explore the use of different vegetation index calculation models available in the VEGINDEX, TASSCAP and PCA modules to analyze vegetation cover. Before continuing, you may find it useful to read or review the Vegetation Indices chapter in the IDRISI Guide to GIS and Image Processing. That chapter provides an extensive overview of many vegetation indices, only some of which will be used in this exercise.

Introduction to Vegetation Indices

Vegetation cover was an early focus of research in natural resources management using space-born satellite images, especially with the release of the Earth Resources Technology Satellites known as Landsat in 1972. Landsat, SPOT and NOAA data offer time series images that are widely used to monitor and assess the status of vegetation at the global, regional, national and local levels. Vegetation indices use various combinations of multi-spectral satellite data to produce a single image representing the amount of vegetation present, or vegetative vigor. Low index values usually indicate little healthy vegetation while high values indicate much healthy vegetation. Different indices have been developed to better model the actual amount of vegetation on the ground. The index that is most appropriate for use in a particular environment can best be determined through calibration with sample measurements of biomass. In the absence of biomass measurements, these index images can be useful indicators of the relative amount of vegetation present.

Vegetation has a characteristic spectral response pattern in which visible blue and red energy is absorbed strongly, visible green light is reflected weakly (hence giving vegetation its green color) and near infrared energy is very strongly reflected. Because of this characteristic spectral response pattern, many of the vegetation index models use only the red and near-infrared imagery bands.

Introduction to the Data and the Study Area

In this exercise, we will assess vegetation cover and its changes in an area of southern Mauritania.

a) Display the image MAUR90-BAND3 with the GreyScale palette and choose to autoscale the image using Equal Intervals.

The area covered by the images in this exercise is near the Senegal/Mauritania border and contains part of the Senegal River flood plain as well as the lower section of the Gorgol River flood plain (partially visible at the upper left corner of the image). This is a tributary of the Senegal River. These sections of the two rivers are covered by riverine vegetation dominated by the *Acacia nilotica* species, the preferred species for fuelwood and charcoal. Other woody species such as *Borassus flabellifer* and *Iphaene thebaica* are used as building material. Rainfed and flood recessional agriculture and grazing are

---

1. Of the 19 vegetation indices produced in the VEGINDEX module of IDRISI, only the RVI and NRVI produce images with high values indicating little vegetation and low values indicating more vegetation. If you are using a vegetation index model not provided in VEGINDEX, you must determine whether the index values are proportional or inversely proportional to the amount of vegetation present before you can properly interpret the image.

also practiced in this region.

Once a relatively humid area, persistent rainfall deficits since the late 1960s have left the study area, as well as more and more of the Sahel, semi-arid. Much vegetation has shifted from savanna to steppe. Relics of the savanna vegetation are only found along river valleys on clay, clay sand and sandy clay soils, since these retain moisture better than other soils in the area. Increasing pressure from populations trying to adapt to the continuous drought conditions has been the main cause of vegetation cover degradation in this environment.

Quantifying the low density vegetation cover that characterizes arid and semi-arid lands is especially challenging because vegetation cover is not complete - most pixels contain an average reflectance of vegetation and bare soil. Some of the vegetation index models we will use have been developed specifically to help account for the effects of background soil reflectance.

The data we will use are Landsat Multi-spectral Scanner (MSS) images. These images were taken on October 10, 1980 and October 12, 1990 by Landsat 4. There are eight images provided in the dataset, four from each year: MAUR80-BAND1, MAUR80-BAND2, MAUR80-BAND3 and MAUR80-BAND4 for 1980; MAUR90-BAND1, MAUR90-BAND2, MAUR90-BAND3 and MAUR90-BAND4 for 1990. These correspond to MSS bands visible green, visible red, near-infrared and a slightly longer-wavelength near-infrared, respectively. Since the two scenes were taken at two different dates, they must be registered to one another if we are to do analysis between them. This task has already been performed using a methodology similar to that described in Exercise 4-5. We will begin the exercise by producing and comparing several vegetation indices for the 1990 scene, then we will analyze changes between the two scenes.

Creating Vegetation Index Images

There are three major families of vegetation indices that we will explore: Slope-Based, Distance-Based and Orthogonal Transformation vegetation indices.

The Slope-Based VI's

The Slope-Based VI's use the ratio of the reflectance of one band to that of another, usually the red and the near-infrared. The term Slope-Based is used because in comparing resulting VI values, we are essentially comparing the slopes of lines passing through the origin and the pixels as plotted on a graph with the reflectance of one band as the x-axis and the reflectance of the other as the y-axis.

b) Before beginning our exploration of vegetation indices, select User Preferences from the File menu and set the "Automatically display the output of analytical modules" feature on. We will always display the VI images with a user-defined palette named NDVI. Go to the Display tab of the User Preferences dialog box and enter NDVI as the Quantitative Palette. Also, choose to show titles, but do not show legends (this will maximize display space). Click OK to save the settings and exit User Preferences.

c) Use the module VEGINDEX (Image Processing/Transformation menu) twice to produce images for two of the slope-based models: Ratio and NDVI. Use MAUR90-BAND2 as the red band and MAUR90-BAND3 as the near infrared band. Call the resulting images 90RATIO and 90NDVI. Examine each of the output images. Consult the on-line Help System for details about the equation used for each index.

1. What similarities and differences do you notice between the two output images? (In answering this question, it may be useful to look at the pair of images with other quantitative palettes as well, such as GreyScale or Quant.) What is the purpose of normalizing the Ratio to create NDVI? (You may wish to consult the Vegetation Indices chapter for help in answering this question.)

The slope-based VI's are simple linear combinations that use only the reflectance information from the red and infrared
bands. In contrast, the second family of Vegetation Indices that we will explore, the distance-based VI's, uses information about the reflectance characteristics of the background soil in addition to the red and infrared bands.

**The Distance-Based VI's**

The reflectance values recorded by the sensor for each pixel constitute an average reflectance of all the cover types in the instantaneous field of view (i.e., the pixel). When vegetation cover is not complete, which is particularly the case in arid and semi-arid regions, the average reflectance values are greatly influenced by the background soil type. The Distance-Based VI's address this problem of separating information about vegetation from information about soils in remotely sensed data.

The distance-based indices are based on the concept of a soil line and distances from that soil line. A soil line is a linear equation that describes the relationship between reflectance values in the red and infrared bands for bare soil pixels. This line is produced by running a simple linear regression between the red and infrared bands on a sample of bare soil pixels. Once that relationship is known, all unknown pixels in an image that have that same relationship in red and infrared reflectance values are assumed to be bare soils. Unknown pixels that fall far from the soil line because they have higher reflectance values in the infrared band are assumed to be vegetation (based on the characteristic spectral response pattern for vegetation where the infrared band reflectance values are relatively higher than those of the red band). Those that fall far from the soil line because their red reflectances are high are often assumed to be water (based on the characteristic spectral response pattern for water where the red band reflectance values are relatively higher than those of the infrared band).

Inputs to the calculation of the Distance-Based VI's are the red band, the infrared band, the slope of the soil line and intercept of the soil line. (In addition, some of these VI's also require a scaling factor.)

The first step in calculating the soil line is to identify a sample of bare soil pixels in the image. We will use the 90NDVI image created earlier to develop a mask image for bare soil. (If better knowledge of the area were available, we could on-screen digitize known bare soil areas.)

2. If you assume that any pixel having a higher infrared than red reflectance is vegetation and everything else is bare soil, what threshold value could you use with the 90NDVI image to separate vegetation from bare soils? (Hint: Use the NDVI equation with some example values to help you answer this question.)

d) Run RECLASS with 90NDVI to create the image SOILMASK. Assign the new value 1 to bare soil areas and the new value 0 to vegetated areas.

Once the bare soil areas have been identified, the values for those areas in the infrared and red bands are submitted to linear regression to calculate the soil line. The soil line calculation is not the same, however, for all the distance-based VI’s. Some are based on a regression where the red band is evaluated as the independent variable, and some are based on a regression where the infrared band is evaluated as the independent variable. Since we will be creating both types of distance-based VI’s, you will need to run the regression twice to determine two soil lines.

e) Run REGRESS (from the GIS Analysis/Statistics menu) twice, between the MAUR90-BAND2 and MAUR90-BAND3 images, using SOILMASK as the mask image. Write down the slope (b) and intercept (a) values for the case in which the red band is treated as the independent variable and for the case in which the infrared band is the independent variable.5

3. What are the slope and intercept when the red band is the independent variable? When the infrared is the independent variable? What is the coefficient of determination ($r^2$)?

---

3. The equation written at the top of the REGRESS display is in the form $y=b+ax$, where $y$=independent variable, $b$=intercept, $a$=slope, and $x$=dependent variable.
The coefficient of determination is quite high, indicating that the relationship between red and infrared reflectance for these bare soil pixels is described well by a linear equation.

f) Run VEGINDEX three times to produce the distance-based VI's PVI, PVI3, and WDVI. For each VI, refer to the Help System section Determining Slope and Intercept Values under VEGINDEX to determine which soil line parameters to use for each particular VI. Also refer to the Vegetation Indices chapter in the IDRISI Guide to GIS and Image Processing for details about the equation used for each index.

4. What are the major differences you see in the displays of the three distance-based vegetation index images produced?

5. Is there a noticeable difference between these three images (on average) and the two slope based images (on average) produced earlier? In other words, would you be able to separate the five output images into two families based solely on the resulting images?

The Orthogonal Transformation VI's

The final group of vegetation indices we will explore are the Orthogonal Transformation VI's. With these VI's, four or more bands of imagery are transformed into a set of new images, one of which describes vegetation. We will explore the use of the Tasseled Cap and Principal Components transformations for producing vegetation images.

The Tasseled Cap transformation uses a set of four MSS multi-spectral images to produce four new images. The Green Stuff or Green Vegetation Index (GVI) image represents vegetation. Other images produced represent Soil Brightness Index (SBI), Yellow Vegetation Index (YVI) and Non-such Index (NSI). The name of the transformation describes the shape of a plot of pixels in GVI-SBI space for an image having vegetation in many stages of development. The Tasseled Cap was developed to represent the most important information from a multi-band agricultural scene in only two images - GVI and SBI.

g) Run TASSCAP from the Image Processing/Transformation menu. Indicate that you will be using MSS data and enter the four bands for the 1990 scene. Give 90 as the prefix for the output files. This will produce four images called 90GREEN, 90BRIGHT, 90YELLOW and 90NOSUCH. Display the four images. (Auto-display is disabled for modules that produce more than one output image.)

6. Why do you think the areas indicated as having high amounts of vegetation in the green vegetation image show low values in the soil brightness image?

The Tasseled Cap transformation uses global constants (i.e., the values don't change from scene to scene) to weight the bands being transformed. Because of this, it may not be appropriate to use in all environments. Principal Components analysis, on the other hand, is a scene-specific transformation of a set of multi-spectral images into a new set of component images. The component images are uncorrelated and are ordered according to the amount of variation they explain from the original band set. The first of these component images typically describes albedo, or brightness, (which includes the background soil) and the second typically describes variation in vegetative cover.

h) Run PCA from the Image Processing/Transformation menu. Choose to calculate covariances directly. Enter 4 as the number of input bands and enter the four 1990 MSS images as input bands. Enter 4 as the number of components to be extracted. Give 90 as the output file prefix. Accept the default to use unstandardized variables. When the processing is finished, display the resulting four images, 90CMP1 through 90CMP4.

The tabular information produced by PCA indicates that the first component describes nearly 93% of the variance in the original set of four bands. All the input bands have high and positive loadings for component one. We might then interpret this component as describing the overall image "brightness." The second component has positive loadings for both

4. The transformation can also be used with six TM images. In this case, three output images are produced, representing greenness, brightness and moistness.
infrared bands and negative loadings for the visible green and red bands. It can be interpreted as an image describing vegetation, independent of the overall scene brightness. Components three and four describe little of the original variance and appear to represent atmospheric and other noise in the images.

The equation used for the GVI image of the Tasseled Cap transformation also weights the infrared bands positively and the visible bands negatively, though the weighting values are somewhat different. It is therefore not surprising to see great similarity between the second component image and the GVI image produced earlier.

### Comparing Vegetation Indices

It is possible to visually compare all of the vegetation index images we have produced. Some obviously have better contrast than others. Some seem to show more variation within the low-value areas. However, without ground-truth information about the status of vegetation in the area in 1990, we cannot determine which indices are most useful. What we will do is analyze the set of images as a whole to see what different characteristics are illustrated by the various indices.

To do this, we will submit all of the VI images we have created in this exercise to a Principal Components analysis. The IDRISI PCA module requires that input data be byte binary. Our vegetation index images are in real format. To convert the vegetation index images to byte binary format, we will use the module STRETCH. We will perform a linear stretch on each image to transform the original range of values to the new range of 0-255.

i) Run STRETCH with each of the seven VI images produced earlier. (Do not use 90NOSUCH or 90YELLOW.) Choose a linear stretch with 256 as the number of levels in the output. Give output filenames that are the input filenames plus a "-BYTE" suffix (e.g., to create 90RATIO-BYTE, 90NDVI-BYTE). To save time, you may wish to use Macro Modeler.

j) Run the PCA module. Choose to calculate covariances directly and indicate 7 input bands. Enter the names of the seven stretched VI images. Choose to extract 4 components. Give VI as the output image prefix and choose to use standardized variables creating integer images (the last option). The output images will be called VICMP1, VICMP2, VICMP3 and VICMP4. Display these images.

The component images describe the most important "patterns" present in the 7 input vegetation index images. The first component image shows that pattern which is most common to all the input images. The second component image shows the next most important pattern remaining after the first has been removed, and so forth. The statistics produced by PCA include information about the percent variance explained by each component and the weightings (loadings) of each input image on each component.

7. Compare VICMP1 with the input stretched VI images. Which resemble it most? Are the loadings of those input images high compared to the others for that component?

Recent research indicated that in a similar study comparing 25 VI images, the first component described a general vegetation index, including elements of greenness and soil background. The second component represented those VI's that corrected for soil background, and the third described soil moisture.

---

5. GVI = \[(-0.386\text{MSS4})+(-0.562\text{MSS5})+(0.600\text{MSS6})+(0.491\text{MSS7})]\] In the naming of the image files for this exercise, MAUR90-BAND1 corresponds to MSS4 in the equation, MAUR90-BAND2 to MSS5 and so forth.

Change Analysis using Vegetation Index Images

We will now undertake an analysis between the two dates of imagery. We will be concerned with identifying areas that have undergone significant change between 1980 and 1990.

k) Display MAUR80-BAND3, the near infrared band of the 1980 image, using the GreyScale palette and autoscaling with Equal Intervals.

Unfortunately, the data we have for 1980 has significant horizontal "striping" effects due to sensor miscalibration. It is, however, the best available data for that time and study area, so we will use it.

l) Choose any one of the vegetation indices you used with the 1990 scene and produce a corresponding image for the 1980 data. If you choose a distance-based VI, you will need to find new soil line parameters for the 1980 data since soil moisture conditions may be quite different between the two dates and areas of bare soil may have changed.

The most elementary of change analysis techniques is visual comparison.

m) Look at the VI image pairs for the two dates and try to determine areas where changes in vegetation are evident. The striping that is apparent in the 1980 scene is an artifact of the sensor system. Use HISTO with the two vegetation images and note the average value for the entire image.

8. Does it appear that there is generally more or less vegetation in 1990 than in 1980?

The closest rain gauge station to this area is the town of Mbout, located outside the image to the East. The station recorded approximately 200 mm of rain in 1980 and 240 mm of rain in 1990. Since rainfall and vegetation cover are highly correlated, we can expect to see generally higher vegetation index values in the area for 1990 than for 1980.

There are many quantitative methods we can use to analyze change between images. Here we will explore only one, simple differencing. For a more complete treatment of change analysis techniques, see the Time Series/Change Analysis chapter in the IDRISI Guide to GIS and Image Processing. You may use the data from this exercise to explore on your own many of the techniques presented in that chapter.

With simple differencing, we merely subtract one image from the other, then analyze the result. The critical issue then becomes one of setting an appropriate threshold for the difference image beyond which we consider real change, as opposed to ephemeral variation, to have occurred. Ground truth information would normally be used to identify these thresholds.

n) Use OVERLAY to subtract your 1990 image from your 1980 image. Call the resulting image 1980-1990. Use HISTO with 1980-1990 and change the class width to be small in relation to the range of values in 1980-1990. (The class width will differ depending on the particular VI you chose to use. Make sure there are at least 100 "bins" or divisions in the histogram.) Note the distribution of values, as well as the mean and standard deviation.

In the absence of ground truth information to guide our selection of a suitable change/no-change threshold, we will use the standard deviation. We will consider that only those pixels lying beyond two standard deviations from the mean in either the positive or negative direction constitute real change and those lying within two standard deviations represent normal variation. In a normal distribution, 90% of the values fall within two standard deviations. By setting this as our threshold, therefore, we are identifying the outlying 10% of pixels as our significant change areas.

o) Use RECLASS with 1980-1990 and the mean and standard deviation values you found above to create a new

7. You may wish to try to mitigate the striping by using Fourier analysis with these 1980 images. Use the forward transform, filter out the horizontal elements, then use the backward transformation. See the chapter Fourier Analysis in the IDRISI Guide to GIS and Image Processing for more information.
What is the distribution of positive and negative change areas in the study area? (Try to disregard change that is due to the sensor miscalibration in the 1980 imagery.)

Optional:
Repeat steps a through o for several other vegetation indices and compare the results. How much does the choice of vegetation index influence the final assessment of change?

**Answers to the Questions in the Text**

1. Answers from this visual analysis will vary. In both images, the drainage patterns show up quite clearly as having higher vegetation index values, as one would expect. The differences between the two images are more subtle. If you reduce the upper display saturation points for both images (in Layer Properties), it appears that the NDVI image has relatively more higher-value pixels in the low vegetation areas and the ratio image has relatively more lower-value pixels in the low vegetation areas. The normalization of the NDVI serves to minimize topographic effects and division-by-zero errors.

2. The equation for NDVI is: \((\text{IR} - \text{Red}) / (\text{IR} + \text{Red})\)

   If we assume that vegetation has higher reflectance in the infrared than in the red, then the NDVI for vegetation would always be positive. Therefore, our identification of bare soil pixels might be NDVI values less than or equal to 0. Because of the "just less than" wording of the RECLASS module, you might assign the new value 1 to values from -1 to those just less than 0.000001 and the new value 0 from 0.000001 to just less than 1. This would include the value 0 in the bare soil category.

3. When red is independent and infrared is dependent: \(Y = -1.01 + 1.00X\), \(r^2 = 98.41\). When infrared is independent and red is dependent: \(Y = 2.36 + 0.98X\), \(r^2 = 98.41\). The first number in the equation is the intercept and the second is the slope (numbers shown are rounded to two decimal places). Your equations will be slightly different if you used different reclassification criteria for the mask image than those given in question 2.

4. These three images are very different. The PVI3 image, in particular, identifies large contiguous areas of relatively higher vegetation that are not identified in the other two images (nor in the slope-based images). More of the low vegetation areas have relatively lower values with the PVI image than with the WDVI images.

5. PVI and WDVI could be distinguished from Ratio and NDVI because the former have relatively lower values in the low vegetation areas. However, the PVI3 image would not easily be grouped with either "family."

6. The areas with much vegetation are darker than areas with less vegetation because vegetation is darker than bare soil.

7. The loadings for the VICMPI image are highest on the ratio, NDVI, PVI and WDVI images. The lowest loading is for PVI3, which verifies the PVI3 as being rather dissimilar from all the others. The highest loading for the second component image is for PVI3.

8. The answer here will depend on the index chosen, but it should appear that the area, overall, has more vegetation in 1990 than in 1980. However, the striping of the 1980 image makes this difficult to assess.

9. The areas that show the greatest increase in vegetation from 1980 to 1990 seem to be those along the drainage system. Much of what is classified as negative change appears to be attributable to the striping problems in the 1980 image.