

Analysis and estimation of deforestation using satellite imagery and GIS

Mesgari Saadi, and Ranjbar Abolfazl,

Geodesy and Geomatics Dept., Faculty of Civil Engineering,

K.N. Toosi University of Technology,

1346 Vali Asr Avenue, Postcode 19697, Tehran, Iran.

Tel. and Fax: +98-21-8786215, 8786216

Emails: mesgari@jtk-ac.org and ranjbar57@yahoo.com

1. Introduction

Forests amongst other natural resources have been degraded during the last decades continuously. The following factors are the main causes for such degradation:

- Changing of landuse from forest into pasture, agriculture and urban, as a result of population growth and general land scarcity,
- Cutting down of forests for timber production and wood industries,
- Use of the wood as a source of heat and energy in economically poor area,
- General degradation of forests caused by industrial growth, environmental pollution, increase in fuel consumption and global warming.

It is obvious that the most important factor in the destruction of forests is the human activities. For a better and sustainable management of such resources we need to know:

- The amount and location of deforestation,
- The rate and speed of deforestation, and
- Reasons and causes of deforestation,

The science and technologies of GIS and remote sensing could be a perfect tool for answering the above questions. Remote sensing can be the basis of fast and inexpensive data collection and the analytical capabilities of a GIS can be used for analyzing the types, location and rates of changes. The final aim in this research is to study and estimate the changes (damages) in the forest areas of Arasbaran, north-west of Iran and to evaluate the importance of different factors in this phenomenon.

2. Study area

The study area, called Arasbaran, is a mountainous area with elevation between 300 and 2700 meter above the sea level. It is in the north of Azarbaijan province and very near to the Caspian sea. The area is located between 38°40' and 39°09' latitude and between 46°42' and 47°03' longitude. It covers a diversity of elevation, slope, population and landuse and includes a variety of sea shore, rivers, etc. Beside the undamaged natural environment in some parts, a big part of the area has been changed by agriculture and grazing activities. This includes the thinly scattered woods, pastures, about 66 villages and differently cultivated areas.

3. The data set used

The satellite images used in this study are a Landsat TM image of 1987 and a Landsat ETM+ image of 2001 with a general resolution of about 28.5 meters. The old 1:50000 topographic maps of the Army's Geographic Organization and the new 1:25000 digital topographic maps of the National Cartographic Center of Iran have been used for georeferencing of the two images. The contour lines of the topographic maps are used for the generation of three maps. These maps represent values of elevation, slope and aspect in the area. Moreover, the location of the villages are extracted and used for generating the map of distance from the population centers.

4. Preprocessing and analysis of the satellite images

Usually three types of errors occur when a satellite image is generated by the satellite sensor. The first is the sensor error. The second is the error created by the atmospheric parameters, which affect the amount of radiation received by the sensor. The third one is the geometric errors related to the curvature of the Earth surface, the Earth rotation, elevation differences, location and situation of the satellite etc. Therefore, These errors should be considered and managed before using the data:

a) Sensor errors

The two images used were already corrected by their providers. Therefore, there was no need for any processing in this regard.

b) Radiometric correction

The Earth atmosphere scatters the shorter wavelengths in a selective manner and this reduces the contrast of the image. The numerical value of each pixel in the image is not a realistic representation of the amount of radiation from the ground surface. These values are changed either by atmospheric absorption or by scattering throughout the atmosphere. In general, atmospheric errors are discussed in three parts: the Haze, Sunangle and Skylight errors. Atmospheric corrections are required in the following situations:

- When we want to compare the images related to different times: When using methods such as image subtraction and image division for change detection, the effect of atmosphere on the two images related to different times are quite different.
- When the ratio of two bands of an image is needed to be calculated, because the atmosphere has different effects on different wavelengths.
- When we want to study spectral characteristics of different phenomena.

If we wanted to use the division or subtraction of images for determining the changes in forest landuse, then we would have to correct for the haze, sunangle and skylight errors. In our approach we compare the results of the landuse classification maps extracted from the two images. The classification of landuse can be done better and more accurate with the raw (unprocessed) images. Therefore, there was not any need for the above corrections in our images.

c) Geometric corrections

The process and analysis of multi-temporal data can be done only when they are geo-referenced similarly, or in another words, when they are geo-referenced to each other. Our images had to be geo-referenced to each other with an accuracy of one pixel. Otherwise, the error coming from different coordinates for similar objects in the two images can be wrongly accepted as a landuse change. In other words, with an inaccurate geo-referencing, a pixel might refer to different objects in the two images and be considered as a landuse or land cover change, which is not realistic.

To prevent such a problem, in comparison of multi-temporal images, the best solution is to geo-reference one of the images using the available topographic maps and then geo-referencing the other images according to the first one, i.e. using image-to-image registration.

In photo/image registration (geo-referencing), the most important task is the proper selection of control points, especially when there is a long time period between the map and the image. Usually, man-made features such as buildings and road intersections are a better choice for control points than the natural ones. The reasons are that they have sharper boundaries and more contrast with their surrounding. Besides, they are geometrically more stable than features such as river/stream junctions. The general rules are that we should try to select more stable features that have longer change periods and the more control points we select the more accurate our registration will be.

We simply used the first order polynomial equations for geo-referencing of the images, which remove the errors related to the rotation and scaling of the image. These are:

$$X = a_0 + a_1x + a_2y$$

$$Y = b_0 + b_1x + b_2y$$

where x and y are the coordinates of a point in the first coordinate system and X and Y are its new coordinates in the new coordinate system.

In this study, the ETM+ image of the year 2001 was first geo-referenced using the information in its header approximately. Then, it was geo-referenced accurately using the available 1:25000 digital maps and the digitized features of the 1:50000 maps of the area. The control points were selected using different color composites with band-combinations of 754, 432 and 543. Afterward, the TM image of 1987 was geo-referenced using the already registered TM image.

For geo-referencing the 2001 image 18 control points were used initially. Every control point with an RMSE or residual error bigger than a pixel size was removed from the calculation and the process of registration was repeated with the rest of the control points. Finally, 10 points with the average error of 16.47 meters remained and were used for registration. For image-to-image registration of the 1987 image 20 control points were initially used. Finally, 6 points were removed and the image was geo-referenced using the remained 14 points with the RMSE of 18.92 meters.

In view of the fact that our images are used for landuse classification, any change to the numeric value of the pixels will introduce some errors and has an undesirable effect on our classification. Therefore, to minimize this effect during geometric correction, the new values of pixels were generated using the nearest neighbor method.

5. Assessment of deforestation by comparison of the classification maps

One of the methods for change detection using satellite images is to compare the results of classification of the images. Two other methods are to calculate the division or subtraction of the two images. The main problem of these methods is that they can only define where some changes are happened. The advantage of the classified-map comparison method in to the other methods is that not only the location but also the nature and type of the changes will be determined. In other words we will define what landuse has been changed to what other.

In this method, first, the images of different times are classified according to the purpose of change detection. Afterward, by overlaying the two classified images with a proper overlay condition, we can determine the location and amount of any changes we are interested in. Because our goal was to determine the deforestation, the only two classes that we considered are the forest and non-forest. The two images are classified using the Maximum-Likelihood method.

By overlaying the results of classifications, the map of the occurred changes are resulted, as is shown in Figure 1. From this map, it can be realized how much of the forest have been damaged and where this has happened. In addition, the pattern and spatial distribution of the phenomenon is properly illustrated. Furthermore, it can be seen where the forest and non-forest classes have been stable and where new forest has been growing.

6. Creation of the logistic regression model

A regression model is a statistical model in which a relation between a phenomenon (a dependent variable) and some of its factors (some independent variables) will be defined based on some observations. These observations are in fact a set of values measured or observed for the dependent and independent variables. Having the model specified and calibrated, the unknown value of the phenomenon can be calculated and predicted on the basis of known values of its factors.

Logistic regression models, a special type of regression models, are used when we want to study the probability of membership in two contradictory classes, such as a forest area being either stable or destroyed. It should be noted that logistic regression can be used to determine the probability of any of the two possibilities (classes) identically. A logistic regression model is usually of the type:

$$\log \text{it}(p) = a + b_1x_1 + b_2x_2 + b_3x_3, \quad \text{where } \log \text{it}(p) = \ln\left(\frac{p}{1-p}\right)$$

Here, 'p' is the dependent variable and shows the probability of one of the two conditions. Dependent variables of 'x₁', 'x₂' and 'x₃' represent the factors defining the phenomenon and 'b₁', 'b₂' and 'b₃' are their coefficients: 'a' is the additive coefficient.

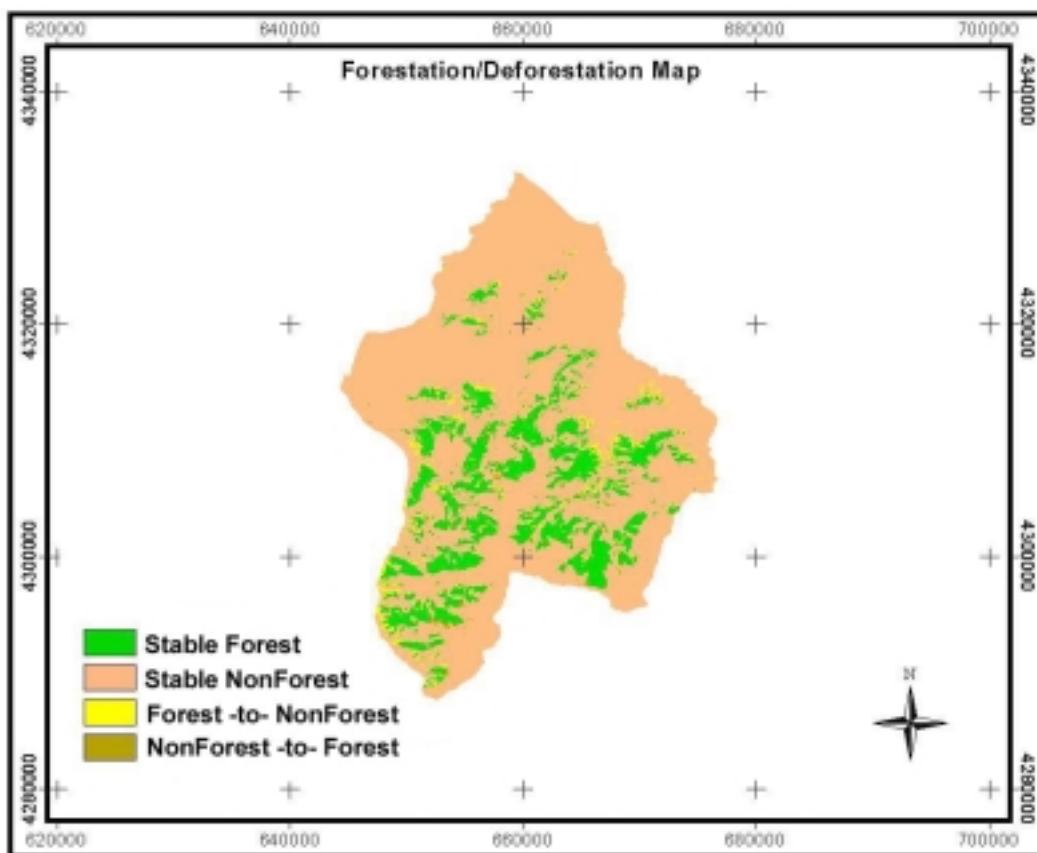


Figure 1 Map of landuse changes (forestation and deforestation) in the study area

In this research, we selected a sampling set of about 5% of the pixels, from the two classes of stable forest and destroyed forests, i.e. forests that have remained forest and forests that have been changed or destroyed. The number of sample pixels is 5106 pixels in total. In these sample pixels, the parameters of elevation, slope, aspect and distance from villages are considered as independent variables and the stability of the forest as the dependent variable. As mentioned, we could use the forest destruction (deforestation) as the dependent variable and get exactly the same results. Independent variables are extracted from the relevant generated maps. The dependent variable, i.e. the forest stability, is represented by the two values of '0' and '1' for the sample pixels. '0' represents the deforested areas and '1' represents the stable forests.

By introducing the sample data to the specified logistic regression model, in the first stage, the variable of distance from population centers entered to the model, improving the χ^2 parameter to the value 601.641. This parameter (called square of 'chi') is a measure for the goodness of the model; a low value for χ^2 means the model is suitable to the data.

The effectiveness of the model in prediction of the phenomenon can be summarized in a simple table (Table 1). From the total 5106 sample pixels, 1256 pixel were changed (deforested) and 3850 pixels were unchanged forest. In every stage of the regression, all pixels are evaluated by the model and a value is predicted for each pixel. The number of both correctly and wrongly predicted pixels for each stage is shown in Table 1. In other words, in this table, the groups of deforested and unchanged forest pixels are compared with what is predicted for them by the model. From the 2nd and 3rd column of the table is clear that 12.18% of the changed (deforested) pixels and 92.47% of unchanged pixels are predicted correctly by the model. This means a total prediction-accuracy of 72.72%.

In the second stage, the elevation variable entered the model and changed the χ^2 parameter to the value 272.826. Having the two parameters of elevation and distance from villages together in the model, 18.55% of the changed (deforested) pixels and 93.82% of unchanged pixels were predicted correctly by the model. This stage showed a total accuracy of 75.30% in the prediction of pixels (Table 1).

In the third stage, the aspect variable entered the model and caused a significant improvement in the χ^2 parameter, changing it into 92.681. The three parameters of elevation, distance from villages and aspect (aspect of slope) being incorporated in the model, 31.93% of the changed pixels and 93.90% of unchanged pixels were predicted correctly by the model. This resulted in a total accuracy of 78.65% in the prediction of pixels, as can be extracted from the last two columns of Table 1.

Table 1 Prediction accuracy of the model in different stages of introducing variables

	Predicted as changed (1st stage)	Predicted as unchanged (1st stage)	Predicted as changed (2nd stage)	Predicted as unchanged (2nd stage)	Predicted as changed (3rd stage)	Predicted as unchanged (3rd stage)
Changed Pixels	153	1103	233	1023	401	855
Unchanged pixels	290	3560	238	3612	235	3615

After the third stage, no other independent variable could enter the model. This means that the only remained variable, i.e. slope, could not cause any significant improvement to the performance of the model and its suitability with the data. This can be either because of the irrelevance of this variable to the phenomenon, or because of its high correlation with other variables incorporated in the model.

Table 2 shows the results of the calibration of the model. The model represents the phenomenon of forest-stability (as opposed to deforestation) on the basis of three factors of distance from population centers, elevation and aspect. The coefficients of these factors in the resulted model are presented in Table 2.

Table 2 Coefficients of variables in the resulted regression model

Factor (variable) name	Coefficient in the logistic regression model
Distance from population centers	0.0010
Elevation	0.0068
Aspect	0.0027
Constant value	- 9.8675

7. Conclusions and future work

The following remarks and recommendations can be concluded from this study:

- In analysis and comparison of images of different times, the error (accuracy) of geo-referencing for the images should be smaller than a pixel dimension. Otherwise, the difference in the geometry and location of any feature in the two images will result in the acceptance of an unrealistic change.
- In this research, we first needed to classify the satellite images. The changes introduced to the pixel-values, by interpolation and during geometric correction, have an undesirable effect on the result of classification. To moderate this effect, the new values of pixels can be generated using the nearest neighbor method.
- As was expected before the research, by moving away from the population centers the stability of forests increase. This is because, in the vicinity of villages, the forests are cut down mainly with the intention of using the land for agriculture and grazing and using the wood as fuel.
- In areas with higher elevation forest is more stable. This is partly because in high areas the environment in general is cleaner and more intact. The higher the area, the less suitable it is for agriculture and the more difficult it is to go.
- The slope aspect has an important role in deforestation in this area. The inclines and hills toward south get more sunlight and therefore are more suitable for agriculture. On the other hand, the east and north directions are enjoying the humidity coming from the Caspian Sea.
- From the beginning, we were aware of the possible dependency among the introduced factors, but the types of dependencies were not clear. It was proved that the slope has a high correlation with other factors, mainly with distance from villages and elevation. Therefore, the importance of comprehensive studies about factors affecting a phenomenon before modeling it has been proved.
- There are many other factors that might be relevant to deforestation and are not covered in this study. Examples of such factors are soil type, distance from the roads and population of the villages.
- In studies related to land use change between years, effort should be made to use the images of the same season and even the same month. When images are from the same month or season, the changes detected from them are more realistic and reliable.

Bibliography

Macleod D. and G. Congalton (1998): *A quantitative comparison of change detection algorithms for monitoring eelgrass from remotely sensed data*, Photogrammetric Engineering and Remote Sensing, Vol. 64, No. 3, pp. 207-216.

Sader A., and C. Winne (1992): *RGB-NDVI color composites for visualizing forest change dynamics*, International Journal of Remote Sensing, No. 13, pp. 3055-3067.

Sermongkontip S., Y. Ali Hussin, and L. Groenindijk (2000): *Detecting changes in the Mangrove forests of southern Thailand using remotely sensed data and GIS*, ISPRS XXXIII, Part B7, pp. 567- 574.

Tabachnick, B. G., and L. S. Fidell (1996): *Using multivariate statistics*, New York, Harper and Row.