Caspian J. Env. Sci. 2009, Vol. 7 No.1 pp. 1~7

CJES
Caspian Journal of Environmental Sciences

[Research]

A model-based approach for mapping rangelands covers using Landsat TM image data

M. Ajorlo[1*], A. Ramdzani Bin [1], H. Ahmad Husni Mohd[2], H. Ridzwan Abd[2], Y. Mohd Kamil[1]

1- Faculty of Environmental Studies, University Putra Malaysia, 43400 (UPM), Serdang, Selangor D.E., Malaysia.
2- Faculty of agriculture, University Putra Malaysia, 43400 (UPM), Serdang, Selangor D.E., Malaysia.
**Corresponding author's E-mail: Ajorlo_m54@yahoo.com*

ABSTRACT

Empirical models are important tools for relating field-measured biophysical variables to remotely sensed data. Regression analysis has been a popular empirical method of linking these two types of data to estimate variables such as biomass, percent vegetation canopy cover, and bare soil. This study was conducted in a semi-arid rangeland ecosystem of Qazvin province, Iran. This paper presents the development of a regression model for predicting rangeland biophysical variables using the original image data of Landsat TM nonthermal bands. The biophysical variables of interest within the rangeland ecosystem were percent vegetation canopy cover, bare soil extent, and stone and gravel which their correlations were analyzed in relation to Landsat TM original data. The results of applying stepwise multiple regression showed that there is a significant correlation between Landsat TM band 2 reflectance values and biophysical variables. The developed models were applied to Landsat TM band 2 and relevant maps were generated. We concluded that such problems as an inexact location of field samples on the image, small size of samples, vegetation heterogeneity may significantly affect the modeling of real rangeland Landsat TM data relationships.

Keywords: Biophysical variables, Empirical model, Multiple regression, Rangeland, Remotely-sensed data.

INTRODUCTION

Total area of Iran's rangelands is about 90 million hectares which cover 52 percent of the country (Technical Bureau of Rangeland, 2000). The total livestock population in rangelands of Iran has reached 125 million based on Animal Unit (AU). While the annual forage dry matter production of rangelands is estimated at more than 10 million tons (Fazilati and Hosseini, 1984). It could provide sufficient forage for only 25 million AU. Therefore, there are 100 million extra hunger livestock in rangelands of Iran. Thus, rangeland degradation is a major significant environmental problem in huge arid and semi-arid areas of Iran. However, little is known of its extent, severity and causative factors (Ajorlo, 2005). In order to establish sustainable rangeland use system, it is important to make attempts to discover the area prone to degradation as well as to monitor the progress of degradation. However, quantitatively reliable data about the extension of rangeland degradation in Iran has not yet been obtained (Ajorlo and Abdullah, 2007). Therefore, it is very necessary to use a precise, repeatable, inexpensive technique for the management of Iran's rangelands.

Remotely sensed data are very useful tool for this purpose. In recent years, many studies have been done on the applicability of remote sensing to acquire rangeland information. Remote sensing data has been expected to provide quantitative information about land cover. This technique has now become the single most effective method for land-cover and land-use data acquisition (Lillesand and Kiefer, 1994). Landsat TM image data consists of seven spectral bands with a spatial resolution of 30 meters for 1, 2, 3, 4, 5 and 7 bands. Spatial resolution of band 6 is 120m. Although it can be argued that spectral response is dependent on vegetation condition and not on the other way around, much of the remote sensing literature reports the vegetation attribute being modeled as the dependent variable (Cohen *et a*l., 2003).

When dealing with rangeland information given by field-measured quantitative variables in combination with remote-sensing imagery, many different types of analysis may be applied (Salvador and Pons, 1998). The value of regression analysis for modeling the relationship between vegetation variables and spectral reflectance value is well established (Guo *et al*., 2000; Danaher *et al*., 2004). Cohen *et al*. (2003) stated that we should be expanding our use of multiple regression over simple regression techniques. Regression models are used to estimate one variable from one or more other variables. Multiple regression is a common technique for estimating sub-pixel cover fractions in satellite imagery, however application is often limited by a lack of field data, and radiometric, spatial and spectral uncertainties of remotely sensed imagery (Danaher *et al*., 2004). Therefore, some restrictions of TM satellite data such as its radiometric, spectral, and spatial limitations, together with restrictions arising from gathering and processing of field data, might have led to poor relations between estimated and observed values in generated models by multiple regression (Salvador and Pons, 1998). Fitzpatrick and Megan (1994) have used regression analysis to find relationship between ground data and satellite image data based on vegetation cover and bare soil. They noted that when we use regression analysis to get a relationship, sampling must be included all range of vegetation cover changes (sparse to dense). They stated that there is a relationship between vegetation cover and Landsat TM data; but regression analysis can not be used to confirm this relationship.

This study is mainly focused on multiple regression analysis since this is a technique that is one of the most commonly used methods applied to estimate rangeland quantitative variables from remotely-sensed data. The objective of this study was to evaluate the relationship between spectral reflectance of Landsat Thematic Mapper (TM) raw data from single date image and rangeland biophysical variables to find the most correlated TM band to rangeland cover in semi-arid areas of Iran.

MATERIALS AND METHODS

Study area

Qazvin Province in northern Iran is characterized by a semi-arid climate with cold winter, dry and warm summer. This study was conducted in semi-arid rangelands of Kolanjin watershed, 120 Km west of Qazvin city. The watershed lies between latitudes 35° 24' 16" and 35° 38' 26" N and longitudes 49° 24' 48" and 49° 31' 48" E. Total area of study watershed is 17654 ha. This watershed is situated at about 2211 m mean elevation. Average annual precipitation is 345 mm over the past 30 years, with an average annual temperature of 10 °C. The area comprises communal rangelands (13 311.8 ha), woodlands, farmlands, gardens (4 261.4 ha) and inhabited areas (81 ha).

Data sources

Landsat TM image with a spatial resolution of 30-m covering the study watershed was applied from Iranian Remote Sensing Center (IRSC) on July 1998 (IRSC, 1998). The path and row of image in Landsat World Reference System (WRS) are 166 and 35, respectively. Topographic maps of the study area in the scale of 1/50000 were obtained from National Cartographic Center (NCC) (NCC, 1995).

Field sampling design and measurement

Field measurements of land cover attributes were taken within a month of the satellite image data acquisition. Vegetation cover, gravel and stone, and bare soil extent as biophysical variables were measured during field survey around the sampling locateons; rural village and stock watering point. Data collection took place in seven field sampling locations; four rural villages and three livestock watering points (Figure 4). Each rural village and stock watering point entailed twelve 200-m transects spaced at 100-m apart and eight 200-m spaced at 50-m apart, respectively. A 3 m² quadrat was put on each transect spaced at 40-m for measuring rangeland surface factors. Thus, six quadrats on each transect were applied for measuring biophysical factors. Quadrats on transects were arranged in a manner that to have one quadrat in each pixel of the image. In general, 414 observation points (quadrats) were recorded in the field work to assist in the development of accurate relationship between field and remotely sensed data. The location of transects was determined using a differential GPS.The areas around sampling locations include a range of land surface conditions ; very sparse to dense vegetation canopy cover, high and low bare soil extent, high degraded area, and low degraded area. These land surface attributes are seen across the rangelands. Hence, sampling around the villages and stock watering points will produce a good representative data set for rangelands surface attributes in the study area (Fitzpatrick and Megan, 1994).

Data Analysis

Image pre-processing

The image was geometrically corrected to have georeferenced coordinates. A polynomial-based model which taking relief effects into account was applied. A root mean square (RMS) value lower than one pixel was achieved. Corrections for atmospheric effects, and illumination effects related to relief, were also applied. Once the image was corrected, the locations of transects were determined on the image using recorded coordinates in the field. Ground control points (GCP) and distinctive landmarks including waterways, trails and roads were used to validate the location of transects. Then, reflectances values of pixels located over the coordinates of field transect were extracted from all six TM bands. No more processing was carried out on the image.

Model development

Stepwise multiple regression approach was applied to develop the most appropriate statistical models for calculation of rangeland cover based on ground based data and Landsat TM image data. Field variables including vegetation cover, stone and gravel, and bare soil extent were utilized as dependent variables. Whereas reflectance values of original Landsat TM bands (excluding band 6) were used together as independent variables for a more in-depth analysis of their correlation with rangeland cover attributes. Then, in the given dataset independent variables which had high significant correlation with dependent variables (field data) were distinguished. Eventually, separate regression models were developed for all variables.

Generating rangeland covers map

Field sampling was conducted on rangelands of the study watershed. In other words, other land uses including farmland, woodland were discriminated on the image in order to evaluate only rangeland cover attributes. Then, regression models were applied to produce vegetation cover, stone and gravel, and bare soil extent maps of rangelands.

MODEL VALIDATION

Independent validation is a critical component of any remotely sensed project; however, acquisition of independent field data to validate estimates of field indicators is prohibitively expensive (Danaher *et al*., 2004). Transect values which have already applied in model development could not be used in model validation test (Cohen *et al*., 2003). To verify the reliability of selected models in estimation of rangeland surface attributes, some tests were carried out with extra transects. To this end, 20 new transects were established on the rangeland (Figure 4). Validation of the model was done using values of rangeland surface attributes on the new transects.

Each model was used to compare variables predicted using model with biophysical variables derived from field measurement. This comparison revealed that there is low correlations between observed bare soil extent, stone and gravel, vegetation canopy cover and estimated data using Landsat band 2 models ($r = 0.43$, $P>0.05$; $r = 0.21$, $p>0.05$; $r = 0.17$, $P>0.05$, respectively).

Preliminary validation of the models in this study suggests that multiple regression is a non-robust technique for finding relationship between Landsat TM data and rangeland biophysical variables. Results of validation trials revealed that the models have non- robust nature to predict biophysical variables using them. Non-robust models are models with very low predictive capabilities.

RESULTS

The results of applying stepwise multiple regression showed that there is a significant correlation between Landsat TM band 2 reflectance values and field data. The criteria used for model development included the adjusted R^2, standard error, and the P value for individual model terms. The model with the highest value of adjusted R^2 and low bias of estimation was the better one to be selected. It should be noted that always R^2 reaches the highest value when the maximum number of independent variables is used, regardless of the remaining degrees of freedom, whereas the adjusted R^2 takes into account the decrease in the degrees of freedom (Rawlings, 1988). As it was explained by Montgomery and Peck (1992), the model with the highest adjusted R^2 is also the model with the lowest value of the residual mean square (the residual variance of the model), which is defined as:

$$MSE=SSE \div (n - p) \quad (1)$$

Where SSE is the sum of the squared deviations between the observed and expected values of the dependent variable and n - p are the remaining degrees of freedom. The results for the selected models are sumarized in Table 1. There were no significant correlation between other bands and field data. Moreover, models selected for their adjusted R^2 seemed to be supported by a very high statistical significance (given by very low values of F test). Regression models were produced according to the output of regression analysis (Table 2). The models were applied only on rangelands in Landsat TM band-2 and relevant maps were generated for vegetation cover, gravel and stone, and bare soil extent (Figures 1, 2, 3).

Table 1. Results of multiple regression analyses with six Landsat TM bands.

Ground cover parameter	Landsat TM band 2			
	R	R Square	Adjusted R Square	P
Vegetation cover	0.87	0.76	0.69	$P<0.01$
Stone and gravel	-0.83	0.67	0.61	$P<0.01$
Bare soil extent	-0.92	0.85	0.80	$P<0.001$

Table 2. Regression model statistics for each model type and dataset.

Model	Slope	Intercept	R^2
Vegetation	1.93	124	0.76
Bare soil	-1.60	-51.98	0.67
Stone and gravel	-0.96	-10.30	0.85

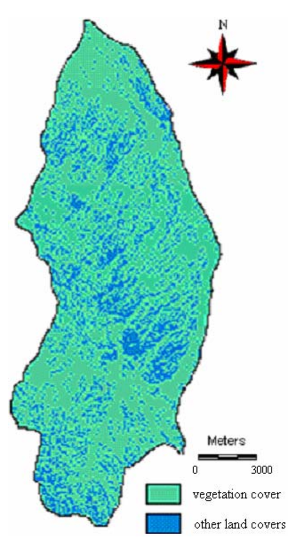

Fig 1. Vegetation distribution map after applying regression model on original band 2.

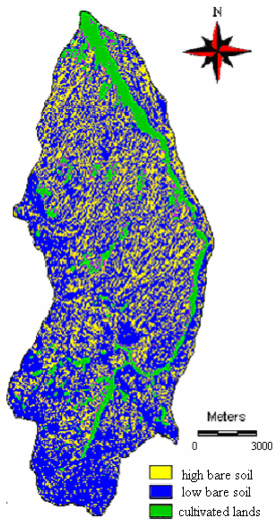

Fig 2. Bare soil extent classification map after applying regression model on original band 2.

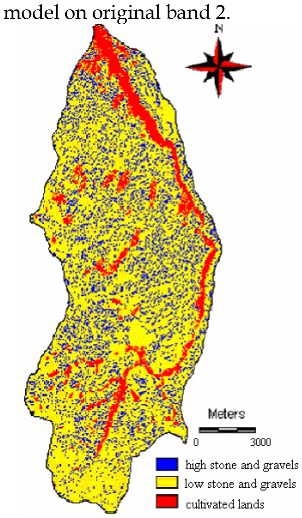

Fig 3. Stone and gravels extent map after applying regression model on original band 2.

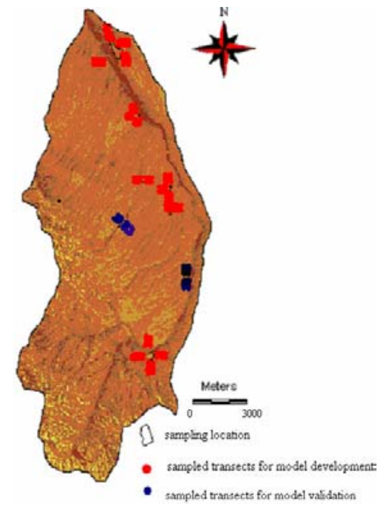

Fig 4. Location of sampled transects in study watershed.

DISCUSSION

The results of the study state that band 2 of Landsat TM contains the highest information on rangeland cover attribute, when multiple regression is used. Rahman *et al*. (2005) found when regression is applied, band 2 of Landsat Enhanced Thematic Mapper Plus (ETM+) contains the highest information on forest biomass. In this study, vegetation indices were not considered. Modelling was done only based on the original band values and field variables using multiple regression. Danaher *et al*. (2004) found that the models based on transformed band values produced similar results and all were better than those based on the original band values. Salvador and Pons (1998) stated that in comparison to simple regression models, reliable results were obtained with multiple regressions using all satellite bands together as independent variables. Danaher *et al*. (2004) observed that multiple regression is a robust technique for mapping woody foliage projective cover (FPC) using landsat imagery in Queensland, Australia. It is also reported that the regression model is performing well in areas with varying soil color and fire-scar backgrounds, and that image stratification is not necessary as the regression model appears to compensate for these factors.

Although models were created with significantly high R^2 and significance level for all variables in this study, validation trials showed that models were non- robust in estimating of rangeland biophysical variables. Salvador and Pons (1998) in a study on the reliability of Landsat TM to estimate forest variables by regression techniques found that the best created models with good fittings (R^2> 0.65) and a statistical significance ($p < 0.0001$) were non- robust models when validation trials were carried out with additional plots. Salvador and Pons (1998) stated two factors as the causes of these inconsistencies between predicted and observed values; a relatively small number of available field plots and a relatively high number of possible independent variables. it is also reported that although results of multiple regressions with original bands were better than simple regression results, but they were not good enough to make quantitative predictions. Furthermore from various studies, different spectral regions as the strongest predictor of field variables are reported when regression is used. Therefore, it is difficult to find a consensus about the effectiveness of models developed by using regression analysis and Landsat TM imagery in the results of different researchers across the world. The results are different according to the regression analysis, environmental condition, sampling method, sample size and so on. It is the matter of further studies to comprehend why it happens.

Taking the results of this study into consideration, we observed some problems such as small size of samples, low number of samples, and sampling on the specific areas are the main reasons of producing non-robust models. Fitzpatrick and Megan (1994) noted that when regression analysis is used to find a reliable relationship, sampling must be included all range of vegetation cover changes (sparse to dense). According to the study, it is also reported that there is a relationship between vegetation cover and Landsat TM data; but regression analysis cannot be used to confirm this relationship. Although field data are important in all remotely sensed projects, but accomplishment of field sampling has some serious restrictions. Field sampling are based on a limited discrete sampling over a continuous spatial dimension. This leads to an unsighted extrapolation when information concerning unsampled areas is requested. Models were based on a relatively low number of plots which were perhaps not sufficient to characterize all environmental conditions. In addition, some restrictions of Landsat TM data, such as its radiometric, spectral, and spatial limitations, together with restrictions arising from gathering and processing of field data, might have led to such relationnships.

In the present study, non-robustness of the models for estimating rangeland biophysical variables in preliminary validation may be attributable to the date of field sampling and low number of plots which were applied for validation. Normally date of field sampling must be coincide to satellite image date. In this study, however, field sampling for model validation was carried out two months later than image date. Another reason for such non- robust models may be attributable to plot size (3 m^2) which was used for gathering field data. Essentially, plot size should be near to pixel size in such studies. In this study, plot size for measuring rangeland cover attributes was determined based on rangeland vegetation measurement rules which proposed size was very smaller than satellite image pixel size.

ACKNOWLEDGMENT

We wish to thank the Iranian Remote Sensing Center (IRSC) for providing us with the satellite imagery.

REFERENCES

Ajorlo, M. and Abdullah, B.R. (2007) Develop an Appropriate Vegetation Index for Assessing Rangeland Degradation in Semi-Arid Areas. In: proceedings of 28th Asian Conference on Remote Sensing. 12-16 Nov, Kuala Lumpur, Malaysia.

Ajorlo, M. (2005) Evaluation and mapping of rangeland degradation using remotely sensed data. *In*: proceedings of international symposium on land degradation and desertification, 12-17 May, Uberlandia, Brazil.

Cohen, B.W., Thomas K.M., Gower S.T. and Turner P.D. (2003) An improved strategy for regression of biophysical Variables and Landsat ETM+ data. *Rem Sens Environ*. **84,** 561-571.

Danaher, T., Armston J. and Collett L. (2004) A regression model approach for mapping woody foliage projective cover using Landsat Imagery in Queensland, Australia. *In*: Proceedings of Geoscience and Remote Sensing Symposium, Australia. pp. 514 – 527.

Fazilati, A. and Hosseini E.H. (1984) Rangelands of Iran and their management, development and improvement. Technical Bureau of Rangeland. Tehran, Iran.

Fitzpatrick, B. and Megan, A. (1994) Relationship between vegetation cover field data and Landsat-TM in the pasture development areas of the Douglas-Daily Basin, Northern territory. *In*: 7th Australian Remote Sensing Conference Proceedings, Melbourne, Australia.

Guo, X., Price K.P. and Stiles J.M. (2000) Modeling Biophysical Factors for Grasslands in Eastern Kansas Using Landsat TM Data. *Trans Kans Acad Sci*. **103**(3/4), 122-138.

Iranian Remote Sensing Center (IRSC) (1998) Landsat 5 TM 10 digital images. Tehran.

Lillesand, T. M. and Kiefer, R.W. (1994) Remote Sensing and Image Interpretation (3rd Ed). John Wiley & Sons Inc.

Montgomery, D.C. and Peck, E.A. (1992) Introduction to Linear Regression Analysis. New York: Wiley, pp. 270–274.

National Cartographic Center (NCC). (1995) Topographic maps (scale 1:50000). Sheets: Karafs 5860 I, Asian 5861 II, Razak 5961 III, Saman 5960 VI. K753 Series. Tehran.

Rahman, M. M., Csaplovics E., and Koch B. (2005) An efficient regression strategy for extracting forest biomass information from satellite sensor data. Int. J. *Remote Sens*. **26**(7), 1511 – 1519.

Rawlings, J.O. (1998) Applied Regression Analysis (a Research Tool). Wadsworth and Brooks. pp. 183–184.

Salvador, R. and Pons, X. (1998) On the reliability of Landsat TM for estimating forest variables by regression techniques: a methodological analysis. *IEEE Trans Geosci Rem Sens*. **36**(6), 1888-1897.

(Received: May 12-2008, Accepted: Jun. 19-2009)