Jour. Agric. Physics, Vol. 8, pp. 29-36 (2008)

Irrigation Induced Soil Salinity Mapping through Principal Component Analysis of Remote Sensing Data

S.K. PATTANAAIK[1], O.P. SINGH[2], R.N. SAHOO[3] AND D.K. SINGH[4]

[1]*College of Horticulture & Forestry (C.A.U.), Pasighat, 791 102, Arunachal Pradesh*
[2]*Division of Agricultural Engineering,* [3]*Division of Agriculture Physics,* [4]*Water Technology Centre, Indian Agricultural Research Institute, New Delhi - 110 012*

ABSTRACT

This paper is based on a research carried out to develop a methodology to delineate severe salt-affected soils using principal component images. For this, fourteen multispectral and multitemporal satellite images were procured for the irrigated areas of Western Yamuna canal command covering Gohana sub-division of Sonepat district of Haryana state. All the data were geometrically and radiometrically corrected and then, resampled to the same spatial resolution. Subsequently, all these data were subjected to principal component analysis. It was observed from the analysis of the Red, Green and Blue (RGB) colour composite, generated from PC1 of NIR is displayed in red and PC1 of Red and PC1 of Green are displayed in green and blue respectively; while the dark tones showed clearly the severe salt-affected areas of an irrigation command.

Keywords: PCA, RGB colour composite, severe salt-affected areas

Introduction

The principal component transformation is a multivariate statistical technique that selects uncorrelated linear combinations (eigenvector weights or loadings) of variables in *n*-dimensional space in such a way that each successively extracted linear combination, principal component (PC), has a smaller variance (Souza Filho and Drury, 1998). This principal component analysis (PCA) has found its application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimensions. Extensive interband correlation is a problem frequently encountered in the analysis of multispectral image data. That is, digital data images generated by various wavelength bands often appear similar and convey essentially the same information. PCA technique is designed to reduce such redundancy in a set of multispectral data. To increase the efficiency of the image manipulation, methods like classification and the image interpretation, reflectance information from all different spectral band images of different satellites, are required to be collected together. But for visual presentation, all different spectral images cannot show together at the same time. Only three different spectral images can perform a colour image, since it is produced from three primary colours which are red, green and blue. Fortunately, there exists a method which can almost combine the total different spectral images into fewer images. This can be achieved by using the principal component analysis. PCA is the method applied for reducing the dimension of image data. So that all treated images with different spectral bands are transformed into a few principal components, which preserve almost the total variance of the original image. Nevertheless, the ground sample distance or pixel resolution of one sensor system will be different from another. Hence, an interpolation method is applied to resembling an image in order to obtain the same ground sample distance or pixel resolution. Principal component analysis (PCA) incorporates spectral information from a number of bands simultaneously, thus accessing the high sensitivity to clear water in the blue to green wavelength and the high sensitivity to wet soils in the red and near infrared wavelength. Sah *et al.* in 1995 developed a methodology to classify salt-affected soils using remote sensing and GIS. For this, they first used principal component analysis to create principal component images taking salt crust as a criterion, Landsat TM with bands other than two and six was found effective, to a greater extent, in classifying the extremely and moderately saline area. The study concluded that integration of GIS with digital image processing of TM was very effective in classifying and monitoring of saline soils. A study was undertaken to delineate waterlogged areas in the Tawa command of Sriram Sagar Project using IRS-1A-LISS-I data. The IRS-1A LISS II data have been proven very useful for the assessment of waterlogging. Density slicing and PCA are useful techniques to make an assessment of waterlogged areas in irrigation command in different seasons. An attempt was made to validate the IRS derived waterlogged area with the available data on depth of water table (Choubey *et al.*, 2000). Dwivedi (1996) used principal component analysis to delineate salt-affected soils of Uttar Pradesh by combining Landsat MSS data for the same ground area from February 1975 and March 1992. He states that PCA can be used to identify temporal change when digitally registered Landsat MSS data are merged and treated as a single data set. The first principal component accounted for 62% of the scene variance and the white patches of soil with salt efflorescence were identified with 98% accuracy. The second principal component accounted for 16.5% of the variance and emphasized the green vegetation. The studies reviewed in this section used a variety of time consuming analysis procedures to detect and map salt-affected soils. With this preamble, a study was conducted to map irrigation induced soil salinity in the Gohana sub-division of Sonepat district of Haryana state under the Western Yamuna Canal Command area by using more numbers of spectral and temporal satellite imageries.

Materials and Methods

Location

The study area lies between 76^0 41' to 76^0 47' E longitude and 29^0 1' to 29^0 7' N latitude, covered under Survey of India topo-sheet Nos. 53(C/12), 53(C/16), 53(D/9), and 53(D/13). It is located in the Western Yamuna Canal irrigation command covering Gohana sub-division of Sonepat district of Haryana State (Fig. 1). The

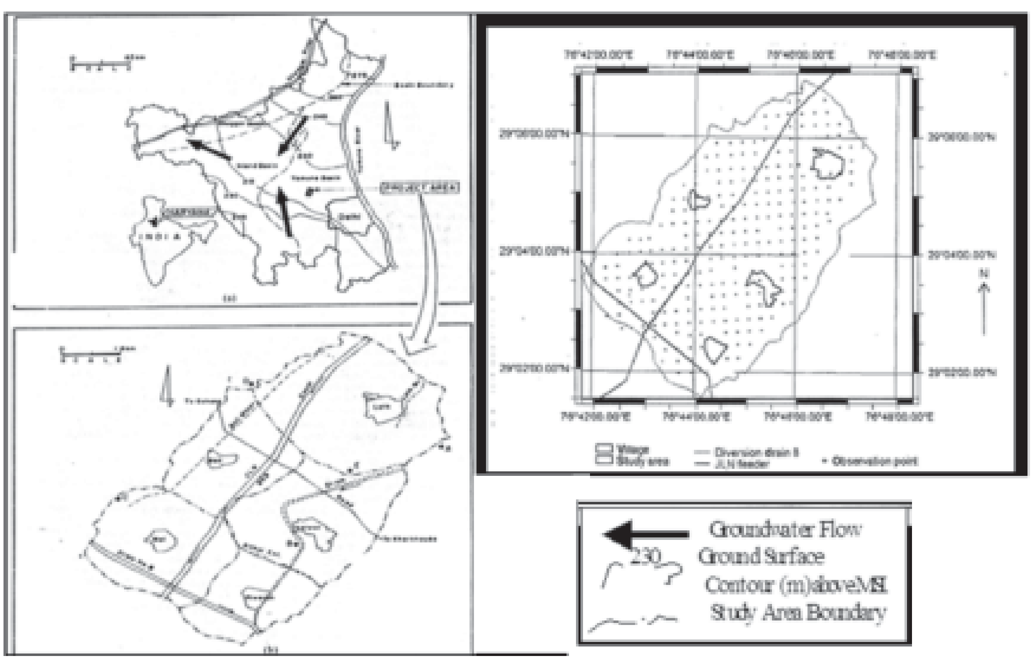

Fig. 1. Site map of the study area

5000 ha study area is spread over to five villages namely Moi, Rewara, Bali, Katwal, and Lath along both sides of the Jawaharlal Nehru (JLN) feeder and parallel to Bhalaut branch. It is bounded by diversion drain no. 8 in the SouthWest.

Physiography

The area has nearly flat topography with reduced levels ranging from 225.8m to 222.5m above mean sea level. The soil is predominantly alluvial with texture varying from sandy loam to silt loam. The area is faced with the problems of waterlogging and salinity.

Climate

The climate of the area is semi-arid. The average annual rainfall is 550mm, out of which 80 per cent occurs during the period from July to September. The average annual evapotranspiration is 1650mm. The maximum temperature varies from 1^0C in winter to 45^0C in summer.

Crops

The major *kharif* crops are paddy, sugarcane and sorghum and the major *rabi* crops are wheat, mustard, sunflower and sugarcane. The cropping sequence followed is Paddy-Wheat, Fodder-Wheat, Fallow-Wheat and Fallow-Sunflower.

Waterlogging and Salinity Problems

Canal water is the main source of irrigation in the study area, although groundwater through shallow cavity wells is also used to irrigate small areas. Due to canal seepage, water table rises and brings up the dissolved salts near or on the soil surface (Fig. 2). The soil has poor water holding capacity and poor natural drainage. As a result, fields remain flooded even after withdrawal of the monsoon and farmers cannot grow *kharif* crop in certain parts. At some places the top layer of the sub-surface water is fresh, and it benefits the subsequent rabi crop, especially under the scarce irrigation water but generally the benefit to *rabi* crop does not compensate for the loss of *kharif* crop. The salts pushed up by the irrigation water from the neighboring fields get concentrated towards the lower regions and the farmers are forced to abandon those parts of their lands.

Data Used

Considering the seasonal as well as spatial nature of salinity problem, the selection of the remote sensing data is an important aspect for the study. For monitoring and simulation of soil salinity affected area in irrigation command, the remote sensing data along with the field data, past history and geographical distributed information in the form of maps, etc. are essentially needed. Accordingly, the pertinent data was collected for the study area on the different aspects including physicochemical properties of soil, topographic details and other secondary data related with flooding and rainfall. For analysis two types of data spatial and associated-attributed-data were collected on different aspects of the study area. The spatial-data representing the geographic location of the features such as location of canal, minors, drain, and village were

Fig. 2. Waterlogged and Salinised Land

from available maps and topo-sheets of the area. The associated-attributed-data for different points providing the descriptive information on EC_e, pH, water-table depth and such other features were collected by conducting field surveys and from the available recent records for the study area. These data were measured at specified grid points. Each block is called a *murraba* having the dimensions of 332m in length and 299m in width. This block is further divided into 25 equal sub-blocks having dimensions of 66.4m length and 59.8m width called *killa*. The *murrabas* were numbered and for each village area boundary the numbering started from 1. The *murraba* lying in two or more villages has a number from both the villages in which it lies. The position of these grid points was taken by the initial of the village followed by the *murraba* number and *killa* number in which it lies. Then their actual locations were determined by GPS.

For the research purpose, consistent routine data regarding the salinity conditions in vast irrigation command are only possible from the remote sensing satellite, therefore selection of proper data becomes very important. For the present study the multi-temporal and multi-spectral remote sensing data were obtained considering the ever-changing field conditions in irrigation command.

The changes in the conditions of the problematic irrigation commands were kept in mind for maximizing the extractable information from multi-spectral measurements. To address this problem, the remote sensing data for the different periods, as listed below, was obtained from different sensors including IRS 1A, IRS 1B, IRS 1D, SPOT and LANDSAT (TM) for the dates as given in Table 1. These satellites have sensors with specific spatial resolutions to meet the objectives of the study. The date of satellite overpass was selected based on crop window, crop calendar, period of waterlogging (pre-monsoon, post-monsoon), exposure of puffy salt crust and percent cloud cover. The different dates of overpass were chosen for study of vegetation dynamics and land use/ cover change.

Software Used

ILWIS Academic 3.0, which is a Window-based, integrated GIS and remote sensing application software, was used.

Methodology

The remote sensing data available from the different satellites (Landsat, IRS and SPOT) were processed. The available satellite images were displayed on the computer screen. Since the data for the different dates were available from different satellites with different resolutions, it was considered necessary that resolution of all the images be made equal in all the cases so that the data for the different dates can be compared with each other. To obtain the same resolution, the data was re-assembled according to the resolution having pixel size of 20m × 20m. The study area as a part of the Western Yamuna Canal near Gohana was selected on the survey of India topo-sheets mentioned above. In another module of the ILWIS package, the subsets of the images were prepared according to the boundary of the study area located on the topo-sheets. Once the subsets of remote sensing imageries were ready for different dates, the next step was to make them geo-referenced. Initially, the PAN imagery of 24th February 2000 was geometrically corrected with respect to topographical map, and then, all the satellite scenes were geometrically corrected with respect to the corrected PAN imagery of 24th February 2000. A coordinate system with the name GOHA 1280 for all the images of the study area was defined. This system was based on non standard grid system followed in India with projection Lambert Conformal Conic and ellipsoid Everest India 1956 having parameters semi major axis= 6974316.6 yards and ratio of semi major axis to difference of semi-major and minor axis as 300.8017 with origin having latitude 26°N and longitude 74°E. The geometrically corrected imageries were subsequently radiometrically corrected by using the procedure as defined by Scott *et al.* (1988).

PCA is a powerful method of analyzing correlated multidimensional data. The data from all of the spectral bands involve a certain degree

Table 1. Details of the Satellite Data Used

Sl. No.	Date	Satellite	Sensor	Spectral Bands (µm)							Spatial Resolution (m) 36.25
				1	2	3	4	5	6	7	
1.	3rd January 1989	SPOT-1	HRV	0.50-0.59	0.61-0.68	0.79-0.89	-	-	-	-	20
2.	12th November1994	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
3.	22nd November1995	Landsat-5	TM	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	10.4-12.5*	2.08-2.35	30(120*)
4.	17th February 1996	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
5.	1st April 1996	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
6.	6th June 1996	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
7.	28th June 1996	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
8.	12th November 1996	IRS-1B	LISS-II	0.45-0.52	0.52-0.59	0.62-0.68	0.77-0.86	-	-	-	36.25
9.	24th February 2000	IRS-1D	LISS-III PAN	0.52-0.59 0.5-0.75	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23 5.8
10.	24th January 2002	IRS-1D	LISS-III	0.52-0.59	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23
11.	18th February 2002	IRS-1D	LISS-III	0.52-0.59	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23
12.	15th March 2002	IRS-1D	LISS-III	0.52-0.59	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23
13.	1st October 2002	IRS-1D	LISS-III	0.52-0.59	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23
14.	15th December 2002	IRS-1D	LISS-III	0.52-0.59	0.62-0.68	0.77-0.86	1.55-1.70	-	-	-	23

of redundancy. Then PCA was used as a data compression technique for solving the mentioned problem. The principal components are based on the eigenvectors of the covariance in the correlation matrix. The variance-covariance matrix C_X can be defined as:

$$C_X = \frac{1}{P-1} \sum_{i-1}^{P} (Xi - M)(X_i - M)^T \qquad \ldots(1)$$

where X is a given of N-dimensional variables with the mean vector M and P is the number of pixel. Each component and P is the number of pixel. Each component Y_i is denote by

$$Y_i = a_{1i}X_1 + a_{2i}X_2 + a_{3i}X_3 + ... + a_{Ni}X_n \qquad \ldots(2)$$

$$Y = a_i^T X \qquad \ldots(3)$$

where, a_i^T is the transpose of the normalized eigenvectors of the matrix C_x. The whole transformation can be shown as

$$Y = A^T X \qquad \ldots(4)$$

where A is the matrix of eigenvectors which gives the covariance matrix Cy of Y by

$$C_y = AC_xA^T \qquad \ldots(5)$$

The Cy matrix will be a diagonal matrix, which the elements are eigenvalues of Cx

$$Cy = \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_1 & ... & 0 \\ . & . & . & . \\ 0 & 0 & ... & \lambda_N \end{bmatrix} \qquad \ldots(6)$$

where $l_1 > l_2 > l_3 > l_N$

The principal component analysis technique was applied to the 14 temporal data sets of NIR, Red and Green bands, respectively to de-correlate possible redundant information of interest i.e. saline land cover into some specific PCs. For channels of multispectral data, the first principal component (PC 1) includes the largest percentage of the total scene variance and succeeding components (PC 2, PC 3,……, PC 14) each contain a decreasing percentage of the scene variance. Furthermore, because successive components are chosen to be orthogonal to all previous ones, the data they contain are uncorrelated (Lillesand, and Keifer, 2000).

Before visual analysis of the RGB colour composite generated from PC1s, careful examination of the eigen-matrix can reveal some important information about the structure of the original data and the results of the transformation (Loughlin, 1991). Each eigenvector can be used to assess the percentage of the original data variance making of successive PCs. In addition, it is possible to roughly approximate the contribution of each data to the individual PC images, by examining the eigenvectors for the original data (Loughlin, 1991). The FCC was then visually studied with the ground truth information to confirm some of the clearly visible land use/ cover in the FCC. The RGB colour composite was generated, in which PC1 of NIR is displayed in red and PC1 of Red and PC1 of Green is displayed in green and blue respectively.

Results and Discussion

It was found that the first three PCs concentrated most of the information of necessity showing highest variance percentage among all other PCs. PC1s showed similar weights for all band combinations of all the respective group of PCs. Their images allowed the discrimination between waterlogged areas (dark blue), waterlogged and saline areas (light dark to blue to bluish white) and severe saline areas (dark/ black) tones.

Having demonstrated the utility of individual PC1s, results have been found to be further improved by combining the fourteen dates PCs of green, red and NIR respectively. Several variations were experimented by combining PCs and it was concluded that a useful practice to delineate severe saline areas within an irrigation command, was to first create PC1s of the respective green, red and NIR bands by taking all the available dates and then by creating the RGB colour composite image from the PC1 of NIR, PC1 of red and PC1 of green by assigning red (R), Green (G) and Blue (B) colours respectively as shown in the figures 3, 4, and 5. Using this notion, FCC was built up to allow the simultaneous assessment of information on a variety of irrigation induced waterlogged and saline areas. The figure 6 shows the FCC image. After visually studying the FCC, it was found that the deep black tones indicated the salt affected areas. It has also been confirmed from the ground truthing and validation with the point soil EC_e and pH data. The statistics showed that about 128 hectares of agricultural land within the culturable command area was found to be highly salt affected as delineated in the figure 7. This might be attributed to the possibility that the soil salinity may be dynamic in

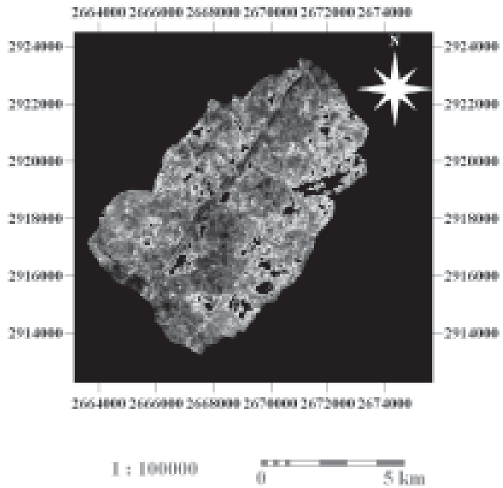

Fig. 3. PC 1 image created from 14 Green bands

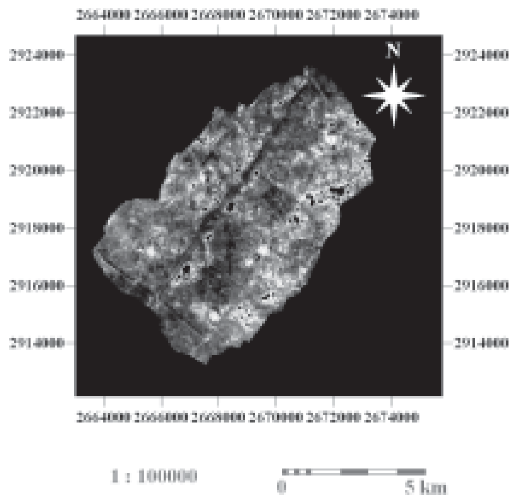

Fig. 4. PC 1 image created from 14 Red bands

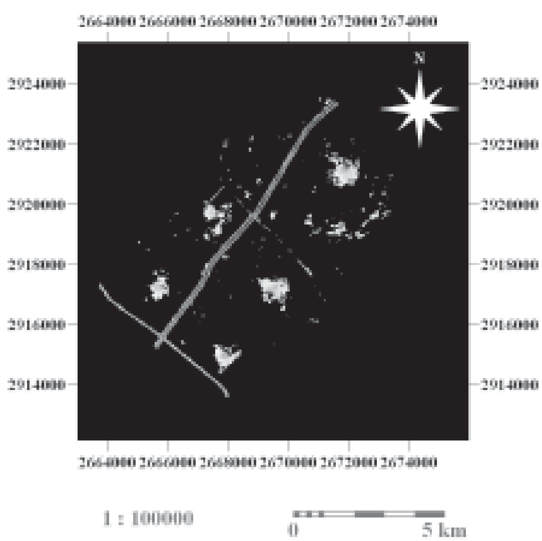

Fig. 5. PC 1 image created from 14 NIR bands

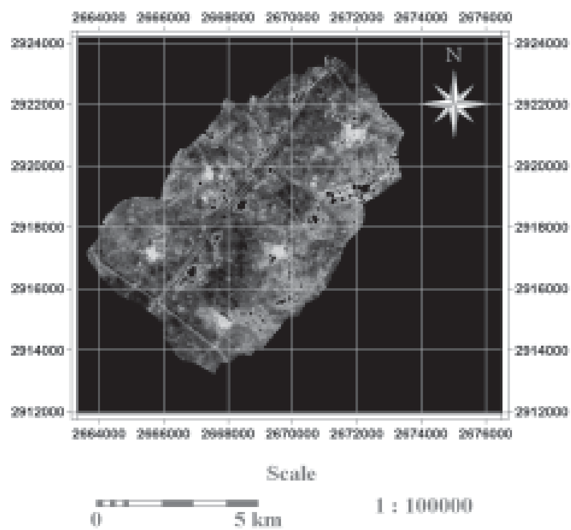

Fig. 6. RGB colour composite created from PC1s

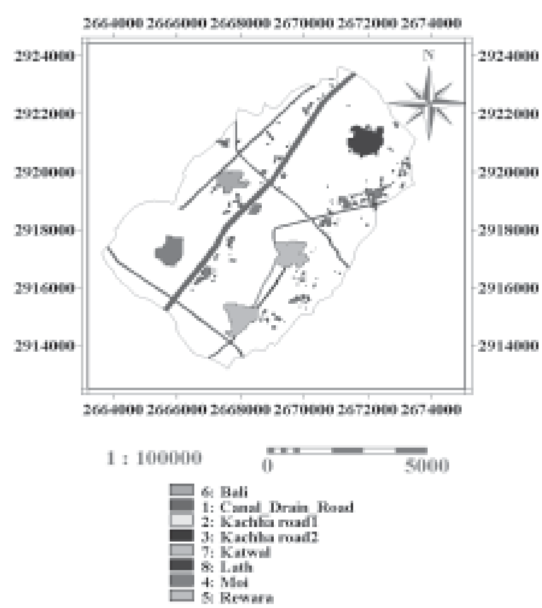

Fig. 7. Delineated saline area from FCC of PC1s

nature with respect to location. However, the highly saline areas in the form of white salt crusts patches are permanent in nature; whereas the medium to low saline areas can not be detected by satellite images directly because of their being dynamic in nature. The spatial distribution of highly salt affected area is shown in light green colour tone.

Conclusion

The use of colour composite created from the first principal component images produced from the principal component analysis of multispectral and multitemporal satellite data for a considerable period can be used successfully in mapping severe saline soils in a large irrigation command.

Acknowledgement

This study is a part of Ph.D. thesis submitted to the Post Graduate School, Indian Agricultural Research Institute, New Delhi, India, and is duly acknowledged.

References

Choubey, V.K. and Choubey, R. 2000. Remote Sensing of Waterlogged Area in Sriram Sagar Command. 8th I.C.I.D. International Drainage Workshop 31st January-4th February, 2000, New Delhi (India).

Dwivedi, R.S. 1996. Monitoring of Salt Affected Soils of the Indo-Gangetic Alluvial Plains Using Principal Component Analysis. *International Journal of Remote Sensing* **17**: 1907-1914.

Lillesand, T.M. and Keifer, R.W. 2000. Remote Sensing and Image Interpretation. John Wiley and Sons, New York.

Loughlin, W.P., 1991, Principal Component Analysis for Alteration Mapping. *Photogrammetric Engineering and Remote Sensing* **57**: 1163-1169.

Sah, K.A., Eiumnoh, A., Shunji, M. and Parkpian, P., 1995, Mapping of Salt-Affected Soils using Remote Sensing and Geograp Information Systems: A Case Study of Nakhon Ratchaslma, Thailand. www.gisdevelopment.net, Pcrs: 1-3.

Souza Filho, C.R. and Drury, S.A. 1998. Evaluation of JERS-1 (FUYO-1) OPS and Landsat TM images for Mapping of Gneissic rocks in Arid Areas. *Remote Sensing Environment* **71**: 158-182.

Scott, J.R., Salvaggio, C. and Volchok, W.J. 1988. Radiometric Scene Normalisation using Pseudo-invariant Features. *Remote Sensing Environment* **26**: 1-16.