Lateritic soil mapping of the Phrae basin, northern Thailand using satellite data

Myint Soe, Krit Won-In, Isao Takashima, Punya Charusiri

a Graduate School of Engineering and Resource Science, Akita University, Akita 010-8502 Japan
b Department of Earth Sciences, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
c Center for Geo-Environmental Science, Akita University, 1-1 Tegatagakuen-cho, Akita 010-8502 Japan
d Department of Geology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

* Corresponding author, e-mail: myintsoe68@gmail.com

ABSTRACT: Landsat 7 Enhanced Thematic Mapper Plus image data was used to identify and map lateritic soil zones in the Phrae basin which is one of the largest intermountain basins in northern Thailand. The lateritic soil zones were discriminated using band ratio and principal component analysis. The lateritic soil detection images were processed by band ratio (band 3 / band 1), principal component analysis of bands 1 and 3, and principal component analysis of bands 1, 3, 4, and 5. The results of these three indices were superimposed using GIS to define a preliminary lateritic soil image of the study area. A threshold method was used for converting a grey scale image into a binary image. Different threshold values were used to find the most probable areas of lateritic soil zones in the image. The threshold values were determined from a published geological map and known lateritic soil areas with good exposure in the image. The quality of the results was evaluated by the normalized difference vegetation index. Field investigation was carried out to substantiate the remote sensing investigation and the laboratory GIS analysis. This method can also be applied to other lateritic soil and iron oxide regions.

KEYWORDS: lateritic soil mapping, remote sensing, band ratio, principal component analysis, thresholding, normalized difference vegetation index

INTRODUCTION

This paper demonstrates how to identify and map lateritic soil using remote sensing data at low cost. Lateritic soils are used extensively in the tropical areas as traditional construction materials and roads. They occur mostly in humid tropical climates between 30° N and 30° S. Lateritic soils consist mainly of the minerals kaolinite, goethite, hematite, and gibbsite, and are rich in silica, aluminium, and iron oxides (hematite and goethite). Primary minerals are altered to characteristic assemblages of the secondary minerals, kaolinite, hematite, goethite, and residual quartz.

The genesis of laterization involves leaching of silicates, formation of colloidal sesquioxide, precipitation of oxides with increasing crystallinity, and dehydration as the soil weathers. Primary minerals in the parent rock, such as feldspar, quartz, and ferro-magnesia minerals are converted into diffuse goethite, followed by well-crystallized goethite, and finally hematite. In the advanced stages of weathering, crystallization leads to the formation of iron and/or aluminium oxide concretions. During laterization, iron oxide and aluminium coat and bind the clay particles leading to a change in the microstructure of the soil.

Facies maps of lateritic soils are important. They can be used to plan for material supply for infrastructure construction, to target mineral deposits, and for safe water supplies. Laterization produces infertile, poor draining. Goethite, formed during laterization, concentrates arsenic which can affect localized ground water. Also, magnesium sulphate and sodium chloride in groundwater from lateritized basement causes diarrhoea. However, epsomite in drinking water may be beneficial in countries where drugs are expensive, as magnesium sulphate benefits women with eclampsia or pre-eclampsia and their babies.

Studies of laterites are important because of their economically viable concentrations of iron, nickel, gold, phosphate, aluminium, titanium, manganese, and rare earth elements. Laterite is an indicator of soil fertility, the usability of an area to cultivate specific crops, and geological age. Lateritic soil is conductive to plant root hairs and is able to absorb other diffused nutrients from the water. It can be used for aquarium plant fertilizers and aquarium plant food. Fish ponds in lateritic soil zones exhibit...
low availability of phosphorus, resulting in a restricted accumulation of various fish food organisms.

The study area covers the Phrae basin which is one of the largest intermountain basins in northern Thailand (18° 00′–18° 30′ N, 100° 05′–100° 20′ E). The basin is 15–25 km wide, 50–60 km long, and covers approximately 1,000 km². The Yom River flows from north to south following the basin axis, and is regarded as a very important channel for transporting sediments into the basin. The lateritic soil is present in low and high terrace deposits and colluvial sedimentary units. Determining the spatial distribution of different types of iron with traditional fieldwork and laboratory analysis is time-consuming and expensive. This study proposes to demonstrate that remote sensing mapping is a useful and cheaper tool to determine the presence of lateritic soil.

Geology of the Phrae basin

The geology of the Phrae basin and its surroundings is shown in Fig. 1. Most of the previous studies were mainly restricted to the geological reports and maps provided by the Geological Survey Division, Department of Mineral Resources of Thailand. Based on these maps (at scales of 1:250,000), the eastern and western sides of the basin is characterized by Permian and Triassic rock units. The rock units in the eastern side of the basin are largely composed of conglomerate, sandstone, siltstone, and shale. The rock units in the western edge of the basin are dominated by volcanic rocks (e.g. rhyolite and volcanic tuff), shale, siltstone, and mudstone.

Sinsakul studied Quaternary geology in the Phrae basin and classified the Quaternary units into four distinctive units. The units are, in descending order, high-terrace, stream-valley, low-terrace (including old alluvial plain), and active Yom River alluvial deposits. So far, no age determination has been done for these Quaternary deposits of the basin. Maneenai reported that the high terrace of the Phrae basin probably belongs to the late Pliocene by correlation with that of the Lampang basin, the age of which has been documented. Maruoka et al proposed that the Phrae formation is of Tertiary deposits which are unconformably overlain by Quaternary alluvial deposits.

Srisuwan et al proposed five sedimentary cycles for the Cenozoic stratigraphy of the Phrae basin based on seismic, lithological, and field data. For each cycle, deposits are dominated by braided channel and flood plain deposits in the early stage, subsequently change to shallow swamp lake and lake-margin deposits, and end with a small delta (or crevasse sprays). Based on palynological data, they suggested that the age of the Phrae basin was probably Late Oligocene to Early Miocene. Won-in proposed six quaternary units on the basis of geomorphologic evidence and the sedimentary deposit age (using the TL dating method). The six quaternary units were identified in descending order as flood plain and active river deposits (6 ka to recent), low terrace deposits (40–12 ka), alluvial fan deposits (16–10 ka and 160–100 ka), middle terrace deposits (88–37 ka), colluvial deposits (≥270 ka), and high terrace deposits (990–170 ka).

METHODS

This study used remote sensing digital image processing techniques and the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image data. Data analysis was carried out using TNTmips version 7.0. The preliminary lateritic soil map is the output derived from using all the evidence from fieldwork in the Phrae basin area. In order to map lateritic soil, two different image processing techniques were performed. These techniques are called band ratio and principal component analysis. After using either technique to process the images, a thresholding method was used by applying a simple Boolean logic algorithm. The threshold values were determined according to known lateritic areas / good exposure areas (Phae Muang Phi National Park and the Phrae city landfill, see Fig. 1). Typically, an object pixel is given a value of 1 while a background pixel is given a value of 0. The quality of the final result was evaluated by the normalized difference vegetation index (NDVI) image. Field investigation was carried out to confirm the satellite analysis. The final result was marked and digitized based on the lateritic pixel density area.

Pre-processing of Landsat 7 ETM+ image

A Landsat 7 ETM+ image (path 130/row 047) of Phrae basin was acquired on 25 Dec 1999. This image has a solar incidence angle of 40.83° and an azimuth angle of 148.05°. The scene is georeferenced in the UTM projection and the WGS-84 ellipsoid. The data was radiometrically converted to sensor reflectance using an image-based correction method. Processing of the image included radiometric correction to remove atmospheric effects.

Reflectance is an intrinsic property of materials and is independent of illumination conditions, slope, sensor, and atmospheric effects. The reflectance data extracted from the digital number (DN) for the VNIR
and SWIR spectral ranges was processed according to the following procedure. Each digital number represents a range of intensity from 0 to 255, where the value 254 corresponds to the maximum intensity and the value 255 corresponds to the saturation value of the sensor. First, DN was converted into a radiance value \( L_{\lambda} \) (in W/m\(^2\) sr \( \mu \)m) using

\[
L_{\lambda} = \frac{(L_{\text{max}} - L_{\text{min}})}{(Q_{\text{CALMAX}} - Q_{\text{CALMIN}})} (\text{DN} - Q_{\text{CALMIN}}) + L_{\text{min}}
\]

where \( Q_{\text{CALMIN}} \) is set to 1, \( Q_{\text{CALMAX}} \) has a value of 255, \( L_{\text{max}} \) and \( L_{\text{min}} \) are determined by the date of acquisition sensor band calibration, and DN is the digital number\(^1\). Second, once converted to
units of radiance, the Landsat 7 ETM+ imagery can be converted into units of the top of atmosphere reflectance using

$$P_P = \frac{\pi L_\lambda d^2}{ESUN \lambda \cos(Q_s)}$$

(2)

where $P_P$ denotes the unitary planetary reflectance, $d$ denotes the Earth Sun distance in astronomical units, $ESUN \lambda$ denotes the mean solar exoatmospheric irradiances (W/m² μm) for each band [B1= 1,969; B2= 1,840; B3= 1,551; B4= 1,044; B5= 82.07; B7= 82.07; B8= 1,368], and $Q_s$ denotes the solar zenith angle.

**Vegetation index**

Vegetation indices derived from satellite data are widely used to interpret the surface vegetation information. Some studies of semi-arid shrublands have successfully used these vegetation indices for vegetation abundance estimation. The main problem in the identification of minerals is interference from vegetation. High vegetation cover decreases the prediction accuracy of lateritic soil mapping. Therefore, the quality of the final result was evaluated by the normalized difference vegetation index (NDVI). The NDVI is calculated using

$$NDVI = \frac{\rho(\lambda_{NIR}) - \rho(\lambda_{red})}{\rho(\lambda_{NIR}) + \rho(\lambda_{red})}$$

(3)

where $\rho(\lambda_{NIR})$ denotes the reflectance in the near infrared band and $\rho(\lambda_{red})$ denotes the reflectance in the red band of the Landsat ETM+ image data. The NDVI gives a measure of vegetative cover on the land surface over wide areas. Dense vegetation shows up very strongly in the image, and areas with little or no vegetation are also clearly identified. The NDVI is a measure of the difference in reflectance between these wavelength ranges, and takes values between −1 and 1. Values more than 0.5 indicate dense vegetation and negative values indicate no vegetation.

**Band ratio method**

Band ratio is a technique that has been used for many years in remote sensing to effectively display spectral variations. Image division or spectral band ratio is one of the most common mathematical operations applied to multi-spectral image data. Ratio images are calculated by dividing the DN values in one spectral band by the corresponding pixel value in another band. Each object has a unique spectral reflectance pattern. The band ratio operation which transforms data and reduces the effects of such environmental conditions can provide unique information not available in any single band which is very useful for determining the nature of surface materials. The band ratio method uses

$$BV_{i,j,r} = \frac{BV_{i,j,k}}{BV_{i,j,l}}$$

(4)

where $BV_{i,j,r}$ is the output ratio for the pixel at row $i$, column $j$, $BV_{i,j,k}$ is the brightness value at the same location in band $k$, and $BV_{i,j,l}$ is the brightness value in band $l$. The range of $BV_{i,j,r}$ is theoretically from 0 to $\infty$.

**Principal component analysis method**

Principal component analysis (PCA) is a multivariate statistical technique that selects uncorrelated linear combinations (eigenvector loadings) of variables in such a way that each successively extracted linear combination or principal component (PC) has a smaller variance. The eigenvector matrix used to perform PCA for each subset was examined to identify which PC contained the target (mineral) information. This technique indicates whether the materials are represented as bright or dark pixels in the principal components according to the magnitude and sign of the eigenvector loadings. Feature oriented principal components selection is based on the examination of PCA eigenvector loadings to decide which of the principal components will extract information directly related to the theoretical spectral signatures of specific targets. This technique indicates whether the materials are represented by bright or dark pixels based on the sign and magnitude of the eigenvectors.

**RESULTS AND DISCUSSION**

The vegetation mapping using NDVI has shown both areas of healthy vegetation and areas of sparse or no vegetation (Fig. 2). Vegetation shows absorption features from 450–680 nm and high reflectance in the near infrared (1,600–2,200 nm) due to the chlorophyll absorption. The absorption bands at 1,400–1,900 nm are related to the water content. Iron oxides and vegetation show similar reflectance in Landsat 7 ETM+ band 1 (450–520 nm) and band 2 (520–600 nm). Therefore these bands are not very useful for separating these materials. On the other hand, band 3 (630–690 nm) shows high reflectance for the iron oxides and a strong absorption for vegetation. This is useful to differentiate oxides from vegetation (Figs. 3A and 3B).

Thematic Mapper bands are too wide to allow the identification of single minerals. Nevertheless, bands in the near and middle infrared serve to identify
groups of minerals belonging to the hydroxyl and sulphate groups which can be used for exploration purposes. Iron oxide and hydroxyl minerals have spectral features in the visible and infrared parts (400–2,500 nm) (Figs. 3A and 3B). In general, lateritic spectra present lows at 400–500 nm and 920 nm, a 650 nm shoulder, and a higher broad reflectance at longer wavelengths (Fig. 3B). These spectral characteristics are indicative of the predominance of goethite as the source of ferric oxide as confirmed by X-ray diffraction analysis. According to the laterite spectral signature, a notable reflection occurred in the band 3 of ETM+ at 700 nm as well as a considerable absorption in band 1 at 400 nm. Band 4 (760–900 nm) is used for vegetation identification, but also contains an absorption band for iron oxides at 900 nm. Both features can be used to separate oxides from vegetation (Fig. 3B). Absorption anomalies at wavelengths less than 900 nm are a good indication of hematite. When the anomalies are at wavelengths above 900 nm, jarosite or goethite are more abundant.

The maximum and minimum reflection orders of the lateritic soil characteristic are easily isolated among the other objects in the image. Laterite gives high reflectance values in band 3 and low values in band 1. The most effective index used to extract the lateritic soil areas is the ratio of band 3 to band 1. When this index was applied to the Phrae basin image data, the highest band ratio values clearly coincided with the lateritic soil situated at the national park (Fig. 4A). The use of this index can be affected by band 3 absorption around 660 nm which is due to chlorophyll and other pigments contained in green vegetation. This fact has been ignored in this study because the vegetation mapping based on NDVI has shown little or no vegetation in the areas considered. Fig. 3A shows that iron oxide minerals have a higher
reflectance at bands 5 and 7 than at band 3. However, this study did not use the band ratio method with SWIR wavelengths (bands 5 or 7). These ratio images can sometimes be confused with OH absorption minerals (kaolinite, alunite) and carbonate absorption (calcite) in minerals (Fig. 3A).

For lateritic soil mapping, the band ratio 3/1 image showed the best results and coincided with known lateritic soil outcrops (Table 1). The ratio between band 4 and 3 showed the thick vegetated terrain as bright pixels. The ratio between bands 5 and 7 showed the clay mineral content. The ratio between bands 5 and 4 gave differences between iron oxide dominance and hydroxyl with areas of high oxides giving brighter pixels due to stronger absorption of the band 4.

PCA has been applied to bands 1 and 3 of Landsat ETM+ image data. The general statistics and principal component eigenvectors and eigenvalues are calculated in Table 2. The resulting images produced by PCA were found to be accurate in the lateritic soil zone. Feature-oriented principal components selection is based on the examination of PCA eigenvector loadings to decide which of the principal components will extract information directly related to the theoretical spectral signatures of specific targets. The lateritic soil minerals give high reflectance values in band 3 and low in band 1. We therefore identify the principal component in which the difference of reflectance is large (Table 2). In PC1 both bands have positive eigenvalues which are not useful for separating bands 1 and 3. Eigenvalues of opposite signs in bands 1 and 3 in PC2 makes the bands separable. According to the spectral characteristics of the laterite, the second component of the principal component analysis of band 3 (0.4219) and band 1 (–0.9066) include the biggest difference (1.3285) between these two bands. This component enables discrimination between the lateritic soil and iron oxides. PC2 has a high and negative loading from band 1 (–0.9066) and high and positive loading from band 3 (0.4219). PC2 is selected as the lateritic soil image and the brightest pixels represent the lateritic soil rich areas (Fig. 4B).

The same PCA technique was applied to bands 1, 3, 4, and 5 based on the spectral characteristics of iron oxides (ferric oxide absorption and reflectance at bands 1 and 3 and ferrous iron mineral absorption and reflectance at bands 4 and 5) and vegetation (Table 3). PC1 had a high loading from band 4 (73 % of the data variance) and this was attributed to albedo, whereas PC2 gave vegetation as bright pixels due to the positive contribution from band 4 and negative contribution from bands 1 and 3. According to the eigenvector statistics (Table 1B), PC4 has a higher negative loading of band 3 (–0.4750) and high and positive loading from band 1 (0.8790), and includes

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Band ratios for lateritic soil and green vegetation based on Landsat 7 ETM+ wavebands.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Band Ratio</td>
</tr>
<tr>
<td>1</td>
<td>Band 3/1</td>
</tr>
<tr>
<td>2</td>
<td>Band 3/4</td>
</tr>
<tr>
<td>3</td>
<td>Band 5/4</td>
</tr>
<tr>
<td>4</td>
<td>Band 5/1</td>
</tr>
<tr>
<td>5</td>
<td>Band 7/4</td>
</tr>
<tr>
<td>6</td>
<td>Band 7/1</td>
</tr>
<tr>
<td>7</td>
<td>Band 4/3</td>
</tr>
<tr>
<td>8</td>
<td>Band 5/7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Eigenvector loadings for lateritic soil mapping based on bands 1 and 3 for Landsat 7 ETM+ of Phrae basin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axis</td>
<td>Landsat 7 ETM+ bands</td>
</tr>
<tr>
<td></td>
<td>Band 1</td>
</tr>
<tr>
<td>PC1</td>
<td>0.4219</td>
</tr>
<tr>
<td>PC2</td>
<td>-0.9066</td>
</tr>
</tbody>
</table>

www.scienceasia.org
Fig. 4 (A) band 3/band 1 ratio image (bright pixels: lateritic soil area). (B) PC2 image from Landsat 7 ETM+ bands 1 and 3 (bright pixels: lateritic soil area). (C) PC4 image from Landsat 7 ETM+ bands 1, 3, 4 and 5 (dark pixels: lateritic soil area).

The biggest difference (1.3540) between these two bands. PC4 gave strong negative loading for band 3 and a relatively strong positive loading for band 1. It was therefore assumed to show lateritic soil pixels as dark pixels (Fig. 4C).

The three analyses of lateritic soil mapping showed mostly the same results. After applying three indices, band ratio image (3/1), PC2 (1, 3), and PC4 (1, 3, 4, 5), the different threshold values were used to find the most probable area of lateritic soil in the image. Thresholding, the simplest method of image segmentation, is an image processing technique for converting a greyscale or colour image into a binary image based upon a threshold value. The threshold values were determined according to known lateritic areas in the image (areas covered mostly by shrubs or grass, i.e. Phae Muang Phi National Park and landfill of Phrae city shown in Fig. 1). For each image a sub-layer image was created. If a pixel in the original image has an intensity value less than the threshold value, the corresponding pixel in the sub-layer image is set to black (value 0), otherwise, it was set to white (value 1). Thus

\[
g(i,j) = \begin{cases} 
1; & \text{if } f(i,j) \geq T \\
0; & \text{if } f(i,j) < T 
\end{cases}
\]

where \( f(i,j) \) is the pixel value in each layer of ratio image (3/1), PC2 (1, 3) or PC4 (1,3,4,5), \( T \) is the threshold value, and \( g(i,j) \) is the corresponding pixel value in the sub-layer. The results of these three indices were overlaid on top of each other using GIS to define the preliminary lateritic soil image. The three extracted lateritic soil layers in binary format were then added together in order to form the final lateritic soil threshold layer (Fig. 5). When using the thresholding method, we could adjust the result from the field GPS data of lateritic soil for the known outcrops.

| Table 3 Eigenvector loadings for lateritic soil mapping based on bands 1, 3, 4 and 5 of Landsat 7 ETM+ of Phrae basin. |
|---|---|---|---|---|---|---|
| Axis | Bands | Eigenvectors | % Variance | Cumulative |
| | 1 | 3 | 4 | 5 | | |
| PC1 | 0.0633 | 0.1572 | 0.7165 | 0.6767 | 73.0267 | 73.0267 |
| PC2 | -0.2124 | -0.4549 | 0.6535 | -0.5664 | 24.8168 | 97.8436 |
| PC3 | 0.4221 | 0.7367 | 0.2438 | -0.4687 | 1.9164 | 99.7599 |
| PC4 | 0.8790 | -0.4750 | -0.0108 | 0.0395 | 0.2401 | 100.0000 |
Lateritic soil area discrimination is not possible when vegetation is present. Therefore, the quality of the final results was evaluated by the NDVI image and field investigation (Fig. 6A). A mask included all these objects applied on the final lateritic soil layers to remove noise and false anomalies. Then, the visual interpretations using the geological maps are incorporated to eliminate the remaining noise to form the final laterite layer. Fig. 6A shows the lateritic soil mapping thresholding results on top of the NDVI image. The final result is marked and digitized based on the lateritic pixel density area (Fig. 6B).

The uses of Landsat TM images for soil mapping has always been hindered by vegetation cover and complex methods are required to remove the influence of vegetation27. The above results show that there are differences in the spectral signatures of vegetated areas, but the results do not show compositional differences and thus provide little information on the soil physical properties. The techniques were based on separation of the areas based on the presence of lateritic soil and vegetation cover. The results showed lateritic soil mapping based on spectral and spatial analysis using the remote sensing application of Landsat 7 ETM+ data.

The presence of vegetation interfered with the spectral response of iron oxide minerals and hence the band ratios did not yield good results. Nevertheless, the band ratio method is easy and useful. The identification of lateritic soil was best accomplished with the principal components analysis. The minimization of the vegetation interference was
obtained with the results of the principal components analysis, which identified the lateritic soil.

ACKNOWLEDGEMENTS

We wish to thank the editor and the reviewers of this manuscript. We are grateful to Kitti Khaowiset, a geologist at the Department of Mineral Resources of Thailand for field support. Field work in Thailand was supported by the Department of Mineral Resources, Thailand. Mr. Paul Moiya Kia, Graduate School of Engineering and Resource Science, Akita University is thanked for detailed and insightful comments that greatly improved the original manuscript. The authors would like to thank Professor Ryutaro Tateishi and his students, Center for Environmental Remote Sensing, Chiba University, for discussions and the developing of image processing skills. The Consortium for Spatial Information which provided the SRTM DEM data. The authors would like to thank the Global Land Cover Facility which kindly provided the Landsat 7 ETM+ data of the study area.

REFERENCES


