The Effects of Haze on the Spectral and Statistical

Properties of Land Cover Classification

Asmala Ahmad

Department of Industrial Computing
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

Shaun Quegan

Department of Applied Mathematics
School of Mathematics and Statistics
University of Sheffield
Sheffield, United Kingdom

Abstract

Haze occurs almost every year in Malaysia and is caused by smoke which originates from forest fire in Indonesia. It causes visibility to drop, therefore affecting the data acquired for this area using optical sensor such as that on board Landsat satellite. The effects of haze on the data can be observed from the spectral and statistical properties of land cover classification. The work presented in this thesis is meant to analyse the statistical properties of land cover classification of hazy dataset. Maximum Likelihood (ML) was found to be a preferable classification scheme in which the effects of haze can be investigated. The study made use of hazy dataset that were simulated based on real haze spectral and statistical properties. By investigating these dataset, the spectral and statistical properties of the land classes can be systematically analysed, in which showing that haze modifies the class spectral signatures and statistical properties, consequently causing the data quality to decline.

Keywords: Haze, ML Classification, Spectral, Statistical, Landsat

1 Introduction

Haze reduces visibility due to the attenuation (i.e. scattering and absorption) of solar radiation by the haze constituents [13]. Most haze consists of aerosols (suspension of fine solid particles or liquid droplets in the atmosphere) and trace gases, ranging in size from a few nanometres to a few micrometers [1], [10]. Studies have shown that haze that is due to biomass burning contains large amounts of hazardous gases, i.e., carbon monoxide (CO), nitrogen dioxide (NO_2) and sulphur dioxide (SO_2), and particulate matter, i.e., PM10 [11], [14], [15]. Compared to gases, aerosol has a more significant impact on visibility. The largest aerosol loading from biomass burning occurs below 5 km in altitude [12]. Atmospheric scattering and absorption depend very much on the wavelength of the radiation and the size of the atmospheric constituents it interacts with. Scattering is usually much stronger for short wavelengths than long wavelengths. Particles with size approximately 0.1 to 10 μm are particularly effective in Mie scattering in the visible wavelength regions (0.4 – 0.7 μm) hence can impair ground level visibilities [4]. In order to quantify the effects of haze, we therefore need to identify a suitable classification method and performance criteria against which to measure these effects.

One of the primary uses of remote sensing data is to classify land covers [5], [7], [8]. A large number of classification methods are available, but our principal aim is to select the method most appropriate to the studies of haze [6], [9]. Our criteria for this selection include:

• simplicity, i.e. the practicality of using a large amount of data. This should involve a smaller number of procedures but should produce reasonably accurate and standard results,
• the ability to select important land covers with an acceptable accuracy, i.e. each pixel will be assigned to the correct land cover on the ground – the performance of the method should not be easily affected by factors such as the complexity of land covers, topographic conditions, etc. and
• objectivity, i.e. not involving tuning by a user to improve performance – the generated classification works straight away without needing any adjustment in terms of the number of classes, training pixels, etc.

These criteria lead us to consider the use of Maximum Likelihood (ML), which is a supervised method [2], [3]. In order to facilitate the use of this method, we can analyse its behaviour from a single image of Landsat 5 TM satellite from 11 February 1999. This image contains the main land covers of Malaysia and has clear sky conditions (free from haze and little cloud cover), and therefore meets the purpose of our study, i.e. to provide a base map for use in studying the effects of haze on land cover classification and how this can be corrected (i.e. does not involve change detection). The Landsat 5 TM was launched in the 1984 and use Thematic Mapper (TM) as the main sensor. The TM comprises of visible band 1

(0.45-0.52 µm), band 2 (0.52-0.60 µm) and band 3 (0.63-0.69 µm), reflective infrared band 4 (0.76-0.90 µm), band 5 (1.55-1.75 µm) and band 7 (2.08-2.35 µm), with 30 m spatial resolution and thermal infrared band 6 (10.40-12.50 µm) with 120 m spatial resolution.

This study aims to analyse the spectral and statistical properties of land cover classification of hazy dataset. The study makes used hazy dataset that were simulated based on real haze spectral and statistical properties. By using these dataset, the spectral and statistical properties of the land classes can be systematically investigated and therefore the effects of haze can be assessed. In achieving the aim, Section 2 describes ML classification of hazy dataset, Section 3 elaborates the statistical and spectral properties of Classes for the Hazy Datasets and finally, Section 4 concludes the study.

2 ML Classification on the Simulated Hazy Dataset

The vector-based structure of a dataset, the hazy band i, $L_i(V)$ can be written as [9]:

$$L_i(V) = \left(1 - b_i^{(1)}(V)\right)T_i + L_O + b_i^{(2)}(V)H_i \qquad \text{... (1)}$$

where T_i is signal component, H_i is haze component, L_O is path radiance due to natural scattering of the atmosphere, $b_i^{(1)}$ is weighting due to signal attenuation and $b_i^{(2)}$ is weighting due to haze intensity.

ML classification was carried out using all 6 bands to produce 11 classes, viz. coastal swamp forest, dry land forest, oil palm, rubber, cleared land, sediment plumes, water, coconut, bare land, urban and industry. To carry out ML classification on the hazy scenes, we need training pixels within the hazy scene. For this purpose, the ROIs (regions of interest) for different land classes (different colours) that were applied on the clear scene were used as a template. Figure 1 shows (a) patches of ROIs for different land classes (indicated by different colours) overlaid on bands 4, 5 and 3 (assigned to red, green and blue) of a 4 km visibility hazy scene used for selecting training pixels from the hazy scene and (b) the ML classification.
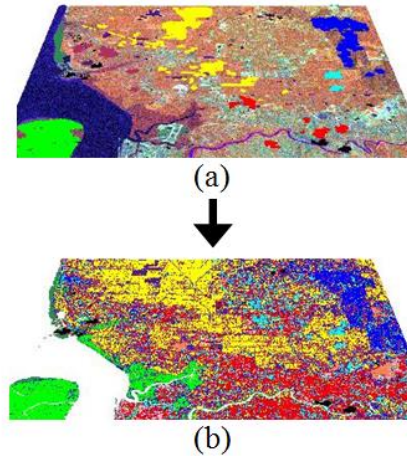
Fig. 1. (a) Patches of different colours are ROIs for different land classes used for selecting training pixels from a 4 km visibility hazy scene and (b) the ML classification.

Figure 2 shows the 4 km hazy datasets before and after ML classification for visibilities 20 km, 10 km, 6 km, 4 km, 2 km and 0 km. These visibilities are chosen to visually show the transition from clear to very hazy conditions. It is obvious that the effects of haze become more severe on bands 3, 2, and 1 (assigned to the red, green and blue channels respectively) as visibility decreases (images on the left). These bands are displayed since they are more affected by the haze than the bands with longer wavelengths. Therefore, classification is much more influenced by the effects of haze in shorter than longer wavelengths. The middle images show the corresponding ML classification using training pixels from the hazy dataset itself. As expected, the ML classification performance degrades as visibility drops. The classes are clearly inseparable at 0 km visibility.

3 Spectral and Statistical Properties of Land Cover Classes in Hazy Dataset

In order to extract the statistics of the classes generated by the ML classification, the classification produced from the clear scene was used as a template to demarcate the pixels in each class and then to compute the class means and correlation between bands in hazy data. To illustrate this, Figure 3 shows plots of mean radiances versus bands for all classes. As expected, the means are more affected at shorter than longer visibilities. At 16 visibility, the difference between
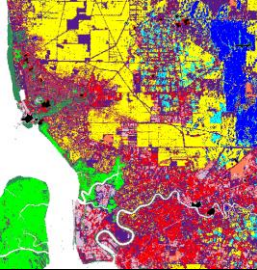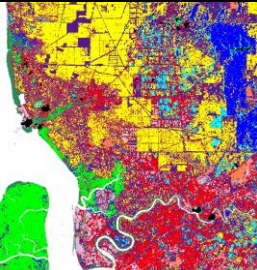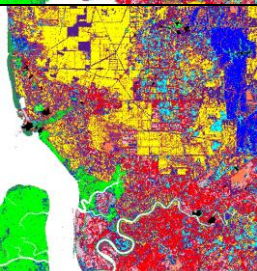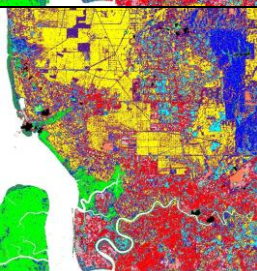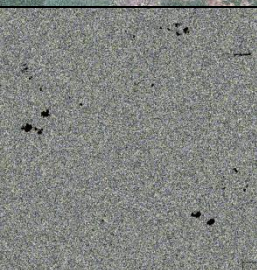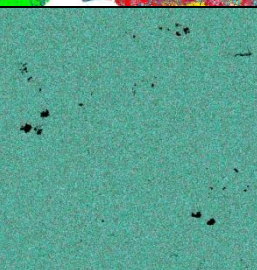
(i) Before Classification	(ii) After ML Classification
(a) 20 km visibility	
(b) 10 km visibility	
(c) 6 km visibility	
(d) 4 km visibility	
(e) 0 km visibility	

Fig. 2. Bands 3, 2 and 1 assigned to red, green and blue channels respectively (left), the ML classification using training pixels from hazy datasets (right) and ML classification using training pixels from clear datasets for (a) 20 km (clear), (b) 10 km, (c) 6 km, (d) 4 km and (e) 0 km visibility.

the original class radiance (red curve) and hazy class radiance (black curve) is very small for most classes. Similarly, there is little difference between the standard deviation of the original class radiance (red vertical bars) and the hazy class radiance (black vertical bars). At 2 km visibility, the haze clearly increases the radiance of bands 1, 2 and 3 for most classes except for bare land and industry, which decrease. The increase in radiance tends to occur for dark classes (e.g. forests and vegetation) because the apparent radiance is dominated by the haze radiance (i.e. radiance scattered directly to the satellite's field of view). A decrease in radiance tends to occur for bright classes because the haze scatters some of the solar radiance out of the satellite's field of view before reaching the ground, and attenuates the reflected radiation on the way back. These effects become more apparent as haze severity increases. This is consistent with Equation (1) that shows that haze increases $L_i(V)$ through the scattering effects on $\beta_i^{(2)}(V)H_i$ but at the same time decreases $L_i(V)$ through the absorption effects on $\left(1-\beta_i^{(1)}(V)\right)T_i$. Hence, the absorption effect is proportional to T_i; therefore the brighter the surface, the higher the absorption, consequently the more $L_i(V)$ decreases. However, it should be noted that this is true in absolute terms but not relative; $\beta_i^{(1)}(V)$ does not depend on T_i. It can also be seen that most classes exhibit an increase in standard deviation as visibility reduces. In other words, the haze increases the variability in the intensity of the class pixels and consequently leads to an increase in the pixels' standard deviation. This is expected from Equation (1); since:

$$\mathrm{Var}\Big[L_i(V)\Big]=\mathrm{Var}\Big[\big(1-\beta_i^{(1)}(V)\big)T_i+\beta_i^{(2)}(V)H_i\Big]$$

$$=\mathrm{Var}\Big[\big(1-\beta_i^{(1)}(V)\big)T_i\Big]+\mathrm{Var}\Big[\beta_i^{(2)}(V)H_i\Big]+2C\Big[\big[\big(1-\beta_i^{(1)}(V)\big)T_i\big],\big[\beta_i^{(2)}(V)H_i\big]\Big]$$

T_i is independent of H_i, so the third term equal to zero. However we cannot discard the T_i term since the target is not constant, so there still exist some variance.

$$\mathrm{Var}\Big[L_i(V)\Big]=\Big[1-\beta_i^{(1)}(V)\Big]^2\mathrm{Var}(T_i)+\Big[\beta_i^{(2)}(V)\Big]^2\mathrm{Var}(H_i)$$

When haze gets more severe, V decreases but $\beta_i^{(1)}(V)$ and $\beta_i^{(2)}(V)$ both increase. Hence, the contribution from the target variance decreases but that from the haze increases. The balance between the two depends on target brightness because bright targets (such as industry and bare land) have larger variance. This is more noticeable in the dark classes (such as vegetation and water) due to the greater difference between the haze and dark class spectral measurement.
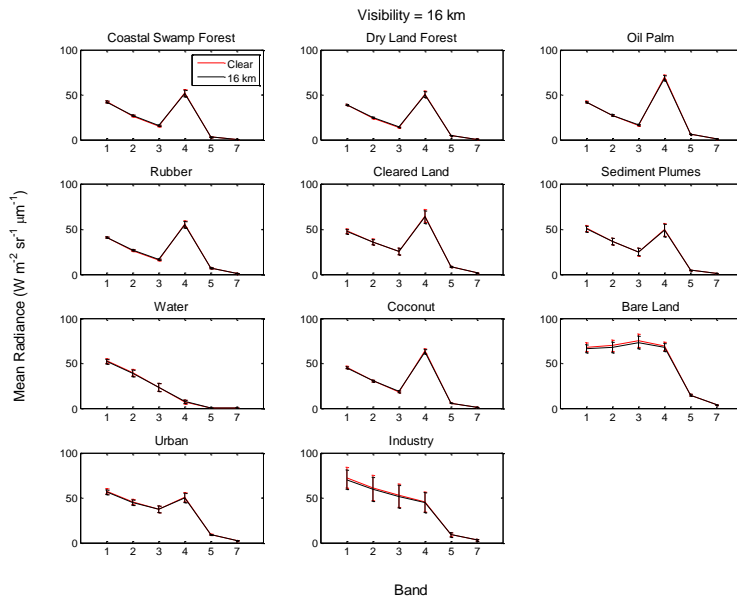
(a)
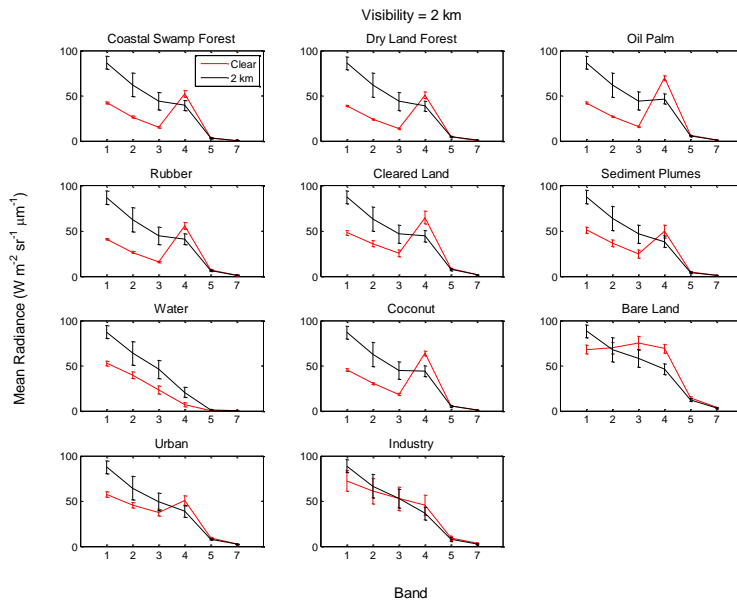
(b)

Fig. 3. Mean radiances versus bands of individual classes for a scene with haze (black) and without haze (red) at visibility (a) 16 and (b) 2 km. Vertical bars indicate standard deviations.

Figure 4 shows plots of the means of all classes versus bands for visibilities 20 km, 16 km, 12 km, 6 km, 4 km and 0 km. At 20 km visibility, all classes exhibit

their true spectral signature curves, but experience little modification by haze at 16 and 12 km visibility. At 4 visibility, these curves are severely modified by the haze and become inseparable as the visibility reduces to 0 km. At 0 km visibility, all curves become very close, approximating the pure haze spectral signature. In other words, during no or light haze, the spectral signature of the classes are evident because the true signal radiance predominates, but as the haze gets severe, the spectral signature of the classes vanishes and is replaced by that of pure haze.
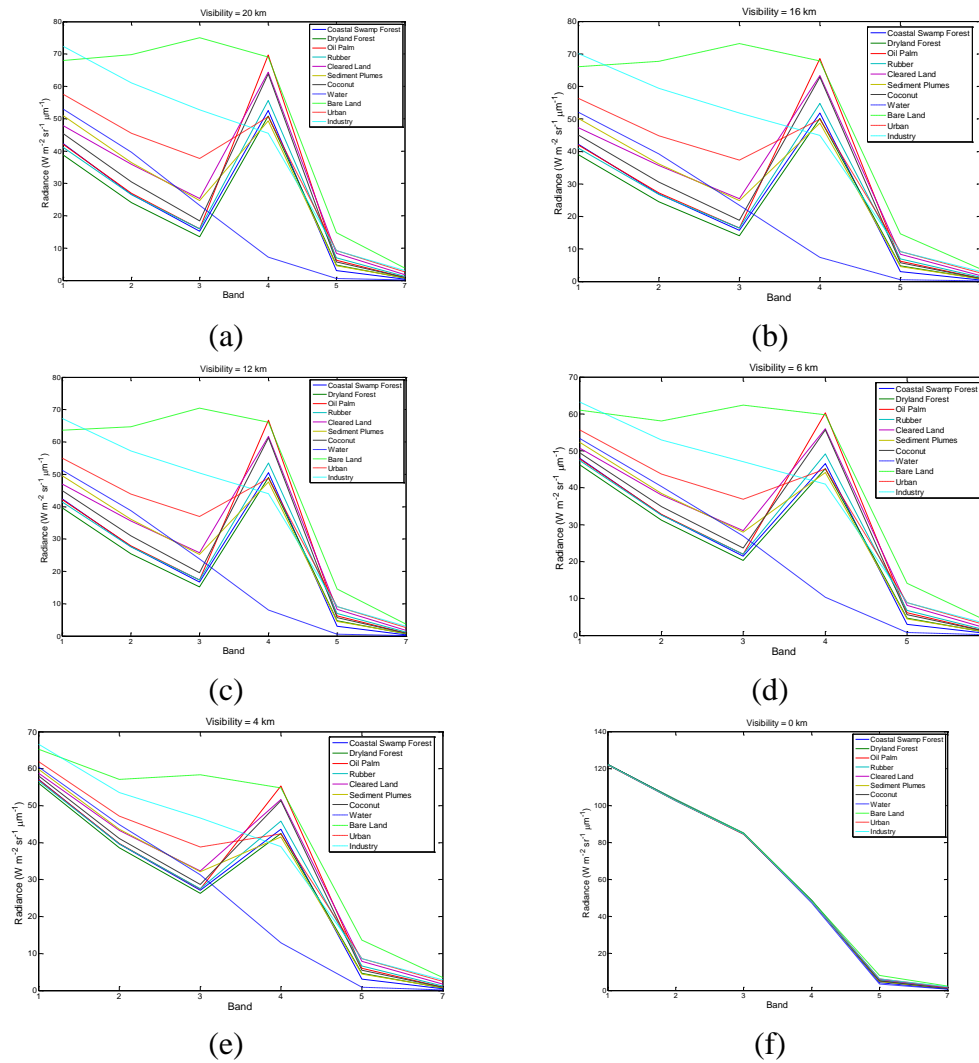

(a)   (b)   (c)   (d)   (e)   (f)

Fig. 4. Mean spectral signatures of the 12 classes at visibilities (a) 20, (b) 16, (c) 12, (d) 6, (e) 4 and (f) 0 km.

For each class, correlations between different band pairs were computed for visibilities running from 20 km to 0 km from the simulated hazy datasets by using ENVI and then checked using Equation (2) with MATLAB; both show a very good

agreement. The correlation between band k and band l of a simulated radiance scene with a particular visibility V can be expressed by the following equation:

$$\rho_S\left(k,l\right) = \frac{C\left[L\left(V,k,l\right)\right]}{\sqrt{Var\left[L\left(V,k\right)\right]\,Var\left[L\left(V,l\right)\right]}} \qquad \text{... (2)}$$

where

$C\left[L\left(V,k,l\right)\right] = \left(1-\beta_k^{(1)}\left(V\right)\right)\left(1-\beta_l^{(1)}\left(V\right)\right) C\left[T\left(k,l\right)\right] + \beta_k^{(2)}\left(V\right)\beta_l^{(2)}\left(V\right) C\left[H\left(k,l\right)\right]$ and

$Var\left[L\left(V,k\right)\right] = \left(1-\beta_k^{(1)}\left(V\right)\right)^2 Var\left[T\left(k\right)\right] + \beta_k^{(2)2}\left(V\right) Var\left[H\left(k\right)\right],$ and

$C\left[T\left(k,l\right)\right]$, $Var\left[T\left(k\right)\right]$ and $Var\left[T\left(l\right)\right]$ are measured from the clear dataset while $C\left[H\left(k,l\right)\right]$, $Var\left[H\left(k\right)\right]$ and $Var\left[H\left(l\right)\right]$ are measured from the pure haze dataset. Here, we assume $\beta_k^{(1)}\left(V\right)$ and $\beta_k^{(2)}\left(V\right)$ are constant throughout the image.

Plots of correlation against visibility for coastal swamp forest, oil palm, urban and water are shown in Figure 5(a-e). Correlations at 20 km visibility represent the classes' original correlation during clear sky condition (i.e. no haze); the correlation of pairs 1-2, 1-3 and 2-3 is higher than other pairs due to their adjacent wavelengths. On the other hand correlations at 0 km visibility represent those of pure haze; e.g. pairs 2-3, 2-4, 3-4, and 5-7 have the highest correlations at 0 km visibility, while pairs 1-5 and 1-7 the lowest. For coastal swamp forest, i.e. a very dark class, the correlation in most pairs starts to increase steadily at longer visibilities (i.e. 18 to 12 km), gets rapid at moderate visibilities (12 to 4 km) but steady again at shorter visibilities (i.e. less than 4 km). This shows the haze significantly modifies the correlations at shorter compared to longer visibilities, with a rapid increase in modification occurs at moderate visibilities. In such case, as haze becomes more severe, $C\left[L\left(V,k,l\right)\right]$ gets bigger and so does $\rho_S\left(k,l\right)$. However, such trend is not so obvious for water because they already possesses quite high correlations at longer visibilities due to the original spectral properties of the classes. For oil palm, the rapid modification occurs at quite short visibilities (i.e. 6 to 2 km) due to the less dark properties of the class; i.e. its spectral properties are influenced by the ground reflectance from the spaces between the oil palm trees. For urban, a rapid increase in correlation occur at slightly longer visibilities compared to bare land, signifying the stronger effects of haze due to the less bright properties of the class.

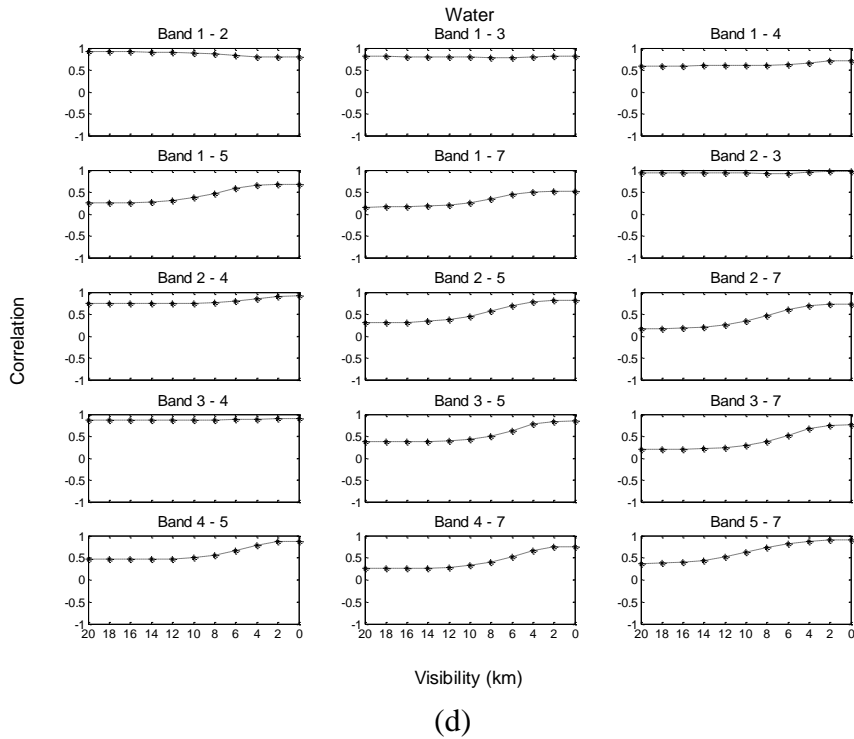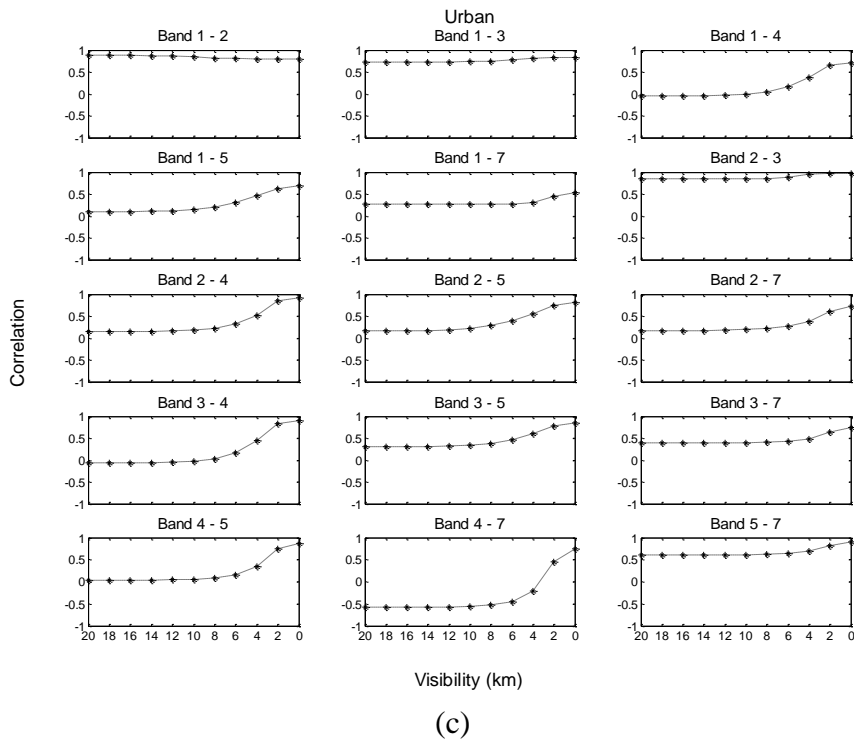Coastal Swamp Forest

(a)
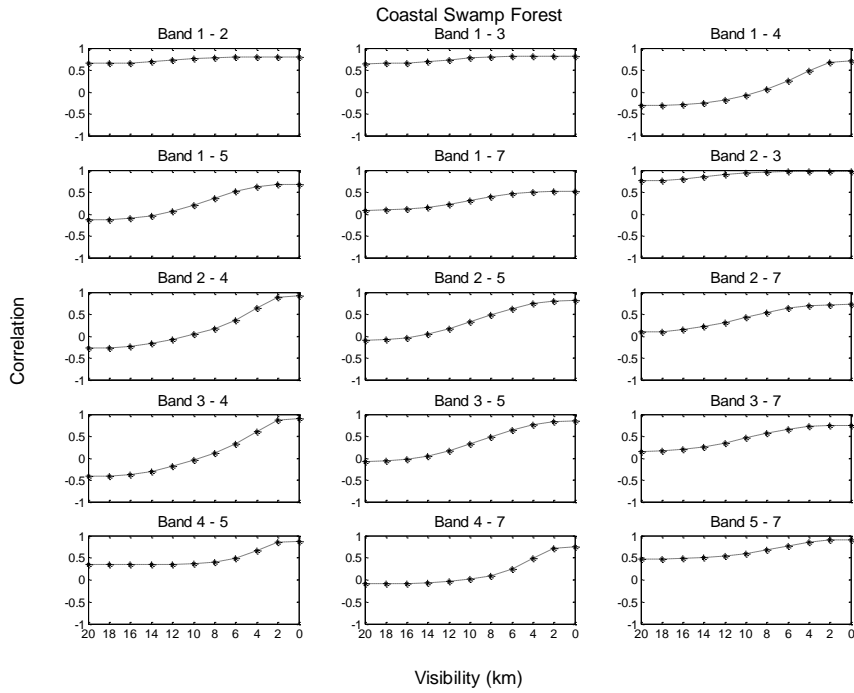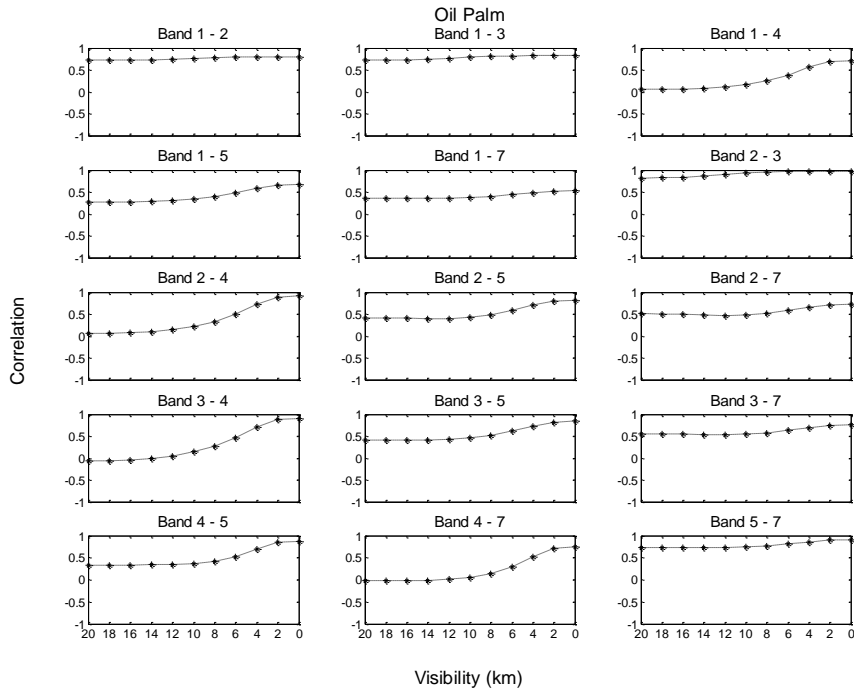
Oil Palm

(b)

Fig. 5. Correlation between bands with reducing visibility for (a) coastal swamp forest and (b) oil palm, (c) urban and (d) water.

4 Conclusions

In this study, we successfully quantified the spectral and statistical properties of land cover classification under hazy conditions. Maximum Likelihood (ML) is found to be the preferable classification scheme in which the effects of haze were investigated. The study made use of hazy dataset that were simulated based on real haze spectral and statistical properties. By investigating these dataset, the spectral and statistical properties of the land classes were systematically analysed, in which showing that haze modifies the class spectral signatures and statistical properties, consequently causing the quality of the dataset to decline.

References

[1]   A. Ahmad and M. Hashim, Determination of haze using NOAA-14 satellite data, Proceedings on The 23rd Asian Conference on Remote Sensing 2002 (ACRS 2002), (2012).

[2]   A. Ahmad and S. Quegan, Analysis of maximum likelihood classification on multispectral data, Applied Mathematical Sciences, 6 (2012), 6425 – 6436.

[3]   A. Ahmad and S. Quegan, Analysis of maximum likelihood classification technique on Landsat 5 TM satellite data of tropical land covers, Proceedings of 2012 IEEE International Conference on Control System, Computing and Engineering (ICCSCE2012), (2012), 1 – 6.
http://dx.doi.org/10.1109/iccsce.2012.6487156

[4]   A. Ahmad and M. Hashim, Determination of haze API from forest fire emission during the 1997 thick haze episode in Malaysia using NOAA AVHRR data, Malaysian Journal of Remote Sensing & GIS, July, 1 (2000), 77 – 84.

[5]   A. Ahmad and S. Quegan, Comparative analysis of supervised and unsupervised classification on multispectral data, Applied Mathematical Sciences, 7 (74) (2013), 3681 – 3694. http://dx.doi.org/10.12988/ams.2013.34214

[6]   A. Ahmad and S. Quegan, Haze reduction from remotely sensed data. Applied Mathematical Sciences, 8 (36) (2014), 1755 – 1762.
http://dx.doi.org/10.12988/ams.2014.4289

[7] U. K. Mohamad Hashim and A. Ahmad, The effects of training set size on the accuracy of maximum likelihood, neural network and support vector machine classification. Science International, 26 (4) (2014), 1477 – 1481.

[8] A. Ahmad, Analysis of Landsat 5 TM data of Malaysian land covers using ISODATA clustering technique, Proceedings of the 2012 IEEE Asia-Pacific Conference on Applied Electromagnetic (APACE 2012), (2012), 92 – 97. http://dx.doi.org/10.1109/apace.2012.6457639

[9] A. Ahmad and S. Quegan, Haze Modelling and Simulation in Remote Sensing Satellite Data. Applied Mathematical Sciences, 8 (159) (2014), 7909 – 7921. http://dx.doi.org/10.12988/ams.2014.49761

[10] A. Asmala, M. Hashim, M. N. Hashim, M. N. Ayof and A. S. Budi, The use of remote sensing and GIS to estimate Air Quality Index (AQI) Over Peninsular Malaysia, GIS development, (2006), 5pp.

[11] A. Heil, A. and J. G. Goldammer, Smoke-haze pollution: a review of the 1997 episode in Southeast Asia. Regional Environmental Change, 2 (2001), 24 – 37. http://dx.doi.org/10.1007/s101130100021

[12] C. Chiang, W. Chen, W. Liang, S.K. Das and J. Nee, Optical properties of tropospheric aerosols based on measurements of lidar, sun-photometer, and visibility at Chung-Li (251N, 1211E). Atmospheric Environment, 41 (2007), 4128 – 4137. http://dx.doi.org/10.1016/j.atmosenv.2007.01.019

[13] M. Hashim, K. D. Kanniah, A. Ahmad, A. W. Rasib, Remote sensing of tropospheric pollutants originating from 1997 forest fire in Southeast Asia, Asian Journal of Geoinformatics 4, 57 – 68.

[14] M. Mahmud, Mesoscale equatorial wind prediction in Southeast Asia during a haze episode of 2005. Geofizika, 26 (1) (2009), 67 – 84.

[15] M. Radojevic, Chemistry of Forest Fires and Regional Haze with Emphasis on Southeast Asia. Pure and Applied Geophysics, 160 (1) (2003), 157 – 187. http://dx.doi.org/10.1007/s00024-003-8771-x