Haze Effects on Satellite Remote Sensing Imagery and their Corrections

Asmala Ahmad¹, Mohd Mawardy Abdullah³
Centre for Advanced Computing Technology (C-ACT)
Fakulti Teknologi Maklumat dan Komunikasi (FTMK)
University Teknikal Malaysia Melaka (UTeM)
Melaka, Malaysia

Shaun Quegan²
School of Mathematics and Statistics
University of Sheffield, Sheffield, United Kingdom

Suliadi Firdaus Sufahani³
Fakulti Sains Gunaan dan Teknik
Universiti Tun Hussein Onn Malaysia
Johor, Malaysia

Hamzah Sakidin⁴
Department of Fundamental and Applied Sciences
Universiti Teknologi PETRONAS
Perak, Malaysia

Abstract—Imagery recorded using satellite sensors operating at visible wavelengths can be contaminated by atmospheric haze that originates from large scale biomass burning. Such issue can reduce the reliability of the imagery and therefore having an effective method for removing such contamination is crucial. The principal aim of this study is to investigate the effects of haze on remote sensing imagery and develop a method for removing them. In order to get a better understanding on the behaviour of haze, the effects of haze on satellite imagery were initially studied. A methodology of removing haze based on haze subtraction and filtering was then developed. The developed haze removal method was then evaluated by means of signal-to-noise ratio (SNR) and classification accuracy. The results show that the haze removal method is able to improve the haze-affected imagery qualitatively and quantitatively.

Keywords—Haze effects; haze removal; remote sensing; accuracy; visibility

I. INTRODUCTION

Haze occurs almost every year in Malaysia and is caused by smoke that originates from forest fires in nearby regions. The recent September 2019 haze was considered among the worst as the air pollutant index (API) in Sri Aman, Sarawak reached 407 indicating ‘hazardous level’ [1]. Previously, the worst haze episode was recorded in 1997 when API in Miri, Sarawak hit 839 (‘emergency level’). This was followed by 2005 (541 in Kuala Selangor), 2010 (432 in Muar), 2013 (762 in Muar) and 2015 (308 in Shah Alam) [2]. Haze causes visibility to drop and therefore affecting the images acquired for this area using optical sensor such as that onboard the Landsat-5 TM satellite. In Malaysia, haze monitoring is carried out by the Malaysian Meteorological Department and Department of Environment Malaysia in terms of visibility and API respectively. Malaysia has a typical tropical monsoon climate characterized by uniformly high mean temperature with temperature approximately 27°C, a relatively high mean annual rainfall that exceeds 2000 mm per year and humidity 70% to 90% throughout the year. The wind over the country is generally mild and variable. However, there are some periodic changes in the wind flow patterns that describe the two monsoon seasons namely the north-east monsoon, known as the wet season that occurs from November to March and the south-west monsoon, known as the dry season that occurs from June to September. The remaining months, April to May and October to November, are known as the transitional periods. Because the wind comes from the south-west and there is much less rain during the south-west monsoon and the second transitional period, smoke from the forest fires in Sumatra remains suspended in the atmosphere for a long time and drifts to Malaysia, causing haze. This paper focuses on the 2005 haze episode that caused a drop in visibility in most places in Malaysia. Fig. 1 shows photos of clear and hazy conditions in Putrajaya, the federal administrative centre of Malaysia, located about 30 km from Kuala Lumpur. Due to the hazardous properties of the haze constituents, a sudden increase in respiratory and eye-related illnesses cases was reported. The drop in visibility conditions also badly affected economy-related activities including tourism, transportation, fisheries and production sectors, which caused a big loss to Malaysia.

Visibility measurement is carried out by the Malaysian Meteorological Department on a daily basis through a network of 149 monitoring stations. For public convenience, haze severity is categorised into five levels; visibilities more than 10 km represent ‘clear’, 5 to 10 km visibilities represent ‘moderate’, 2 to 5 km visibilities represent ‘hazy’, 0.5 to 2 km visibilities represent ‘very hazy’ and visibilities less than 0.5 km represent ‘extremely hazy’ (Table I).

The Department of Environment Malaysia operates a network of 51 stations, where 36 stations are in West Malaysia (or Peninsular Malaysia) and 15 in East Malaysia. Due to the potential harm to human health, five main pollutants are measured which are SO₂, NO₂, CO, O₃ and PM₁₀ [4]. Based on their locations and the types of pollutant measured, the stations are categorised into Residential (20 stations), Industrial (12 stations), Traffic (1 station), Background (1 station) and PM₁₀ (2 stations). The difference between these categories is the types of pollutant measured (Table II). The location of air quality monitoring stations in West Malaysia and a typical monitoring station are shown in Fig. 2 [5].
and were derived based on human health data and Malaysian guidelines represent the safe level for each pollutant averages), and O\textsubscript{3} and SO\textsubscript{2} (hour running averages). For reporting purposes, the same averaging times are used: PM\textsubscript{10} and SO\textsubscript{2} (24-hour running averages), CO (8-hour running averages), and O\textsubscript{3} and NO\textsubscript{2} (1-hour running averages). The Malaysian guidelines represent the safe level for each pollutant and were derived based on human health data and recommendations from the World Health Organisation (WHO). For example, a PM\textsubscript{10} concentration of 150 µg/m\textsuperscript{3} corresponds to 100 API, and is the upper limit for the safe level; PM\textsubscript{10} concentrations exceeding this are likely to cause adverse health effects. Conversion of the PM\textsubscript{10} concentration from µg/m\textsuperscript{3} to API can be done using the equations shown in Table V.

**TABLE III. API RANGE, STATUS, LEVEL OF POLLUTION AND HEALTH MEASURE [4]**

<table>
<thead>
<tr>
<th>API Range</th>
<th>Status</th>
<th>Level of Pollution</th>
<th>Health Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 50</td>
<td>Good</td>
<td>Low, no ill effects on health</td>
<td>No restriction of activities to all groups.</td>
</tr>
<tr>
<td>51 to 100</td>
<td>Moderate</td>
<td>Moderate, no ill effects on health</td>
<td>No restriction of activities to all groups.</td>
</tr>
<tr>
<td>101 to 200</td>
<td>Unhealthy</td>
<td>Mild aggravation of symptoms and decreased exercise tolerance in persons with heart or lung disease.</td>
<td>Restriction of outdoor activities for high-risk persons. General population should reduce vigorous outdoor activity.</td>
</tr>
<tr>
<td>201 to 300</td>
<td>Very Unhealthy</td>
<td>Significant aggravation of symptoms and decreased exercise tolerance in persons with heart or lung disease.</td>
<td>Elderly and persons with known heart or lung disease should stay indoors and reduce physical activity. General population should reduce vigorous outdoor activity. Those with any health problems to consult doctor</td>
</tr>
<tr>
<td>300 – 500</td>
<td>Hazardous</td>
<td>Severe aggravation of symptoms and endangers health.</td>
<td>Elderly and persons with existing heart or lung disease should stay indoors and reduce physical activity. General population should reduce vigorous outdoor activity.</td>
</tr>
<tr>
<td>&gt; 500</td>
<td>Emergency</td>
<td>Severe aggravation of symptoms and endangers health.</td>
<td>General population advised to follow the orders of National Security Council and always follow announcements through the mass media.</td>
</tr>
</tbody>
</table>

**TABLE IV. AIR QUALITY MEASUREMENT GUIDELINES [4]**

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Averaging Time</th>
<th>Malaysian Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(ppm)</td>
<td>(µg/m\textsuperscript{3})</td>
</tr>
<tr>
<td>O\textsubscript{3}</td>
<td>1 hour</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>8 hours</td>
<td>0.06</td>
</tr>
<tr>
<td>CO</td>
<td>1 hour</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>8 hours</td>
<td>9</td>
</tr>
<tr>
<td>NO\textsubscript{2}</td>
<td>1 hour</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>24 hours</td>
<td>0.04</td>
</tr>
<tr>
<td>SO\textsubscript{2}</td>
<td>1 hour</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>24 hours</td>
<td>0.04</td>
</tr>
<tr>
<td>PM\textsubscript{10}</td>
<td>24 hour</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1 year</td>
<td>50</td>
</tr>
</tbody>
</table>
TABLE V. EQUATIONS FOR API CALCULATION BASED ON PM$_{10}$ 24-HOUR RUNNING AVERAGES [4]

<table>
<thead>
<tr>
<th>PM$_{10}$ concentration, C ($\mu$g/m$^3$)</th>
<th>Equation used for conversion to API</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C \leq 50$</td>
<td>API = $C$</td>
</tr>
<tr>
<td>$50 &lt; C \leq 350$</td>
<td>API = $50 + (C - 50) \times 0.5$</td>
</tr>
<tr>
<td>$350 &lt; C \leq 420$</td>
<td>API = $200 + (C - 350) \times 1.43$</td>
</tr>
<tr>
<td>$420 &lt; C \leq 500$</td>
<td>API = $300 + (C - 420) \times 1.25$</td>
</tr>
<tr>
<td>$C \geq 500$</td>
<td>API = $400 + (C - 500)$</td>
</tr>
</tbody>
</table>

Fig. 3 shows visibility and PM$_{10}$ intensity against Landsat overpass date in 2005 for Klang Port, Petaling Jaya, Kuantan and Kota Bharu. The sudden increase in PM$_{10}$ and drop in visibility in August 2005, particularly in Klang Port and Petaling Jaya, was associated with the occurrence of haze in that year. It can be seen that extreme haze occurred between 6 and 22 August 2005. Klang Port and Petaling Jaya, which are located on the west of Malaysia, with average visibility and PM$_{10}$ concentration approximately 11 km and 70 API respectively, experienced lower visibility and higher PM$_{10}$ intensity than Kuantan and Kota Bharu, with average visibility and intensity approximately 14 km and 40 API respectively, which are located on the east. Since extremely hazy and very hazy conditions are quite rare in Malaysia, we are more concerned on a more frequently occurring conditions, specifically moderate; in which later the haze removal will be tested onto an image with moderate haze.

Fig. 4 shows scatterplots of visibility for Petaling Jaya vs. Klang port, Petaling Jaya vs. Kuantan, Petaling Jaya vs. Kota Bharu and Kuantan vs. Kota Bharu, together with linear fits to these plots. It is clear that the visibility correlation between nearby stations, Petaling Jaya and Klang Port (0.708) is much higher than non-neighbouring stations, Petaling Jaya and Kuantan (0.04), Petaling Jaya and Kota Bharu (0.02) and Kuantan and Kota Bharu (0.08). In this thesis, the testing of the developed haze removal method will be carried out over Bukit Beruntung area, by using PM$_{10}$ measurements from Petaling Jaya station.

II. HAZE EFFECTS ON REMOTE SENSING IMAGERY

A. Simulating Hazy Images

6SV1 is the vector version of the 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) [7], [8], though it also works in scalar mode. The vector version is introduced to account for radiation polarisation, due to Rayleigh scattering in a mixed molecular-aerosol atmosphere, which is to be used when performing atmospheric correction. In our study, the 6SV1 was used in simulating haze effects, therefore the radiation polarisation effect was assumed negligible. Hence, our interest was in the scalar mode of 6SV1, which was similar to 6S.

B. ML Classification on the Simulated Hazy Images

ML classification was carried out using all 6 bands to produce 11 classes, which were coastal swamp forest, dry land forest, oil palm, rubber, cleared land, sediment plumes, water, coconut, bare land, urban and industry [9], [10], [11]. To carry out ML classification on the hazy scenes, we need training pixels within the hazy scene. For this purpose, the ROIs for different land classes (different colours) that were applied on the clear scene were used as a template. Fig. 5 shows bands 3, 2 and 1 assigned to red, green and blue channels respectively (left). ML classification using training pixels from hazy images were performed for 20 km (clear) down to 0 km visibility (right).

C. The Effects of Haze on the Hazy Images

We carried out spectral signature analysis by plotting mean radiance versus band 1, 2, 3, 4, 5 and 7 for all 11 land covers and for visibility 20 km (clear) down to 0 km [12]. The outcome for visibility 16 km and 2 km are shown in Fig. 6. Mean radiances versus bands of individual classes for a scene with haze (black) and without haze (red) were simultaneously plotted to signify the effects of haze. Vertical bars indicate standard deviations in km visibility. It is clear that the spectral signature of the land covers is severely modified at 2 km compared to 16 km visibility. At 16 km visibility the modification can be hardly seen due to the insignificant scattering that occurred and moreover it still considered as good visibility. On the other hand, significant modification can
be seen at 2 km visibility indicating severe scattering particularly for bands with shorter wavelengths (band 1, 2 and 3) compared to longer wavelengths (band 5 and 7). When the plotted together for all 11 land covers (Fig. 7), the spectral signatures get narrower at 2 km compared to 16 km visibility indicating that the land covers are becoming visually inseparable as visibility gets very low.

To understand better the effects of haze on remote sensing images, band correlation versus visibility was plotted for all possible combinations and for all 11 land covers (Fig. 8). It can be seen that for oil palm and urban, correlation gets close to ‘1’ as visibility reduces. This is due to the fact that haze has greater effects at low visibility due to the very severe scattering that makes the radiance becomes very similar although when measured from different bands and eventually giving correlation close to ‘1’ signifying highly correlated.

![Image](https://example.com/image.png)

**Fig. 5.** Bands 3, 2 and 1 Assigned to Red, Green and Blue Channels Respectively (Left), the ML Classification using Training Pixels from Hazy Images (Middle) and ML Classification using Training Pixels from Clear Images for (a) 20 km (Clear), (b) 10 km, (c) 6 km, (d) 4 km, (e) 2 km and (f) 0 km Visibility. Note that Images a(ii) and a(iii) are the Same and are Displayed for Convenience. Black Patches are Cloud and its Shadow (Masked Black).

**Fig. 6.** 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 in km Visibility.

**Fig. 7.** Mean Spectral Signatures of the 12 Classes at Visibilities (a) 16 and 2 km.
Subsequently, classification accuracy versus visibility was plotted for all 11 land covers as shown in Fig. 9 [13], [14]. As expected, a drop in accuracy can be seen as visibility gets low. However, the drop trend varies for different land covers. Due to the similar physical properties, oil palm and coconut show somewhat similar trend. Similarly, a common trend also can be seen for dryland forest and rubber. A faster drop can be seen for industry compared to others particularly due to mixed pixels issue. It is also obvious that accuracy drop for water only occurs at 2 km visibility due to the fact that water has somewhat uniform spectral properties and tend to be classified correctly at 20 km down to 2 km visibility.

The reason for the drop in accuracy can be clarified by plot of pixels versus classes for oil palm and urban as shown in Fig. 10. Different coloured lines indicate different visibilities. In an ideal situation, when classified as oil palm. Oil palm pixels should be 100% at all visibilities as for 20 km visibility. However, migration of oil palm pixels to other land cover classes seems to occur as visibility reduces causing a decrease in classification accuracy.

To understand the overall trend on how haze affects classification, overall classification accuracy versus visibility was plotted for 20 km down to 0 km visibility images. Overall classification accuracy and Kappa coefficient versus visibility are shown in Fig. 11.
III. HAZE CORRECTION

This section attempts to correct for the haze effects based on the findings in the previous sections. The simulated hazy images undergo mean subtraction and three types of filtering, namely average, median and Gaussian. The signal-to-noise (SNR) of the filtered images versus visibility were plotted for each filtering type. The outcomes are shown in Fig. 12.

To examine how far the haze correction is able to improve the hazy images, we plotted classification accuracy versus visibility for filtered and unfiltered images for average, median and Gaussian filtering. The accuracy difference between after and before filtering was also determined to predict the improvement made for the hazy images. Fig. 13 shows classification accuracy against visibility for average, median and Gaussian filtering. Fig. 14 shows accuracy difference between filtered and unfiltered images for all the three types of filtering.

IV. IMPLEMENTATION ON REAL HAZY IMAGES

Based on the outcomes from the previous sections, a procedure was developed to implement the haze correction on real hazy images. The procedure can be represented by a flowchart as shown in Fig. 15. The study area within the real hazy images was located in Selangor, Malaysia. Based on the flowchart, the haze correction procedure involves two key phases, namely haze removal and quality assessment.

A. Haze Removal

The real hazy images were initially examined to check whether having uniform haze or not. If the haze was uniform, the weighted haze mean was to be estimated based on pseudoinvariant features (PIF) technique and subtracted from the hazy images [15]. On the other hand, if the haze was not uniform, the haze was to be segmented first using minimum noise fraction (MNF) technique. Subsequently, Gaussian spatial filtering was performed onto the hazy images [16].

B. Quality Assessment

In this phase, the performance of the haze correction was to be quality assessed. For such purpose, ML classification was first performed onto the hazy images. Finally, classification accuracy was calculated and examined.

The result is of the implementation is shown in Fig. 16. The top row shows the colour composite image of band 3, 2 and 1 assigned to red, green and blue respectively. The middle row shows ML classification while the bottom row shows the corresponding enlarged version. The left and middle column shows the image before and after haze removal while the right column shows the clear image representing non-hazy...
condition. Visually, in the top row, the middle image shows that most of the haze have been eliminated, making it looked very close to the clear image in the right column. This is more obvious as can be seen from the images in the third row. A more significant effect can be seen on dark areas [16], [17], [18], [19]. In the second row, the misclassification between urban, rubber and cleared land as can be seen in the first column has been corrected making the classification very close to that of the right column. In the third row, the effects before and after the haze correction can be seen more clearly in the selected area. The outcomes also suggest that dealing with homogenous haze is likely to be simpler compared to heterogenous haze [20], [21], [22], [23].

V. CONCLUSIONS

In this paper, haze effects on remote sensing imagery and their corrections have been studied. To understand the effects of haze on remote sensing imagery, haze was initially simulated over real remote sensing images to represent hazy images. Spectral signature analysis, correlation analysis and classification accuracy analysis have been carried out to assess the effects of haze on the simulated hazy images. Subsequently, a methodology for correcting haze has been developed involving weighted mean subtraction and filtering for correcting haze. For assessing the quality of the corrected images, SNR and classification accuracy have been used. Eventually the developed haze correction has been tested onto real hazy images. The result shows that the method was able to remove most haze from the hazy images effectively. The developed method is expected able to increase the reliability of remote sensing data in haze affected areas. However, the performance of the method might be hampered by heterogenous haze since the haze mean reflectance cannot be easily subtracted. Future studies should seriously look into this matter to ensure remote sensing imagery can be used optimally in various applications for the benefit of mankind.
ACKNOWLEDGMENT

The Authors are grateful to Universiti Teknikal Malaysia Melaka (UTeM) for providing the funding and support for this research.

REFERENCES


