Haze Optimized Transform on LANDSAT 8 Imagery for Thin Cloud Detection and Removal

Mark Edwin A. Tupas

Department of Geodetic Engineering
University of the Philippines Diliman, 1101 Quezon City, Philippines

Abstract—Landsat images, being optically captured, contain one of the most fundamental remote sensing issues-- cloud and haze contamination. Various algorithms have been developed through the years to correct haze contamination and maximize the use of archived Landsat images since its launch in 1972. One such algorithm is the Haze Optimized Transform (HOT). HOT identifies spatially varying haze thickness on the premise that clear sky conditions can be characterized from the regression of highly correlated blue and red bands; and features diverging from this relation indicate thickness of haze using their orthogonal distance (HOT values) from clear sky function. A modified Dark Object Subtraction is then performed based on histogram matching per HOT value versus the clear sky case.

This paper presents modifications in applying the HOT algorithm considering the effects of increased radiometric resolution and new coastal blue band in Landsat 8 were tested on two separate images with different dates of acquisition from a test site in the Davao Oriental province of southern Philippines, which was selected due to its prevalent cloud cover condition throughout the year. The effects of the increased radiometric resolution and new coastal blue band in Landsat 8 were tested on two separate images with different dates of acquisition. Haze correction using the coastal blue band demonstrates noticeable difference in adjustment for certain land cover types. On the other hand, the increase in radiometric resolution shows exponential effects to HOT value ranges which translates to finer haze depth estimation but at the expense of performance. Moreover, applying the algorithm demonstrates a higher rate of over correction, which is then compensated by applying a clear aerosols fraction adjustment.

The corrected images are then further processed to compute Normalized Difference Vegetation Index and Supervised Classification to show the effectiveness of the HOT correction algorithm. This study shows that the HOT algorithm with the presented modifications can be efficiently and effectively implemented on Landsat 8 images, and obtain the desired results.

Keywords— Remote Sensing, Relative Radiometric Correction, Haze Removal, Landsat 8, Haze Optimized Transform

1. INTRODUCTION

1.1 Background of the Study

The Landsat program, one of the legacy programs in remote sensing, has provided satellite imagery with almost worldwide coverage of the earth's surface since 1972 (USGS, 2012). In the last decade Landsat images has been available to all users at no cost, this has made it as one of the most used satellite imagery for research, development and operational use. The latest iteration, Landsat 8, was

---

Correspondence to: Mark Edwin A. Tupas | Department of Geodetic Engineering, University of the Philippines Diliman, 1101 Quezon City, Philippines | Email: matupas@up.edu.ph

launched on Feb. 11, 2013. Landsat 8 images have four visible bands, along with the three heritage optical bands for Red, Green, and Blue the new coastal blue band (coastal aerosol) was added for near shore applications (USGS, 2012). Another variation of significant interest is Landsat 8 images' radiometric resolution of 16-bit unsigned integers (USGS, 2014), provides a higher digital number (DN) range of 0 - 65535 which is a great improvement from the its predecessors that have only 0-255 range of DN values. Further changes in Landsat 8 include the addition of 9th band, which is described as the cirrus band which can be used to identify cirrus clouds.

Cloud and haze contamination has always been an issue when dealing with remotely sensed satellite imagery especially those acquired in the optical region. On satellite images thick clouds completely covers the underlying ground features, while other atmospheric contaminants such as haze partially covers objects on the surface. Zhang et al. (2002) defines haze as spatially varying, semitransparent cloud and/or aerosol layers on an image. Clouds and haze share a similar characteristic, in that both increase the radiometric (DN) values of pixels regardless of the occluded feature. This distorts the expected response of a given pixel, making spectral analysis difficult, if not impossible. While haze only partially covers the underlying pixels it also distorts an image's contrast (Du et al., 2002) thus image analysis methodologies such as spectral indices e.g. NDVI and image classification are largely affected.

Optical remote sensing images for most applications require clear skies to be useful. Even with the increasing number of available satellites and remote sensing missions, it is still a challenge to acquire imagery with low cloud contamination. To be able to maximize the use of cloud contaminated satellite images various algorithms have been developed for haze identification, and removal or correction.

1.2 Review of Related Literature

One of Landsat 8's aims is to continue the worldwide coverage it has started for earth observation (USGS, 2012). However, Landsat 8 represents a significant change over its predecessors for being the first to utilize a push broom sensor (Knight and Kvaran, 2014) among other changes. In terms of spectral response e. g. TOA reflectance a systematic difference between the previous two Landsat iterations (Flood, 2014). Such changes in its imaging systems prompt the need for evaluation of spectral response based algorithm applicability.

The ideal case for correcting haze contaminated images is by absolute radiometric correction, this requires the complete solution to the radiative transfer problem which requires in situ measurements to ascertain the actual atmospheric conditions prevailing at the time of image capture (Du et al., 2002). However, in situ measurements require a priori knowledge of image acquisition, if not constant measurements of areas of interest; because of this difficult requirement, relative radiometric correction is more often employed to adjust for haze contamination. Relative radiometric correction implies the use of correction parameters solely based on the relative DN values within an image scene (He et al., 2009).

Because of this, this method is more often applied to images taken from curated and archived image repositories.

Some traditional algorithms assume homogeneous haze depth all throughout an image. Most notable is the DOS (Dark Object Subtraction) which assumes that, within an image, there exists a dark object with theoretically zero DN value that has been raised due to the additive effect of haze contamination (Chavez, 1988). Thus, the DN value of that pixel is the estimated effect of haze throughout the image. However, haze on an image is seldom homogeneous, and often has spatially varying thickness. The crux in removing heterogeneous haze largely depends on the initial step which is its identification and characterization. For this case, the DOS concept is modified by succeeding algorithms that utilize histogram matching of haze thickness level versus the ideal clear sky case (Chavez, 1988; Ricther, 1996, Zhang, 2002; Du, 2002; Li et al., 2014).

The conventional method in identifying objects in optical satellite imagery leverages the different spectral characteristics of image features to differentiate them from one another. Haze and thick cloud contamination raises reflectance, and consequently DN values in raw images, due to increased scattering of incident path radiance. Aside from high DN values in the optical regions, thick clouds are characterized by low temperatures (low DN values) in the thermal region, this characteristic makes it ideal to differentiate thick clouds from white objects on an image such as built-up areas, white sand, etc. However, thin clouds or haze exhibits the same rise in DN values in the optical region, do not always have low temperatures, making haze identification more complicated (Li et al., 2014).

To leverage spectral characteristics, one common theme in haze identification is the utilization of subsequent thresholds on varying band combinations and transformations to isolate clouds (Richter, 1996; Irish et al., 2006). On the other hand spatial characteristics have also been utilized in object based algorithms for haze identification such as HAWAT which uses wavelet decomposition based approach (Du et al., 2002), image segmentation based on thresholds (Zhu and Woodcock, 2012), and localized dark object selection to generate Haze Thickness Maps (HTM) (Makarau et al., 2014).

A number of exhaustive cloud detection and characterization processes have been employed on Landsat imagery, most notable of which is the Automated Cloud Cover Assessment (ACCA) (Irish et al., 2006) and Function of mask (Fmask)by Zhu and Woodcock (2012). While, ACCA is mainly used to generate nominal cloud cover percentages without the cloud mask output, whereas Fmask aims to generate a binary cloud mask. However, these algorithms specifically utilize most of Landsat bands including optical bands, near infrared, and brightness temperatures, and are tuned to their specific spectral responses making them not directly adaptable to other images.

The optical region specifically the red and blue wavelength bands of satellite imagery are highly correlated (Zhang, 2002); and the increase in DN values brought about by cloud contamination adversely affects this relationship. In this regard, several algorithms have utilized this premise to identify and characterize cloud contamination based on image transformations with bands that highly correlate to haze. This was first observed from Richter (1996) when he utilized the haze band from the

Tasseled Cap Transformation (TCT) while a subsequent study by Li et al. (2014), suggests that the 2nd component of a Principal Component Transformation (PCT) demonstrates the same feat.

While TCT and PCT transformations were not originally targeted for haze identification, Zhang et al. (2002) improves on the TCT to devise the Haze Optimized Transformation (HOT). With HOT spatial distributions and haze depth becomes more apparent to allow for better characterization. Guidon et al. (2003) argues further the effectiveness of HOT as tested from several Landsat Imagery and comparisons with other haze removal techniques. Moreover, HOT has also been applied to various imagery of varying spectral and spatial characteristics; such as high resolution images like IKONOS (Dal Moro, 2007), Quickbird, and even on CBERS imagery (Wen et al., 2009). More recently, the HOT algorithm has been extended to algorithms characterize imagery shadows (Zhang et al., 2014) as well as detailed aerosol characterization (Razali, 2015).

Of all the available algorithms reviewed, HOT seems to provide a quick and robust methodology for both characterization and haze removal for optical images without the need for additional information making it a via procedure even for RGB images. This paper will therefore focus on the application of HOT procedure in removing haze in Landsat 8 imagery.

1.3 Objectives

The main objective of this study is to apply the aforementioned Haze Optimized Transform for Landsat 8 imagery and determine its effectiveness in detecting and removing cloud and haze from the images. Since the main goal of haze removal is to correct images that are otherwise useless, the results can be gauged by subsequent processing of those imagery. Spectral analyses, namely, image classification and Normalized Difference Vegetation Index (NDVI) will be done to demonstrate the effectiveness of HOT in cloud detection and removal.

To reach this main goal the following specific objectives were set forth. First, to determine if coastal blue (aerosol) band in Landsat 8 is a viable substitute to the heritage blue band in computing the HOT values. Due to the similarity of both bands, it is surmised that coastal blue band may be a suitable substitute to the heritage blue band. On the other hand, the lower wavelengths of coastal blue band are more affected by scattering due to aerosols in the atmosphere, it is hypothesized that Band 1 of Landsat 8 may result in more sensitive HOT image. Second, to determine the effect of increased radiometric resolution to the HOT algorithm; and to develop a better histogram matching for correction function in lieu of increased dynamic range of HOT values. It is surmised that an increase in DN values would result in finer haze depth estimation; hence, better haze correction.

2. METHODOLOGY

2.1 Test Site

The study site for this study is located in Davao oriental Province of southern Philippines, covering the municipalities of Cateel, Bagangga and Boston, shown in *Figure 1*. The area is located on the eastern seaboard of the island of Mindanao, facing the Pacific Ocean, and known to be susceptible to typhoons and Inter-Tropical Convergence Zones (ITCZ). The test site is of interest to hydrologic studies, specifically flooding caused by extreme rainfall events. The prevalent land cover and topography consists of mountain ranges with mostly primary forests, to flat lands with agricultural lands and coastal areas with a small number of rural settlements. The test area is found between the longitudes from 126' 8" 7.14º to 126' 41" 16.94º, and latitudes from 7' 26" 22.10º to 7' 55" 56.55º and is covered by a single Landsat scene.
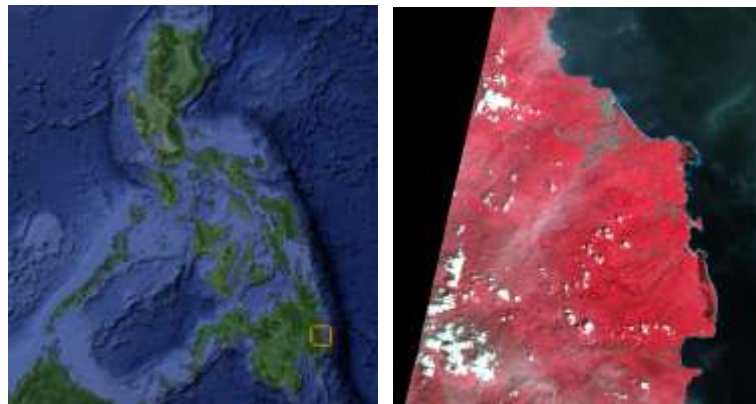

**Figure 1.** The Test Site: Landsat Imagery of Davao Oriental

2.2 Materials

Archived 16-bit GeoTIFF format Landsat 8 images were downloaded from Earth Explorer interface of USGS (http://earthexplorer.usgs.gov/) and used to test the algorithm. Like its predecessors capture images over a 185 km swath and gather data at an altitude of 705 km. The study area is covered by the Landsat image in the World Reference System (WRS- 2) path 111, rows 55. Landsat 8 is composed of Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) with spectral bands shown in Table 2. For this study images utilized for haze removal and subsequent processes were limited to OLI bands 1 – 5 of images from October 2, 2014, and January 22, 2015 (false color image shown in *Figure 1*), dates were utilized in this study, all of which are covered by thick clouds and haze to test the proposed procedure.

Table 2. Landsat 8 (OLI) Spectral Bands

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43 - 0.45	30
Band 2 - Blue	0.45 - 0.51	30
Band 3 - Green	0.53 - 0.59	30
Band 4 - Red	0.64 - 0.67	30
Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
Band 6 - SWIR 1	1.57 - 1.65	30
Band 7 - SWIR 2	2.11 - 2.29	30
Band 8 - Panchromatic	0.50 - 0.68	15
Band 9 - Cirrus	1.36 - 1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

HOT has been implemented by various Remote Sensing softwares such as Panchroma (*http://www.pancroma.com*) and ENVI/IDL extension (Hu et al., 2009). However, for his study the modifications to the algorithm were implement using python 2.7 scripts using GDAL and Numpy libraries for being free and open source, and readily available. The implementation was wrapped in a QGIS python plug-in for ease of use, which is currently under development and called Dehaze Image (shown in Figure 2), for more convenient processing. The processing was performed using a Ubuntu 14.10 computer running on 8Gb RAM and Intel® Core™ i5-2450M CPU @ 2.50GHz $\times 4$ processor.
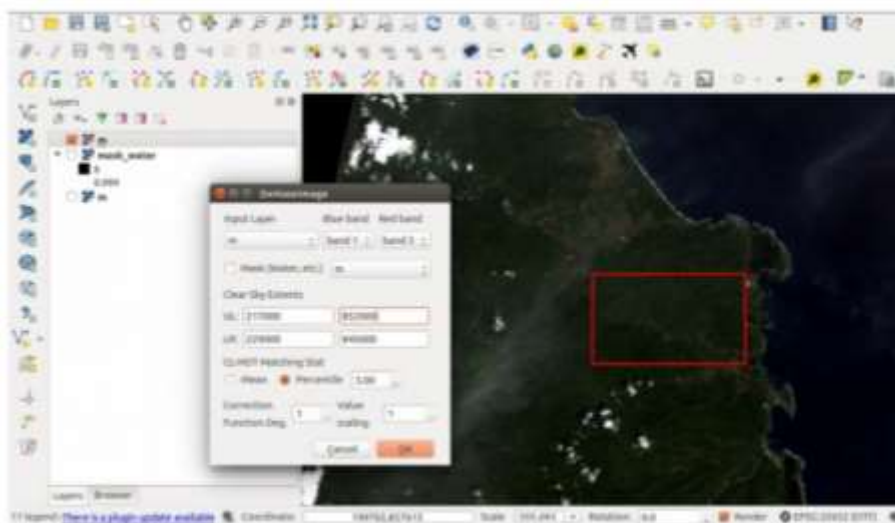
Figure 2. DeHaze Image, QGIS plug-in

2.3 Haze Optimized Transform (Theoretical Framework)

Zang et. al (2002) point out that at clear sky conditions there is high correlation between Red and Blue DN band values for most land cover types. They argued that a clear sky line (CL) based on these two bands can be regressed from a set of pixels selected within the scene, and hazy pixels would deviate from this line. The transformation determines the orthogonal distance from the computed clear sky line. This displacement from the clear sky line can be computed using Equation (1), and from here on be referred to as HOT values. HOT values then relatively indicates the thickness of haze, with higher HOT values indicating larger haze depth.

$$HOT = DN\_Blue(\sin\Theta) - DN\_Red(\cos\Theta) \qquad\qquad (1)$$

where DN_Blue and DN_Red, denote the digital numbers for the respective bands, and Θ the slope of the clear sky line.

HOT haze removal algorithm can then be thought of as a supervised algorithm in the sense that clear sky pixels are selected to generate the regression of the clear sky line previously mentioned. Since this selection procedure can be subjective, an understanding of clear sky pixels is required, which in some cases requires a priori knowledge of the land cover class present, this has been cited as a limitation of HOT (Makarau et al., 2014). To reduce subjectivity, analysis of correlation such as Pearson's r between the red and blue bands of selected clear sky pixels can be performed. Selection with the highest Pearson's r will then be used to perform the regression of clear sky line. On the other hand, He et al. (2009) suggest adopting the whole image scene when a clear sky region is not present.

As pointed out by Zhang et al. (2003) the algorithm does not fare well for certain land cover types, such as built-up, bare soil, and silted water. They noted that underlying silted water and built-up areas tends to have elevated HOT values, whereas bare soils have slightly lower HOT values. It can be surmised that HOT performs best for vegetated land cover types. To address this concern, a modification to the original algorithm was developed where bad faring land cover types are masked before the actual HOT algorithm is implemented; after which an additional step to correct the masked pixels is performed (Dal Moro, 2007). Another modification to the HOT process, which they call advanced HOT or AHOT varies the computation by adjusting HOT values bias per land cover class (He et al., 2009).

$$HOT = DN\_Blue(\sin\Theta) - DN\_Red(\cos\Theta) - a \qquad\qquad (2)$$

where DN_Blue and DN_Red, denote the digital numbers for the respective bands, Θ the slope of the clear sky line, and *a* is the intercept of the clear sky line.

Equation (2) indicates the modified HOT calculations for the AHOT process. The intercept ofthe clear sky line is utilized aside from the slope angle. The intercept was added to the calculations to shift the HOT values of clear sky pixels (mainly vegetated and asphalt land cover types) to almost zero (He et al., 2009).

Based on the original algorithm adjustment values were computed by matching the lowest DN value for a HOT value histogram to the lowest value in the clear sky histogram. The difference (which serves as the correction) is to be subtracted from the original DN value of the image (Zhang, 2002). Alternatives to this have been presented based on the premise that lower bound values are not stable, instead other histogram statistics were explored such as mean, percentile statistics (Dal Moro, 2007), and VCP (Virtual Cloud Point) or the mean of minimum and maximum regressed lines (He, 2009).

Zhang et al. (2002) point out that a linear function is a 'good first estimate' of the correction although no further possible correction functions were discussed. Further review of literature on this aspect has turned futile, thus from this it can be presumed that linear regression of the function may still be the best estimator.

As observed, even non-hazy pixels could possibly have significant HOT values. The application of the correction then should be limited to that of the hazy pixels. While the boundary of hazy and non-hazy is not exactly defined using the HOT algorithm, non-application of DOS on "clear sky pixels" results in uneven image after correction. On the other hand, non-removal of clear sky pixels results in overcorrection, although all HOT values resulted in a left shift in the dehazed histogram. As observed from data presented by from literature (Zhang, 2002; Dal Moro, 2007; He et al., 2009) dehazed histogram is often slightly shifted to the left, or moved to lower values. Makarau et al, (2014) also cites over correction in the intermediate steps of their HTM algorithm, the overcorrection is pointed at removal of clear sky aerosols. They surmise that even clear sky imagery have aerosol layer and the registered correction over compensates and removes this layer as well. It is further cited that the clear scene fraction is negligible in lower radiometric resolutions it is more apparent 16 bit images such as Landsat 8. It is supposed that HOT values computed for CL pixels is can be attributed to the same "clear sky aerosols fraction".

2.4 Methods

In the following section, details of the haze removal strategy is presented and discussed.

Subsets of the Landsat 8 scenes described above were prepared for HOT processing. To investigate the effect of radiometric resolution on the HOT images, varying radiometric resolution images were generated from top of the atmosphere reflectances images. This was performed by multiplying the TOA reflectance images by integer values: 100, 255, 1000, 10000 and 65535 respectively. Using a common set of clear sky pixels, analysis of HOT values generated from the

original heritage blue vs red bands, and the HOT calculation using new coastal blue vs red bands were performed by image difference of the computed dehazed imagery. The at each run of varying radiometric resolution and blue band were performed using the general methodology outline as follows.
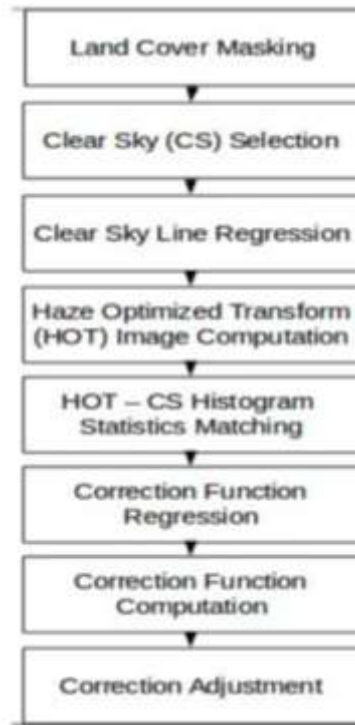

Figure 3. HOT Methodology

Land Cover Masking – Similar to Dal Moro's HOT implementation where land cover pixels with notorious HOT responses are initially filtered out (Dal Moro, 2007). In this particular study, water pixels were masked out due to the erroneous HOT values and due to the relative size of water pixels in the image scene. Significant drift in the correction function is expected due to the large number of elevated HOT values from silted water, hence they were target for exclusion. On the other hand, there are almost insignificant built-up areas in the study site which were not included in the masking.

Clear Sky Selection – Selection of clear sky pixel were done by rectangular bounding boxes as assisted by software used. Image enhancements and stretching were performed to visually isolate non-hazy areas. Instead of measuring Pearson's R to vet clear sky pixels, a number of iterations were performed until HOT images with sufficient contrast is achieved. This is performed by investigating HOT value ranges from the results after each iteration and histogram stretching. The a clear sky area per image was selected to be consistent for the different iterations.

Clear Sky Line Regression – from the selected pixels linear regression of the blue and red bands are used to calculate the clear sky line, more specifically the clear sky line slope angle (*Θ in Formula 1*). Scatter plots of the blue and red band combinations suggests a linear relationship between each. The coastal blue band is tested and compared with the heritage blue band. Initial comparison of correlation and regressed slope angle is taken to ascertain difference from heritage blue band.

HOT Image Computation – A HOT image is then generated by applying Formula 1. Instead of using the clear sky line intercept (as indicated by Formula 2). However, the HOT values were shifted using the minimum HOT value in the image. This was done to optimize the process in reducing the HOT values to be histogram matched. HOT images using both blue bands were compared and contrasted.

HOT – CS Histogram Statistics Matching – matching statistics taken form histograms of per HOT value for each optical band is compared with its corresponding clear sky case. The selection of matching statistic is viewed as a parameter for the algorithm execution, this is in part due to the inconclusive analysis on best performing matching statistic. The following matching statistics were explored for implementation: 1) Minimum value or histogram lower bound, 2) Mean values, 3) Percentile which indicates the relative position in percentage of the cumulative counts of values. Percentile is the most flexible since using 50th percentile indicates the median and 100th percentile indicates the maximum.

In terms of computational complexity, the analysis of the histogram matching would indicate that the minimum value is the least complex which requires on average only less than a single pass of all values to perform matching. Followed by mean statistic which requires exactly one pass to perform the matching this indicates a complexity of N. On the other hand, percentile statistics requires more than one pass. The straight forward algorithms in computing a samples percentile produces a complexity of N^2, better implementations suggest an NlogN complexity. This indicates an almost negligible effect in actual computational times.

(a) minimum   (b) mean

(c) maximum   (d) 5th percentile
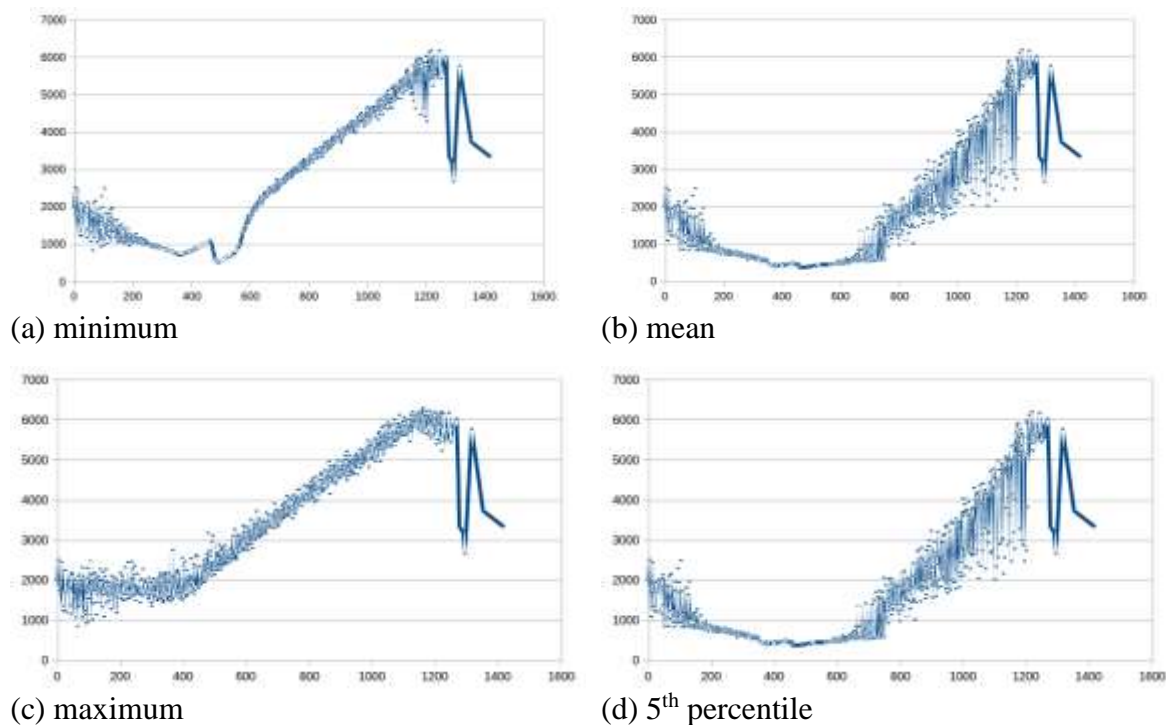
Figure 4. Scatter plots of histogram matching. X-axis: HOT values, Y-axis: Correction values (10000 DN image)

Figure 4, illustrates scatter plots from the various matching statistics of the January 2015 image. Surprisingly, the minimum value demonstrates the lowest variability among the tested statistics. In the same manner lower value percentile like the 5th percentile indicated up to the median shows similar scatter plots than that of the mean statistic. Moreover, the maximum statistics the least similarity with the other plots. However, while the minimum function represent less variability between local points, it shows relatively larger discrepancy with best fit lines due to well defined inflections at the 400 – 600 DN value range where most of the pixels are concentrated. For consistency in presentation the 5th percentile solution will be shown for subsequent presented results.

Correction Function Regression – The difference in histogram matches are linearly regressed to determine a correction function with HOT value as the required parameter. The assertions of linear correction function were tested for the same imagery. Regardless of the radiometric resolution similar graphs were computed, hence only 10000 DN image was shown for clarity. As shown in Figure 4, lower HOT values show a decreasing trend regardless of the histogram matching used; and at some point it will up tick to an increasing trend. Overall the DN values indicate an increasing trend when a trend line is used, further analysis of the samples indicates that most of the haze samples are concentrated near the inflection point from Figure 8, at around 400 - 500 DN value range.

Application of Correction Function – The computed function is applied using the HOT image, subtracting the results to correct the additive effects of haze. While the graph of the HOT vs DN values suggests that piecewise or even a parabolic correction function may be a better fit due to an inflection point, majority of the hazy pixels fall within the initial up tick, making the linear assumption valid.
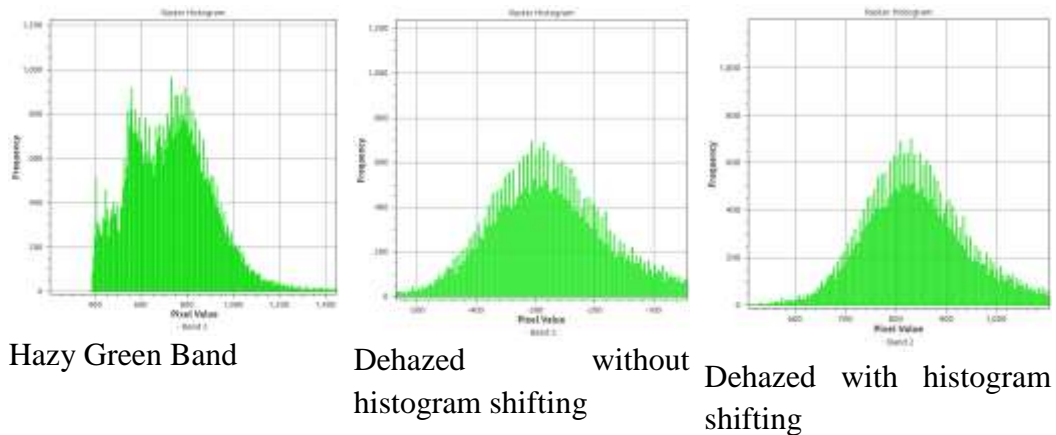

Hazy Green Band   Dehazed   without   Dehazed   with   histogram
                      histogram shifting   shifting

Figure 5. Histogram Shifting

Histogram Shifting – The resulting histogram is compensated for the over correction, where the values are shifted based on the clear sky histogram. Figure 5 illustrate the concerns with overcorrection of DN values. Compensation based on the concept of clear sky aerosol function (Makarau et al., 2014) is used to perform the histogram shifting. For simplicity the histogram shift was derived from the mean value matching of the selected clear sky pixels and the intermediate dehazed pixels.

3. RESULTS AND DISCUSSION

3.1 Haze Removal Results

After variations to algorithm parameters such as radiometric resolution, blue band, histogram matching statistics, and correction function were performed. This section details the most viable results obtained and inferences on the parameter utilized are made. Figure 6 shows haze removal of the 16-bit January 2015 Landsat 8 image using heritage blue band and linear regression using 5th percentile histogram matching.
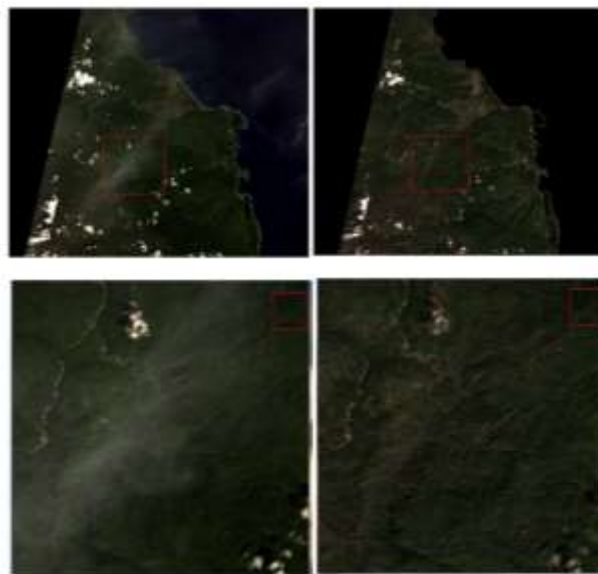

Figure 6. Haze Removal Results

It is observed from the per band images, correction function, and RGB true color image that Band 1 or the blue band was corrected the most, while decreasing correction in both the subsequent bands this is consistent with literature indicating lower wavelength bands having higher susceptibility to haze. A slight color distortion of RGB image is observed, while the individual band correction maybe deemed effective the process may have affected the relative responses between band histograms, this surmised as caused by the simple histogram shifts performed where only the mean.

3.2 Radiometric Effects

Corrections of images of varying radiometric resolution are shown in Figure 7. Due to the transformation procedure the number of bins of resulting HOT images or as supposed the discreet levels of haze thickness were inferred. Figure 8 shows a direct exponential relationship with the original resolution to that of the resulting HOT levels. Reduced number of discreet levels implies over or under

corrections of pixels broadly grouped by the procedure resulting in blocky haze corrections. As observed from the processed images 1000 bin images already show similar refined detail to that of the 65535 bin imagery.
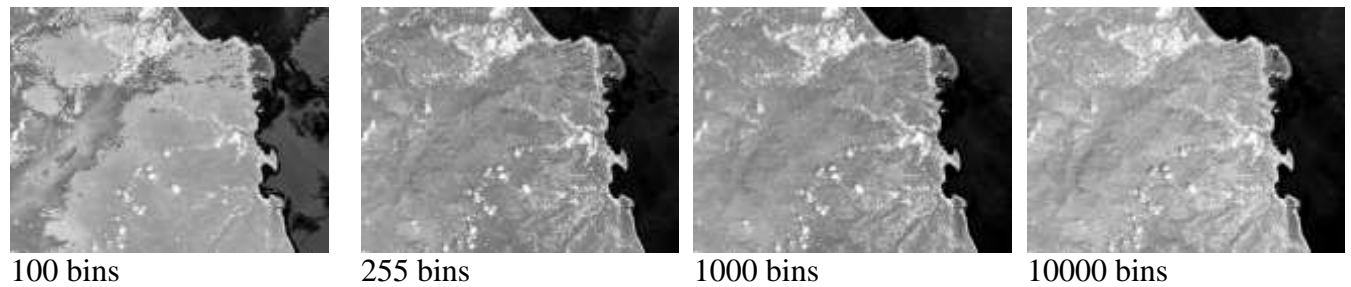

100 bins   255 bins   1000 bins   10000 bins

Figure 7. Radiometric Resolution Effects (Green Band)
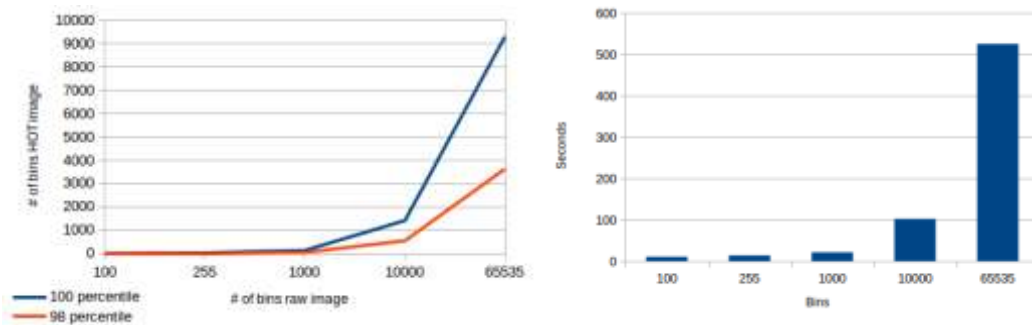

Figure 8. Resulting HOT Image bins and Algorithm Performance versus raw image bins

As expected the haze depth variation is better captured by higher radiometric resolutions. Furthermore, the exponential increase in the resulting HOT levels indicates better effectiveness of the algorithm. However, the algorithm performance is directly related to the number of bins due to the nature of histogram matching procedure performed. This suggests that increasing the radiometric resolution of images increases effectiveness of the algorithm in terms of characterizing haze effects. While Landsat 8 images can be processed at default radiometric resolution, the effectiveness is not much greater than that of lower scaled image values.
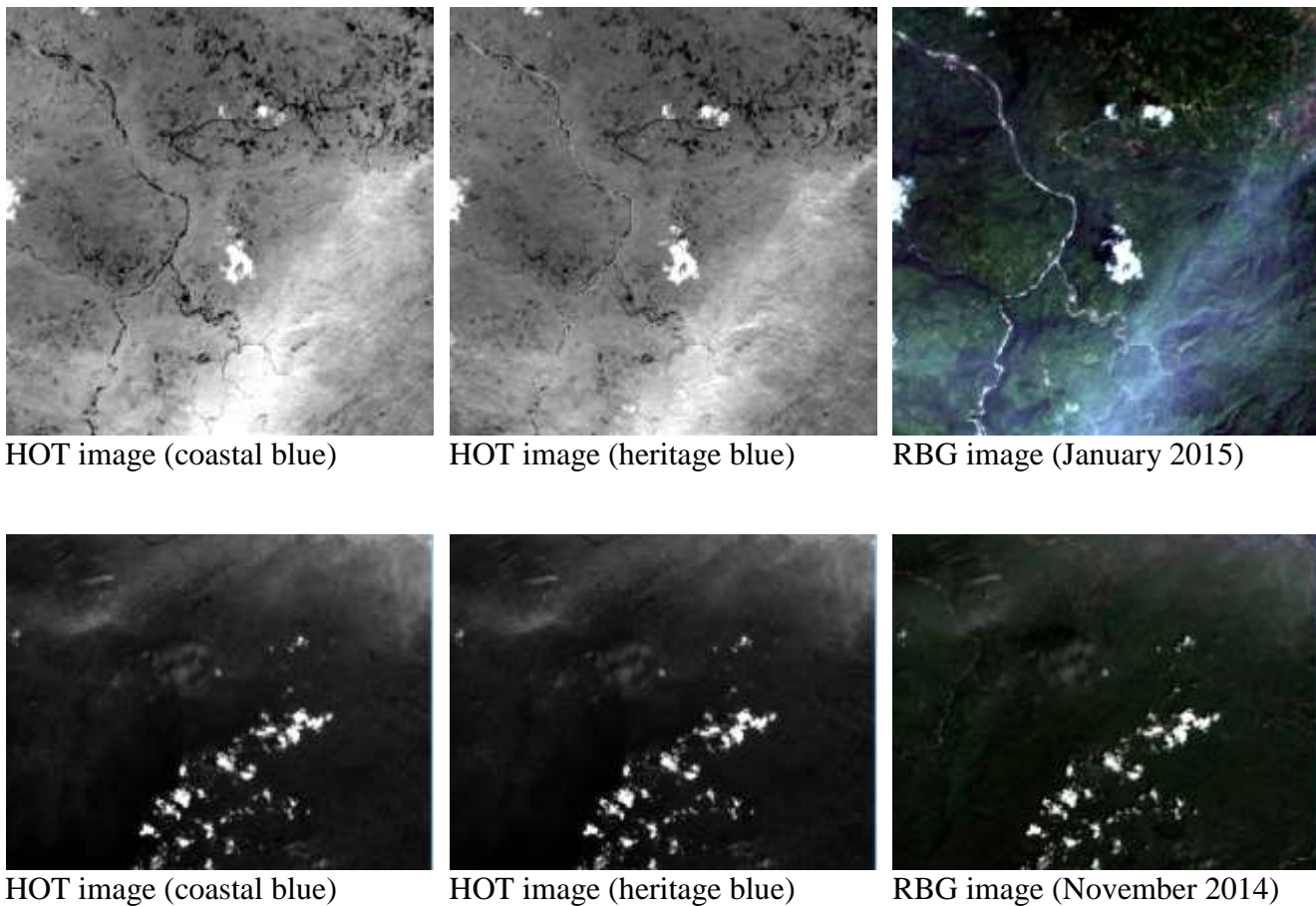
3.2 Coastal Blue vs. Heritage Blue

Using the the native Landsat 8 images clipped and masked to include OLI bands were used to test coastal blue HOT images. The correlation (Pearson's r) of red and blue band compared with red and coastal blue band shows similarly high correlation (see Table 3). The heritage blue band exhibits slightly higher correlation than that of the coastal blue band. Linear regression of the given indicates that there is a significant difference in the slope angle computed. This suggests that HOT images from using coastal blue would have significantly change as well.

Table 3. *Comparison Pearson's R and Clear Sky Line Slope Angle*

	coastal	heritage	
Pearson's R	*0.98418*	*0.98900*	Jan 2015
	0.99519	*0.99683*	Nov 2014
Slope angle (degrees)	*54.40537*	*53.69029*	Jan 2015
	49.68670	*47.84904*	Nov 2014

HOT images were computed for both coastal and heritage blues. Visual inspection of the images suggests higher contrast for heritage blue HOT image. It is also worth pointing out that land cover classes exhibiting elevated HOT values such as silted water and built-up areas are less elevated in coastal blue HOT images. In terms of range of values, no significant conclusions can be made, on average the Jan 2015 image had a 38.45% decrease in HOT value range from using coastal blue band, while a 0.55% increase in HOT value ranges for the Nov. 2015 image.

HOT image (coastal blue)   HOT image (heritage blue)   RBG image (January 2015)

HOT image (coastal blue)   HOT image (heritage blue)   RBG image (November 2014)

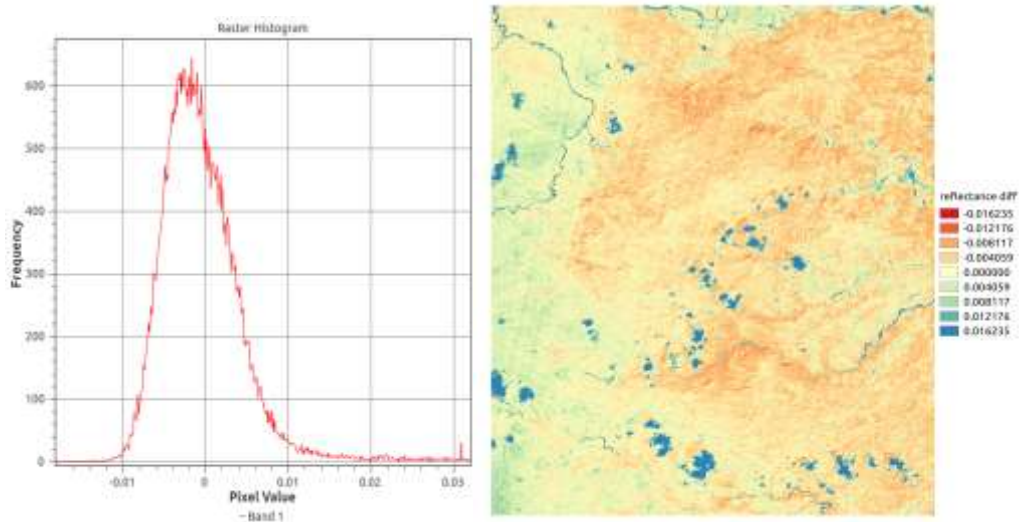***Figure 9.*** *Comparison of HOT images generated*

Figure 10. Reflectance Difference (DH image coastal blue – DH image heritage)

The haze correction procedure was continued without masking in land water features for both HOT images, the resulting dehazed imagery was then calibrated to TOA reflectances. A difference image, where heritage blue HOT dehazed image was subtracted from coastal blue HOT dehazed image, was computed from the reflectance images can be seen in Figure 9. Most pixels agree or have similar corrected values, surprisingly most of these pixels are pixels with which are thought of as hazy pixels as observed from the HOT image.

Moreover, the difference in the images are noticeable to the extent that land cover features can be identified such as, water pixels, clouds, and fallow or bare soil areas. This suggests that coastal blue band HOT has significant change in its response that are dependent on these land cover types. The largest positive difference were observed on clouds and silted rivers, the land cover features that exhibit elevated values for the original HOT. This indicates greater correction for these land cover types happens in the heritage blue band, conversely the coastal HOT have lesser response to higher haze levels. Utilizing the coastal blue band therefore reduces over correction at the least for water land cover. However due to limited built-up pixels in the image scene it's effect was not investigated. In all reduced correction (negative values in Figure 9) was observed for most other land cover types, based on the slightly skewed histogram.
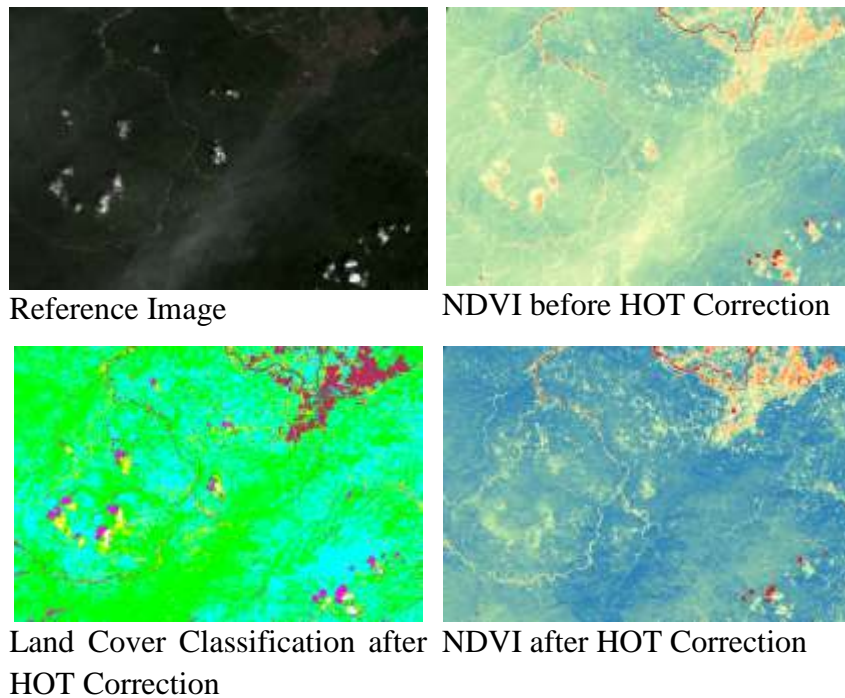
3.4 Post Processing

Reference Image   NDVI before HOT Correction

Land Cover Classification after   NDVI after HOT Correction
HOT Correction

Figure 11. Post HOT Correction Image Analysis

The end goal haze removal is to prepare the haze filled images for further processing. To gauge the HOT methodology in haze removal effectiveness subsequent NDVI and image classification was performed. From the image it can be observed that the process was able to correct most thin clouds existing in the image. While thick cloud fringes were corrected the most of cloud remains intact. For most objects affected by haze, greatly improved NDVI and classification results was observed, where hazy pixels originally indicating mostly low NDVI values now spatially varies according to the underlying land cover.

4. CONCLUSION AND RECOMMENDATIONS

The study shows the applicability of HOT thin cloud removal algorithm on Landsat 8 images. The following conclusion are made based on the performed methods:

Analysis of matching statistic and correction function are inconclusive and in some cases inconsistent with literature results. The best results from were obtained from an iteration of the process with varying parameter inputs (see Fig. 4). In general, 5th percentile with linear correction function was used for most of the processing performed. The application of histogram shifting compensates for the nominal DN values of corrected pixels; however, this does not address apparent color distortion.

The default radiometric resolution of Landsat 8 results in cumbersome processing of HOT images due to the histogram matching step of the algorithm. As observed increasing radiometric resolution increases the HOT levels computed, this result is finer correction which is desired. However, corrected images using the natural 16-bit resolution displays similar results to that of images rescaled to 10000 bin resolution, this suggests algorithm performance can be optimized by rescaling Landsat 8 radiometric resolution and attain similar results.

Coastal blue band of Landsat 8 shows significant correlation to the red band to merit consideration for HOT application. It was shown that coastal blue band of Landsat 8 imagery is a viable substitute to heritage blue band for HOT calculations coastal blue band of Landsat 8 shows significant correlation to the red band to merit consideration for HOT application, this results in similar images with slight variations that is observed to be land cover dependent. It was also observed that it performs better for water land cover for the images utilized for this study. In general, there is a slight increase in haze correction when coastal blue was used.

The corrected images are then used for calculation of NDVI and Image classification to show HOT correction algorithms effectiveness. In all, it can be concluded that the HOT algorithm can be efficiently and effectively implemented on Landsat 8 images with the demonstrated modifications.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] U.S. Geological Survey, 2012, Landsat-A Global Land-Imaging Mission: U.S. Geological Survey Fact Sheet 2012–3072, 4 p. (Revised July 19, 2012).

[2] USGS/EROS. LSDS-809. (July 2014) Landsat 8 (L8) Level 1 (L1) Data Format Control Book (DFCB). Ver. 08. Retrieved: January 11, 2015 from: http://landsat.usgs.gov/documents/Landsat8_Level1_Data_Format_Control_Book_LSDS-809.pdf

[3] Chavez , P.S. J R , 1988, An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment, 24, pp. 459–479.

[4] R. Richter, 1996. Atmospheric correction of satellite data with haze removal including a haze/clear transition region. Computer and Geosciences, 22, 675 – 681.

[5] Y. Du, B. Guindon, and J. Cihlar, 2002. Haze detection and removal in high resolution satellite image with wavelet analysis, IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 1, pp. 210–217.

[6] Y. Zhang, B. Guindon, and J. Cihlar, 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images, Remote Sens. Environ., vol. 82, no. 2/3, pp. 173–187.

[7] B. Guindon, and Y. Zhang, 2002. Robust haze removal: an integral processing component in satellite-based land cover mapping. Proceedings, Geospatial Theory, Processing and Applications, ISPRS Commission IV Symposium.

[8] Y. Zhang, and B. Guindon, 2003. Quantitative assessment of a haze suppression methodology for satellite Imagery: effect on land cover classification performance. IEEE Transactions on Geoscience and Remote Sensing, Vol. 41, pp. 1082-1089.

[9] R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, 2006. Characterization of the Landsat-7 ETM + Automated Cloud-Cover Assessment (ACCA) algorithm. Photogrammetric Engineering and Remote Sensing, 72(10), 1179–1188.

[10]   G. Dal Moro, and L. Halounova, 2007. Haze Removal and Data Calibration for High-Resolution Satellite Data. International Journal of Remote Sensing. Vol. 28 , Issue 10, pp 2187-2205.

[11]   J. Hu, W. Chen, X. Li, et al., 2009. A haze removal module for mutlispectral satellite imagery. Urban Remote Sensing Event, Vol 9, pp. 1-4.

[12]   X. Wen and X. Yang, 2009. Haze removal from the visible bands of CBERS remote sensing data. Proc.Int. Conf. Industrial Inform. Syst., pp.118 -121.

[13]   Xing Yuan He , Jian Bo Hu , Wei Chen and Xiao Yu Li, 2010. Haze removal based on advanced haze-optimized transformation (AHOT) for multispectral imagery, International Journal of Remote Sensing, 31:20, 5331-5348

[14]   Z. Zhu and C. E. Woodcock, 2012. Object-based cloud and cloud shadow detection in Landsat imagery,” Remote Sens. Environ., vol. 118, pp. 83–94.

[15]   Huifang Li, Liangpei Zhang, and Huanfeng Shen, 2014. A Principal Component Based Haze Masking Method for Visible Images. IEEE Geoscience And Remote Sensing Letters, Vol. 11, No. 5.

[16]   E. Knight and G. Kvaran, 2014. Landsat-8 Operational Land Imager Design, Characterization and Performance. Remote Sens. 2014, 6, 10286-10305; doi:10.3390/rs61110286

[17]   N. Flood, 2014. Continuity of Reflectance Data between Landsat-7 ETM+ and Landsat-8 OLI, for Both Top-of-Atmosphere and Surface Reflectance: A Study in the Australian Landscape. Remote Sens. 2014, 6, 7952-7970; doi:10.3390/rs6097952

[18]   Makarau, R. Richter, R. Muller, and P. Reinartz, 2014. Haze Detection and Removal in Remotely Sensed Multispectral Imagery. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 9, pp. 5895 – 5905.

[19]   Y. Zhang, B. Guindon, and X. Li, 2014. A Robust Approach for Object-Based Detection and Radiometric Characterization of Cloud Shadow Using Haze Optimized Transformation. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 9, pp. 5540 – 5545.

[20]   M. F. Razali, A. Ahmad, O. Mohd, N. Iman, and S. Bahari, 2015. Quantifying Haze from Satellite Using Haze Optimized Transformation (HOT). Applied Mathematical Sciences, Vol. 9, 2015, no. 29, 1407 - 1416