Visibility Estimation Due To Forest Fire Smoke Using Backward Elimination Multiple Regression of Himawari_8 Satellite Data over Sumatera and Borneo Island Indonesia

H Ismanto\textsuperscript{1,2}, H Hartono\textsuperscript{2}, M A Marfai\textsuperscript{2*}

\textsuperscript{1} Center for Aeronautical Meteorology, Agency of Meteorological Climatological and Geophysical (BMKG), Kemayoran, Jakarta 10720, Indonesia
\textsuperscript{2} Fakultas Geografi, Universitas Gadjah Mada, Bulaksumur, Yogyakarta 55281, Indonesia

*Corresponding author: arismarfai@ugm.ac.id

Abstract. Land and forest fires smoke has been regularly revealed over Sumatera and Borneo Islands for past two decades. Various sector have been severely impacted, including air transportation. Air obscurity caused by smoke reduce horizontal visibility and profoundly affects aviation regularity, efficiency and safety. The study aimed to develop a visibility estimation of an orbital satellite using remote sensing and observation data. Orbital satellite is one of the technological advances that have potentials for supporting and improving the ability of meteorological service provider to meet user’s needs related to adverse weather. The Automatic smoke detection using classification tree analysis was used as the first step. Then backward elimination for multiple regression was used to produce visibility estimation based on Himawari_8 satellite-smoke-detection and actual observation of visibility in airports. The results showed that the correlation between the estimated visibility in satellite-detected smoke and observed visibility was high ($R^2 = 0.68$). This approach very sensitive to thick smoke with horizontal visibility of shorter than 2000 meters.

1. Introduction
Smoke from land and forest fires in Indonesia has attracted international attention due to the high emission of greenhouse gases that significantly impact weather and climate. Furthermore, smoke exposure in Southeast Asia has been reported to negatively affect health, tourism, transportation and economic sector [1]. The 2015 forest fire smoke is the worst smoke disaster in the last two decades in Indonesia. Almost 43 million people were affected by smoke particles, and a hundred thousand were afflicted by respiratory diseases [2].

Smoke aerosol particle size is very small, approximately 140 ±20 nm in Northern America (Kaufman and Fraser, 1997 in [3], 14 – 2500 nm in Beijing China [4]. According to the Rayleigh scattering theory, transmission at short wavelength is relatively sensitive to smoke plumes in opposite long wavelength transmission is relatively insensitive [3][5].

“Smoke is a combination of the airborne solid and liquid particulates and gases generated when a material undergoes combustion and the air that they are mixed with” [6]. One of the meteorological
factors that significantly impact capacity is visibility. The variability introduced by meteorological factors is exacerbated by the fact that aircraft have different weather capabilities ranging from almost all-weather to completely inhibited from operating. The resulting airport capacity, in terms of arrival and departure rates, can vary considerably and abruptly [7].

The definition of visibility for aeronautical purposes is “the greater of the greatest distance at which a black object of suitable dimensions, situated near the ground, can be seen and recognized when observed against a bright background; and the greatest distance at which lights in the vicinity of 1 000 candelas”. It’s revealed in the METAR messages, aerodrome (airport) routine meteorological reports. Instead of visibility, METAR message contain such as information: wind, present weather, cloud and pressure. In some case also contain trend forecast of weather element [8]. The increased concentration of aerosol particles in the atmosphere causes elevated extinction coefficient and low visibility [9]. Low visibility in the atmosphere impacted by hydrometeor (rain, fog, mist, snow) or lithometeor (smoke, dust) [10]. Visibility is possible as one of the parameters used to quantify smoke hazard level.

Himawari_8 Satellite with the Advanced Himawari Imager (AHI) was successfully launched onboard of Himawari-8 by the Japan Meteorological Agency (JMA) on 7 October 2014. It has 16 multispectral bands, including six visible and near infrared (VNIR) and ten IR bands. High spatial and spectral resolutions revealed at the first AHI images of the earth captures on December 2014 comparing with its predecessor MTSAT-series satellite images. AHI have three spatial resolution in nadir: 0.5-km for B3 (0.64 µm), 1-km for B1 (0.47µm), B2 (0.51µm), B4 (0.86µm), and 2-km for B7 to B16. [11]. Satellite Himawari-8 has ten minutes temporal resolution for Indonesian Area.

Classification tree analysis (CTA) is the one of the methods grouping/partioning the multidimensional data into as homogenous as possible [12]. This method become implemented over various field, such as medical or specifically pathogenic [13], transportation [14], smoke detection [15, 16]. This method is the new approach in multispectral satellite data classification which kind of machine learning of the multivariable data.

Visibility can be estimated by using satellite imagery besides from automatic in-situ observation (Runway Visual Range, RVR) or manual observation. The darkest object method was used for high-resolution satellite imagery (Landsat, ASTER, SPOT, IKONOS dan Quickbirds) because object like lakes, rivers looks dark at blue channel [17, 18]. The problem is small water body cannot clearly and detailed seen from satellite medium resolution such as Himawari_8 satellite data.

The research aim is to develop the spatial visibility estimation based on remote sensing and actual visibility observation (METAR). This approach offers many advantages for various sectors in life such as air transportation. The research area focuses on Sumatera and Borneo Island, Indonesia.

This paper is arranged as follows. Section II presents the research data and area, while Section III describes the research methods. Section IV is the result and discussion, containing the spectral analysis of Himawari_8 bands and their correlation with horizontal visibility, estimated visibility from the empiric model and its validation, and the sample implementation of spatial visibility estimation over small airport area. Lastly, Section V comprises the conclusion.

2. Data and research area

2.1. Data
The data used in this research were Himawari_8 satellite data and aerodrome routine meteorological report (METAR) data over Sumatera and Borneo Island Indonesia. Himawari_8 satellite data with 2-km spatial resolution, 10 minutes temporal resolution and 16 channels were used. This level 1b data have already been calibrated and validated, meaning that it can be used directly without any further correction.
METAR is the information of actual weather observations at an aerodrome made by aeronautical meteorological stations. This information is disseminated to the aerodromes/airports beyond the origin aerodrome. Actual weather condition such as wind, visibility, weather phenomena, cloud condition, dewpoint, temperature and air pressure are reported in METAR [8].

2.2. Research Area
The research area is displayed in figure 1. Data from 47 airports were used to develop and to validate the estimated visibility from the satellite. The areas where the airports are located are the regions that have mostly been covered by smoke each year in the past two decades. The visibility estimation model will be implemented to one sample airport (WIKT airport at Tanjung Pandan Island) to get more high spatial analysis.

![Figure 1 Research Area](image)

Figure 1 Research Area. White stars are the Airport location. Map of research area is develop using python-basemap and airports location. Red square is the sample of visibility-estimation implementation area (WIKT, airport over Tanjung Pandan Island).

3. Research Methods
The methods employed in this research include (1) smoke classification and separation with other backgrounds using the Classification Tree Analysis (CTA) based ROI sampling—which have been carried out by Ismanto et al. [16]; (2) correlation assessment between Himawari_8 channels and horizontal visibility data; (3) the correlation assessment between Himawari_8 bands value and horizontal visibility over smoke-affected areas; (4) the development of the visibility estimation models using backward elimination for multiple regression. The output of the visibility estimation model was then possible to be implemented over all airports in the study area, but only one sample was introduced here at Tanjung Pandan airport area (WIKT) that represent remote location from the existing forest and land fires.

As the key ingredient in the development of the visibility estimation model, the smoke detection relied on the Classification Tree Analysis with Entropy as the impurity measure. CTA Entropy-10 (figure 2) is spatially, logically and computation performance analyzed and become the best model in
3.1. Spectral And Correlation Analysis Of Himawari_8 Bands With Actual Horizontal Visibility Over Airports

This study employed Pearson correlation and combined data specifically from the areas where smoke occurrences had been detected by the satellite-based CTA model and METAR from 47’s airports. The data were sampled from the following dates: July 30, 2015 03.00 UTC; August 18, 2015 01.00 UTC; September 29, 2015 08.00 UTC; October 19, 2015 02.00 UTC; October 19, 2015 05.00 UTC; October 21, 2015 02.00 UTC; November 02, 2015 01.00 UTC.

3.2. Backward Elimination Multiple Regression

Backward elimination (or backward deletion) is a reverse process in which all independent variables are first inserted into the equation and deleted one at a time if they do not contribute to the regression equation. Correlation analysis firstly as to be a consideration to choose the possible and potential Himawari_8 bands in developing visibility estimation models. The visibility estimation models were multiple regression of combination and function of Himawari_8 satellite bands. The decision or selection of backward elimination was managed by its significant level < 0.05 and correlation value. Only models that fulfilled this condition will be accepted as the best models to estimate the horizontal visibility. The flowchart of this step revealed in figure 3.
4. Results and Discussions
The results of this study include (1) the correlation between Himawari_8 channels and surface visibility, (2) the spatial visibility estimation models and its correlation with surface visibility and (3) the spatial analysis of visibility estimation model due to smoke at a sample airport over small island, Tanjung Pandan Indonesia.

4.1. Backward Elimination Multiple Regression
Figure 4 shows the correlation values between each Himawari_8 channel and horizontal visibility in the observed airport. These values were generated from a set of data that had been filtered to only contain the smoke area classified from Himawari_8 images and according to observe visibility of airports. The length of the bar chart shows the correlation values, while the colour marks are the Himawari_8 bands.

A relatively good correlation between Himawari_8 satellite channels and visibility is shown in band 7 (short wave infrared (MID-IR)/3.9 \( \mu m \)); band 3 (reflectance) and thermal infrared (TIR) band 16. The reflectance band can capture the occurrence of smoke or low visibility because the small particles of smoke (submicron) are scattering the incoming wavelengths. These negative correlation values indicate that lower visibility (high concentration of smoke particle in the air) results in a more significant scattering of sunlight (shown in Figure 5a and Figure 5b).
A good relationship between Himawari_8 satellite and visibility was also identified in visible and thermal bands. The most significant one was found in band 7 (MID-IR). Band 7 (3.9 µm) is effective to monitor the low cloud, fog and forest fires [19]. The spectral values of smoke pixel in thermal infrared (TIR) bands appear because the smoke particles have a specific temperature that distinguishes them from other backgrounds and other surfaces. In general, an increase in thermal infrared temperature is associated with decrease in visibility (abundant of smoke-aerosol quantity) released from forest fires emission. Longer wavelength more sensitive with lower temperature differentiation, in which the temperature differentiation of smoke with other background is small. The characteristic value of smoke in TIR underlies the potential of this band as an input in multiple regression of the visibility estimation model.

4.2. The Visibility Estimation Model Generated By Multiple Regression Using Backward Elimination

Table 1 shows the empirical models for estimating surface visibility utilizing the combination of Himawari_8 satellite bands. These models have met the requirement of 95% confidence level (probability < 0.05). Therefore, they can generate the quantity of smoke aerosol through the proxy of visibility with a correlation value (R^2) of more than 0.50.

The best multiple linear regression is the visibility model 1. It is a combination of band 1, 7, and 16, which represent the visible and thermal bands (MID-IR and TIR). These band combination is similar to the Ismanto et al. studied [20], that also built of the combination of visible, MID-IR and TIR bands. The high correlation value (R^2) in this study is inline with that of the visibility model studied by Kessner et al [21] which uses the GOES-5 Satellite derivative products and creates a model that sufficiently assesses visibility modeling.

4.3. The Output Of The Smoke Detection Model Applied In Sampled Airport Witk In Tanjung Pandan

| Model No | Model | Correlation Value (R^2) |
|----------|-------|-------------------------|
| 1        | \( Y = -1,786e+04 + (-1477,362xB1) + (104,562xB7) + (-47,476xB16) \) | 0.68 |
| 2        | \( Y = -1,797e+04 + (-1217,242xB2) + (102,8822xB7) + (-45,6036xB16) \) | 0.67 |
| 3        | \( Y = -2,476e+04 + (-1535,224xB1) + (116,247xB7) + (-33,585xB15) \) | 0.66 |
| 4        | \( Y = -2,643e+04 + (-1699,221xB1) + (126,776xB7) + (-38,1896xB14) \) | 0.65 |
| 5        | \( Y = -1,541e+04 + (104,371xB7) + (-58,068xB16) \) | 0.62 |
| 6        | \( Y = -3,341e+04 + (-1934,067xB1) + (112,888xB7) \) | 0.56 |
Tanjung Pandan Airport is a sample that can provide an output of the visibility model on a detailed scale due to its small area. This airport was selected because of several considerations, including its proximity to the source of forest fire (mainly in Sumatera and Borneo Island) [22] and its suitable geographic position for testing the performance of the visibility estimation model (a small land area or island surrounded by the ocean) (figure 6 and figure 7).

The CTA model detected smoke area very clearly by CTA model output (Figure 6; October 24, 2015, 05.00 UTC). The smoke area was revealed as the color of visibility estimation. The smoke covered more than 50% of the areas located within radius 8 km and 16 km from airport perimeter. Visibility estimation value around 600 to 2000 meters.

According to the ICAO document number 8896 manual on aeronautical weather observation, weather condition for air navigation information must be representative of the aerodrome area (within radius 8 km to radius 16 km from aerodrome center point). Aerodrome weather information of representative area around radius 8 km used as based data of local routine report (information especially for takeoff and landing), meanwhile representative area around 16 km used as based data of METAR report (information of actual weather over the aerodrome and disseminated beyond the aerodrome).

The disadvantage of satellite-based smoke detection includes low sensitivity to the thin smoke. On October 25, 2015, 02.00 UTC, there was no smoke detected by our algorithm in the aerodrome (Figure 7), but the METAR data showed a presence of smoke with a visibility 3500 m (referred to as thin smoke). This condition indicates that the spectral value of satellite data for aerosol is mixed with that of the underlying surface. Otherwise, the CTA model shows that cloud presence along 10-20% of aerodrome area that in line with METAR cloud information (FEW020 means lowest layer cloud presence almost 10-20% area of sky). This result was not good enough if compares with Xie et al [3], thick and thin dispersed smoke can be captured by the back propagation neural network detection algorithm. Even though, these paper results still have opportunity to be implemented over aviation needs due to flight operation only sensitive in low visibility in instrument flight rule.

![Figure 6](image)

**Figure 6** Visibility estimation model performance in thick smoke condition over small airport WIKT at Tanjung Pandan Island.
Figure 7 Visibility estimation model performance in thin smoke condition over small airport WIKT at Tanjung Pandan Island.

4.3. The Output Of The Smoke Detection Model Applied In Sampled Airport Wikt In Tanjung Pandan

These results are useful for air transportation, including weather information support airport capacity. It is often variable and dependent on several aspect one of them weather. Meteorological factors that significantly impact capacity include wind direction and intensity, precipitation, visibility, hazardous clouds and more severe hazards. The variability introduced by meteorological factors is exacerbated by the fact that aircraft have different weather capabilities ranging from almost all-weather to completely [23].

5. Conclusions

The results of this study are as follows. Himawari_8 satellite is sufficiently good at detecting the presence of thick smoke with visibility value less than 2000 meters. The best empirical model of visibility estimation combines by three main part variable that are reflective, shortwave infrared and thermal infrared bands. This combination (bands 1, 7, and 16) found as the best visibility estimation combination with a coefficient correlation value of 0.68. This result has advantage information for aviation users. Spatial characteristic of horizontal visibility over the airport is useful for the service provider, operators, and users to improve their situational awareness related to the smoke hazard for aviation safety, regularity and efficiency.

Acknowledgement

This paper was funded by Rekognisi Tugas Akhir – Universitas Gadjah Mada (RTA - UGM) and as part of the Phd doctoral dissertation result.

References

[1] D Murdiyarso, Lebel L, Gintings A, Tampubolon S, Heil A and Wasson M 2004 Policy responses to complex environmental problems: insights from a science–policy activity on transboundary haze from vegetation fires in Southeast Asia *Agriculture Ecosystem & Environmental* pp 47-56
[2] McKirdy E 2015 CNN World 29 October 2015. [Online]. Available: http://www.edition.cnn.com/2015/10/29/asia/southeast-asia-haze-crisis/index.html.

[3] Xie Y, Qu J, Xiong X, Hao X, Che N and Sommers W 2007 Smoke plume detection in the eastern United States using MODIS International Journal of Remote Sensing pp 2367-2374

[4] Wang Y H, Liu Z R, Zhang J K, Hu B, Ji D S, Yu Y C and Wang Y S 2015 Aerosol physicochemical properties and implications for visibility during an intense haze episode during winter in Beijing Atmospheric Chemistry and Physics pp 3205-3215

[5] Li X, Song W, Lian L and Wei X 2015 Forest fire smoke detection using back-propagation neural network based on MODIS data Remote Sensing pp 4473-4498

[6] SKYbrary 2018 Fire Smoke and Fume 4 December 2018 [Online] Available: https://www.skybrary.aero/index.php/Smoke#Definition.

[7] Kistan T, Gardia A, Sabatinia R, Ramasamy S and Batuwangala E 2017 An evolutionary outlook of air traffic flow management techniques Progress in Aerospace Sciences pp 15-42

[8] ICAO 2011 Manual on Aeronautical Meteorological Practice pp 4473-4498 (Canada: International Civil Aviation Organization)

[9] Anodina T G and Ibragim S M 2016 Study of aerosol humidification impact Естеств. Науки pp. 43-46

[10] Mali R and Vashistha R 2005 Visibility Measurement technique and Its Application in Aviation Services at International Airports in India [Online] Available: https://www.wmo.int/pages/prog/www/IMOPpublications/IOM-82-TECO-2005/Posters/P1(08)_India_2_Mali.pdf.

[11] Yu F and Wu X 2016 Radiometric inter-calibration between Himawari-8 AHI and S-NPP VIIRS for the solar reflective bands Remote Sensing pp 1-16

[12] Thiller W, Araujo M B and Lavorel S 2003 Generalized models vs classification tree analysis: predicting spatial distributions of plant species at different scales Journal of Vegetation Science pp 669-680

[13] Dominick G M, Papas M A, Rogers M L and Rakowski W 2015 Classification tree analysis to examine influences on colorectal cancer screening Cancer Causes & Control pp 443-454

[14] Shaaban K and Pande A 2016 Classification tree analysis of factors affecting parking choices in Qatar Case Study on Transport Policy pp 88-95

[15] Ismanto H, Hartono H and Marfai M A 2019 Classification tree analysis (Gini-Index) smoke detection using Himawari_8 Satellite data over Sumatera-Borneo Maritime Continent Southeast Asia International Conference on Environmental Resources Management in Global Region, Yogyakarta pp. 1-9

[16] Ismanto H, Hartono H and Marfai M A. Automatic smoke detections using classification tree analysis using Himawari_8 satellite data over Sumatera and Borneo Island Indonesia Acta Geophysica pp 1-14
[17] Hadjimitsis D, Clayton C and Toulios L 2010 Retrieving visibility values using satellite remote sensing data *Physics and Chemistry of the Earth* pp 121-124

[18] Retalis A, Hadjimitsis D G, Michaelides S, Tymvios F, Chrysoulakis N, Clayton C R I and Themistocleous K 2010 Comparison of aerosol optical thickness with in situ visibility data over Cyprus *Natural Hazards Earth System Science* pp 421-428

[19] Bessho K, Date K and Hayashi M 2016 An Introduction to Himawari_8/9-Japan's New-Generation Geostationary Meteorological Satellites *Journal of the Meteorological Society of Japan* pp 151-183

[20] Ismanto H, Hartono H and Marfai M A 2018 Smoke detection and visibility estimation using Himawari_8 satellite data over Sumatera and Borneo Island Indonesia *Spatial Information Research* pp 1-12

[21] Kessner A L, Wang J, Levy R C and Colarco P R 2013 Remote sensing of surface visibility from space: A look at the United States East Coast *Atmospheric Environment* pp 136-147

[22] Ismanto H, Hartono H and Marfai M A 2019 Spatiotemporal visibility characteristics impacted by forest and land fire over airports in Sumatera and Borneo Island, Indonesia *Quaestiones Geographicae* pp 5-16

[23] Kistan T, Gardi A, Sabatini R and Ramasami S 2017 An evolutionary outlook of air traffic flow management techniques *Progress in Aerospace Sciences* pp 15-42