

IDENTIFIKASI AWAN PADA DATA TIME SERIES MULTITEMPORAL MENGGUNAKAN PERBANDINGAN DATA SEKUENSIAL

(CLOUD IDENTIFICATION FOR MULTITEMPORAL TIME SERIES DATA USING SEQUENTIAL COMPARISON)

Anis Kamilah Hayati^{1,a}, Wismu Sunarmodo^{1,a}

¹Pusat Teknologi dan Data Penginderaan Jauh, LAPAN

^aKontributor utama

e-mail: anis.kamilah@apan.go.id

Diterima : 2 Februari 2021; Direvisi : 12 Juli 2021; Disetujui : 15 November 2021

ABSTRACT

Cloud identification is an important pre-processing step of remote sensing data. Generally, cloud identifications could be classified into single-date and multi-date methods. Furthermore, the single-date method could be divided into physical-rules-based and machine-learning-based. Physical-rules-based method generally need data with sufficient spectral resolution while machine-learning-based method depend on training dataset. While the multi-date method usually using clear data as a reference. The clear data itself could be a whole scene or built from many scenes. Processing cloud-free data is a challenge in areas with high cloud coverage such as Indonesia. In this paper, a cloud identification method using multi-date time series scenes is proposed. This method only uses RGB channels which are common in remote sensing visual data. In addition, this method does not require or process cloud-free data mosaics in advance. A pixel value from an examined scene is compared to other pixel values from other scenes in the same position. The other scenes are the scenes that were acquired before and after the examined scene. The value differences between the examined pixel and it's before and after then evaluated using some thresholds to determine whether the pixel is a cloud or not. Assessment is done by using L8 Biome as a reference. The result shows that using some thresholds in our proposed method has a Kappa coefficient higher than 0.9.

Keywords: *cloud identification, multitemporal, time series, remote sensing, sequential comparison*

ABSTRAK

Identifikasi awan merupakan langkah penting dalam pengolahan data citra satelit penginderaan jauh optis Landsat-8. Secara umum, identifikasi awan dapat diklasifikasikan dalam dua metode, yaitu *single-date* dan *multi-date*. Metode *single-date* dapat dibagi menjadi dua kategori, yaitu metode yang berdasarkan ciri-ciri fisik dan metode yang berdasarkan pembelajaran mesin. Metode yang berdasarkan ciri-ciri fisik pada umumnya membutuhkan data dengan resolusi spektral yang baik sementara metode berdasarkan pembelajaran mesin sangat bergantung pada *training dataset*. Sementara metode *multi-date* pada umumnya menggunakan data yang bebas awan sebagai referensi. Data bebas awan tersebut bisa merupakan satu *scene* secara keseluruhan maupun dibangun dari beberapa *scene*. Pengolahan data bebas awan merupakan suatu tantangan terutama di area dengan liputan awan yang tinggi seperti Indonesia. Makalah ini membahas metode *multi-date* untuk identifikasi awan dengan menggunakan data *time series*. Metode ini hanya menggunakan kanal RGB yang umum pada data visual penginderaan jauh. Selain itu, metode ini tidak memerlukan atau memproses mosaik data bebas awan terlebih dahulu. Suatu nilai piksel dari suatu *scene* yang diperiksa dibandingkan dengan nilai piksel dari *scene* yang berbeda pada lokasi yang sama. *Scene* yang berbeda yang dimaksud adalah data yang diakuisisi sebelum dan sesudah data yang diperiksa. Perbedaan nilai piksel dari data yang diperiksa dan data

yang diakuisisi sebelum dan setelahnya itu kemudian dievaluasi menggunakan *thresholds* untuk mengkategorikan piksel tersebut sebagai awan atau non-awan. *Assessment* dilakukan dengan menggunakan L8 Biome sebagai referensi. Hasil dari *assessment* menunjukkan metode yang diusulkan memiliki koefisien Kappa lebih besar dari 0,9.

Kata kunci: *identifikasi awan, multitemporal, time series, penginderaan jauh, perbandingan sekuensial*

1 PENDAHULUAN

Keberadaan awan merupakan hal yang tidak terpisahkan dari data penginderaan jauh optis multispektral termasuk Landsat-8. Keberadaan awan tersebut dapat mempengaruhi analisis fenomena permukaan bumi sehingga identifikasi awan merupakan hal yang penting dilakukan sebagai langkah pengolahan awal pada data citra penginderaan jauh.

Penelitian terkait identifikasi awan telah banyak dilakukan. Zhu et al. (2019) membagi metode identifikasi awan menjadi *single-date* dan *multi-date*. Metode *single-date* kemudian dibagi lagi menjadi metode yang berdasarkan ciri-ciri fisik atau *physical-rules-based* dan metode yang berdasarkan pembelajaran mesin atau *machine-learning-based*. Metode *single-date* dengan *physical-rules-based* pada umumnya sangat bergantung pada analisis kanal spektral. Metode ini memanfaatkan karakteristik awan seperti terang dan dingin (Huang et al., 2010; Irish et al., 2006; Li et al., 2017; Sun et al., 2018; Zhu & Woodcock, 2012). Dengan demikian, semakin banyak kanal yang dimiliki oleh suatu data, semakin akurat hasil identifikasi awan yang dihasilkan. Adapun kanal yang paling umum digunakan sebagai acuan identifikasi awan adalah kanal termal.

Metode *single-date* dengan menggunakan pembelajaran mesin pada umumnya menunjukkan hasil yang baik. Sebagai contoh, Hughes & Hayes (2014) menggunakan *neural network classifier* dan memiliki akurasi keseluruhan sekitar 99%, sementara Zhou et al. (2016) memanfaatkan *support vector machine* sebagai *classifier* dan memberikan hasil akurasi lebih dari 90%. Meskipun memberikan hasil yang baik, metode dengan menggunakan pembelajaran mesin sangat bergantung pada *training dataset* (Huang et al., 2010). *Training dataset* yang baik akan menghasilkan luaran yang baik dan sebaliknya. Adapun

untuk membangun *training dataset* yang baik, diperlukan kemampuan dan pengetahuan tertentu terkait awan.

Metode *multi-date* pada umumnya menggunakan referensi berupa data yang bebas awan, baik itu satu *scene* keseluruhan (Jin et al., 2013), maupun dibangun dari beberapa data dengan menggabungkan area yang bebas dari awan (Hagolle et al., 2010), atau dengan menentukan piksel bebas awan dari suatu kelompok data (Goodwin et al., 2013). Hal ini sulit dilakukan di daerah seperti Indonesia yang memiliki liputan awan yang tinggi sepanjang tahun. Oleh karena itu makalah ini mengusulkan suatu metode identifikasi awan dengan menggunakan data multi-temporal yang tidak bergantung pada *scene* bebas awan. Hal ini dilakukan dengan membandingkan nilai piksel pada suatu *scene* dengan nilai piksel pada *scene* lain hasil akuisisi tanggal sebelumnya dan tanggal sesudahnya pada lokasi piksel yang sama. Perbedaan nilai piksel kemudian dievaluasi dengan menggunakan *threshold* untuk mengkategorikan piksel tersebut sebagai awan atau non-awan.

Asesmen dilakukan dengan menggunakan referensi dari L8 Biome *cloud cover assessment set* (U.S. Geological Survey, 2016). Selain itu data referensi pun dibandingkan dengan hasil identifikasi awan dari metode FMask (Zhu & Woodcock, 2012). Metode yang diajukan diharapkan memiliki akurasi lebih dari 0,85 sesuai batas yang direkomendasikan USGS (Anderson et al., 1976).

2 METODOLOGI

2.1 Lokasi dan Data

Data yang digunakan dalam makalah ini adalah data Landsat-8 *path 113 row 063* sepanjang tahun 2014 sebanyak 20 *scene*. Data ini meliputi wilayah Sulawesi Tenggara. Adapun liputan awan seluruh

scene yang digunakan dijelaskan pada Tabel 2-1. Persentase liputan awan ini diperoleh dari metadata Landsat-8.

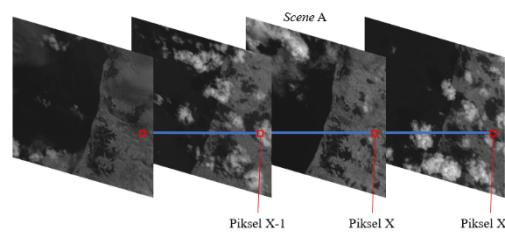
Tabel 2-1: Tanggal Akuisisi Dan Liputan Awan Dari Data Yang Digunakan.

Tanggal Akuisisi	Liputan Awan (%)
1 Januari 2014	84,16
17 Januari , 2014	89,66
6 Maret 2014	60,35
22 Maret 2014	69,7
7 April 2014	49,05
23 April 2014	18,39
9 Mei 2014	52,45
25 Mei 2014	57,48
10 Juni 2014	66,81
26 Juni 2014	57,38
12 Juli 2014	50,69
29 Juli 2014	11,76
13 Agustus 2014	11,97
29 Agustus 2014	8,65
3 September 2014	4,94
16 Oktober 2014	15,63
1 November 2014	4,84
17 November 2014	17,11
3 Desember 2014	76,21
19 Desember 2014	66,29

Kanal yang digunakan dalam penelitian ini adalah kanal visual yang meliputi kanal *Red*, *Green* dan *Blue* (kanal 4, 3, dan 2). Kanal visual digunakan untuk mengetahui apakah metode ini juga dapat digunakan untuk data yang hanya memiliki kanal visual. Penelitian ini dilaksanakan menggunakan data satelit Landsat-8 level 1 (L1T dan L1G) dengan resolusi radiometrik 16 bit.

2.2 Metode Penelitian

Data *time series* digunakan dengan asumsi bahwa objek tutupan lahan pada data tersebut tidak berubah drastis di antara periode akuisisi data. Dengan demikian piksel pada posisi yang sama seharusnya memiliki nilai yang tidak jauh berbeda karena jenis tutupan lahan yang sama. Namun ketika piksel tersebut merupakan awan atau bayangan awan, maka nilai piksel tersebut akan jauh berbeda.

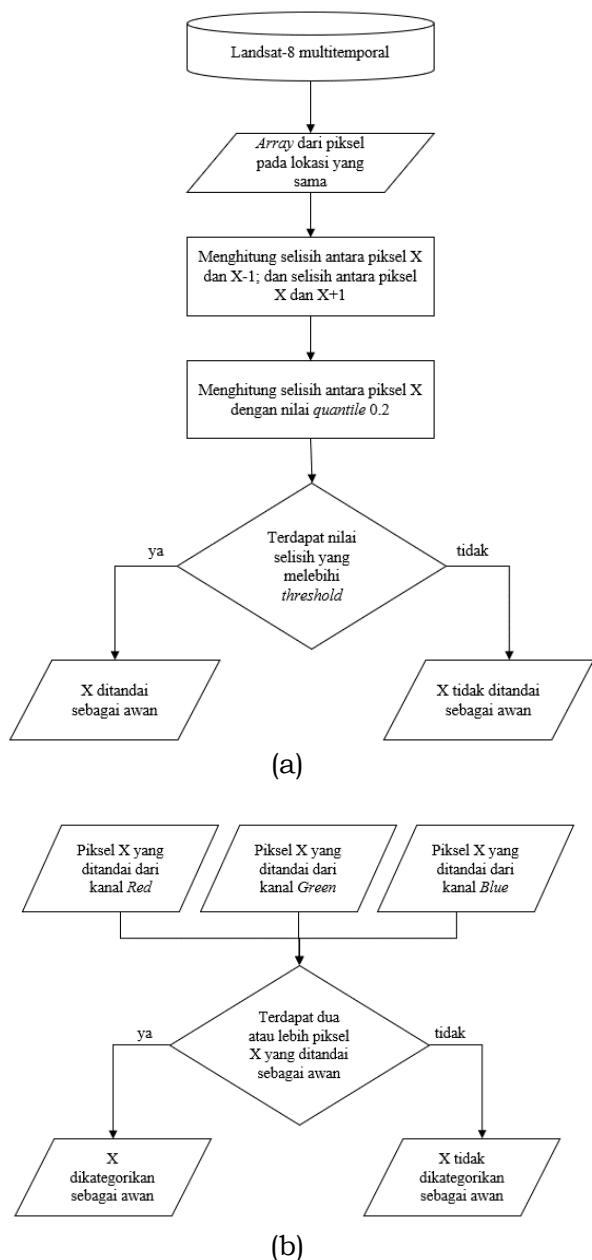


Gambar 2-1: *Scene* yang ditumpuk, kotak merah menunjukkan piksel pada lokasi yang sama.

Seluruh *scene* yang digunakan dalam penelitian ini ditumpuk berdasarkan tanggal akuisisi. Kemudian piksel yang berada pada lokasi yang sama dikumpulkan menjadi satu *array*. Saat menganalisis suatu piksel dari suatu *scene* (piksel X, *scene* A) maka piksel X tersebut akan dibandingkan dengan piksel dari *scene* yang diakuisisi sebelum dan setelah *scene* A. Untuk memudahkan penamaan, dalam makalah ini piksel dari *scene* yang diakuisisi sebelum *scene* A disebut X-1 dan piksel dari *scene* yang diakuisisi setelah *scene* A disebut X+1 (Gambar 2-1). Jika selisih antara nilai piksel X dengan X-1 atau antara pixel X dengan X+1 melebihi suatu *threshold* tertentu maka piksel X ditandai sebagai awan.

Ketika terdapat piksel awan yang berurutan, maka nilai selisih piksel tersebut tidak akan memenuhi nilai *threshold*. Oleh karena itu untuk menangani kasus dimana piksel X-1 atau X+1 juga merupakan awan, digunakan *quantile* 0,2 sebagai acuan. Dengan asumsi *quantile* 0,2 secara statistik dapat mewakili nilai piksel tutupan lahan atau objek selain awan dan tutupan awan. Nilai minimum tidak digunakan di sini karena dari percobaan yang dilakukan pada beberapa data, nilai minimum yang didapatkan merupakan nilai bayangan awan. Dengan demikian, jika perbedaan antara nilai piksel dengan *quantile* 0,2 lebih besar dari *threshold*, maka piksel X ditandai sebagai awan.

Proses ini dilakukan pada setiap kanal (kanal 4, 3, dan 2). Hasil final ditentukan berdasarkan hasil *voting* dari sejumlah kanal. Jika dua atau lebih kanal menandai piksel X sebagai awan, maka hasil final akan mengkategorikan piksel X sebagai awan. Gambar 2-2 menunjukkan algoritma yang digunakan dalam penelitian ini.



Gambar 2-2: Algoritma yang digunakan, (a) Proses pada masing-masing kanal, dan (b) Hasil dari *voting* setiap kanal.

Dalam penelitian ini digunakan beberapa *threshold* untuk mengetahui nilai *threshold* yang memberikan hasil yang paling baik. Ketika data untuk penelitian ini dikumpulkan, ditemukan bahwa selisih antara piksel yang berurutan berkisar antara ratusan hingga sekitar 10.000. Adapun selisih antara piksel yang berurutan yang menunjukkan keberadaan awan dimulai dari kisaran 2.000.

Dengan mempertimbangkan hasil temuan dari beberapa data tersebut,

penelitian ini membatasi kisaran *threshold* yang digunakan yaitu antara 2.000-5.000. Dengan demikian dalam penelitian ini diujicobakan empat *threshold* yaitu 2.000, 3.000, 4.000, dan 5.000.

2.3 Metode Asesmen

Keempat *threshold* tersebut menghasilkan empat dataset. Keempat dataset tersebut kemudian dibandingkan dengan hasil dari Fmask dan L8 Biome. Adapun perbandingan yang dilakukan meliputi asesmen secara kualitatif dan kuantitatif. Asesmen kuantitatif dilakukan dengan menghitung nilai akurasi dan Kappa koefisien.

Nilai akurasi dihitung berdasarkan jumlah piksel yang terdeteksi sesuai dengan referensi sesuai persamaan 2-1.

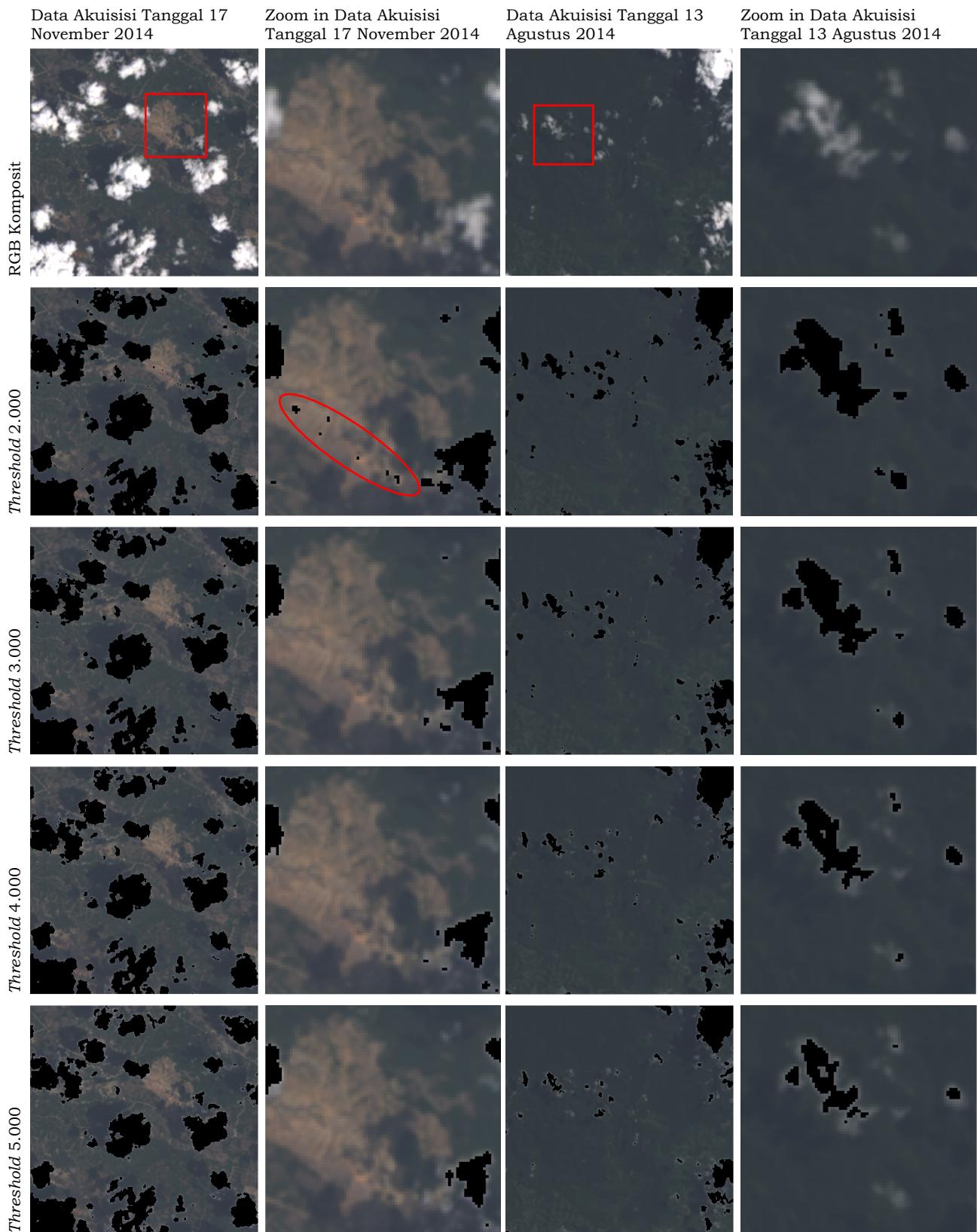
$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2-1)$$

Dalam hal ini TP (*true positive* yaitu piksel awan yang terdeteksi sebagai awan) ditambah dengan TN (*true negative* yaitu piksel non-awan yang terdeteksi sebagai non-awan) dibagi jumlah piksel keseluruhan termasuk FP (*false positive* yaitu piksel non-awan yang terdeteksi sebagai awan) dan FN (*false negative* yaitu piksel awan yang terdeteksi sebagai non-awan).

3 HASIL PEMBAHASAN

Secara visual, terlihat bahwa dataset dengan *threshold* 2.000 menunjukkan hasil dengan *omission error* yang paling rendah. Namun juga terlihat bahwa dataset ini memiliki *commission error* paling tinggi dibanding dataset lainnya (lingkaran merah). Sebaliknya, dataset dengan *threshold* 5.000 merupakan dataset dengan tingkat kegagalan deteksi awan paling banyak.

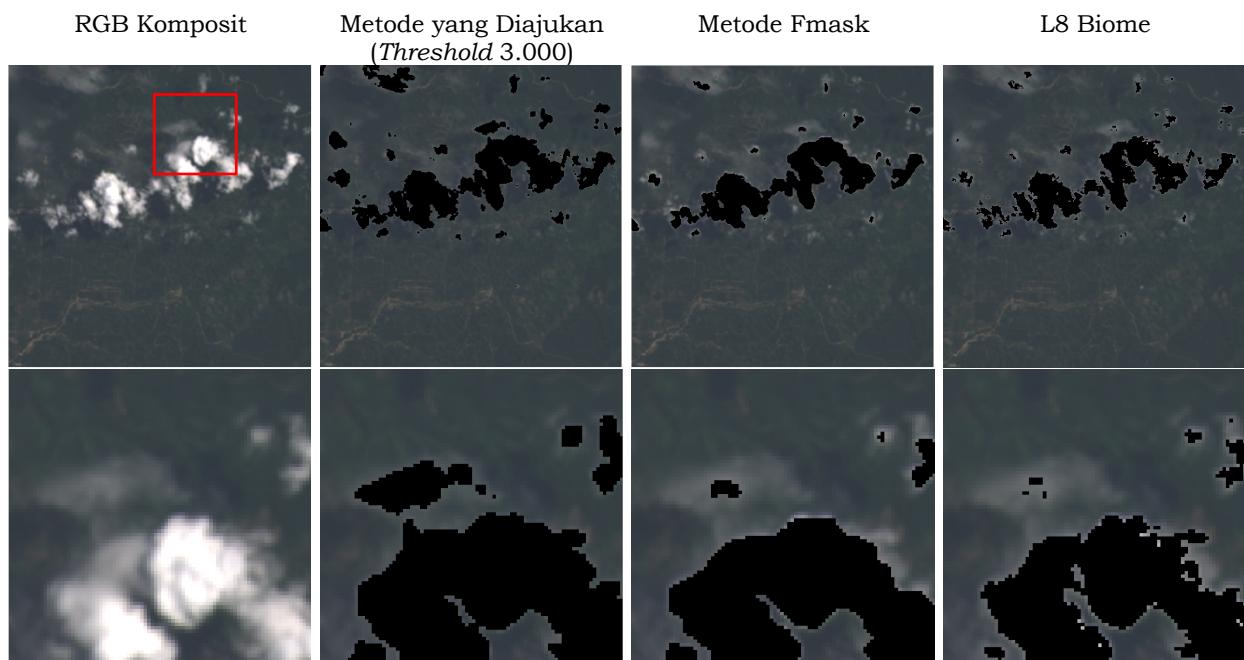
Gambar 3-1 menunjukkan hasil identifikasi awan dari keempat *threshold* yang digunakan. Kotak merah menunjukkan area yang diperbesar (*zoom-in*). Awan yang teridentifikasi pada masing-masing dataset ditunjukkan dengan piksel warna hitam.



Gambar 3-1: Perbandingan hasil identifikasi awan dengan menggunakan empat *threshold* yang berbeda.

Jika dibandingkan dengan data referensi dari L8 Biome dan Fmask, secara visual penelitian ini menunjukkan hasil yang cukup baik. Hal itu dapat dilihat dari Gambar 3-2, bahwa dengan

menggunakan *threshold* 3.000, hasil penelitian ini mampu mendeteksi awan-awan tipis yang tidak terdeteksi oleh algoritma Fmask dan juga luput pada L8 Biome.



Gambar 3-2: Perbandingan hasil identifikasi awan dari metode yang diajukan dengan metode Fmask dan L8 Biome.

Asesmen secara kualitatif dilakukan dengan menghitung akurasi dan Kappa koefisien dari data hasil penelitian dibandingkan data L8 Biome. Tabel 3-1 menunjukkan hasil dari perhitungan kualitatif. Dataset dengan *threshold* 5.000 memiliki akurasi dan Kappa koefisien yang paling tinggi dibanding dataset lainnya. Meskipun demikian, perlu diingat bahwa sebagai referensi data L8 Biome pun secara keseluruhan memiliki kesalahan akurasi sekitar 7% (Zhu et al., 2019). Selain itu, secara visual pun L8 Biome memiliki kekurangan terutama pada awan tipis yang tidak

terdeteksi. Dengan demikian, perbandingan seluruh dataset dengan L8 Biome pun memiliki kemungkinan kesalahan.

Selain membandingkan L8 Biome dengan metode yang diajukan, pada penelitian ini diujicobakan juga dengan data hasil algoritma Fmask. Dengan membandingkan data dari L8 Biome dan Fmask, didapatkan hasil pada Tabel 3-1. Nilai akurasi dan Kappa koefisien data hasil dari metode yang diajukan (*threshold* 4.000 dan 5.000) lebih baik dibandingkan Fmask.

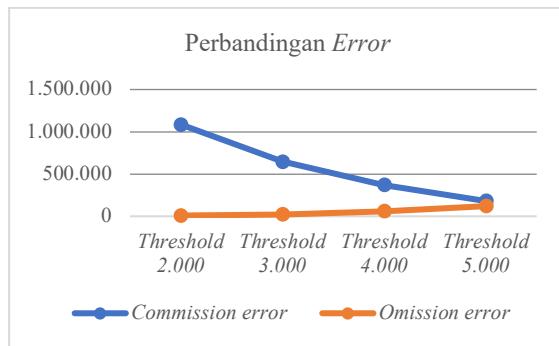
Tabel 3-1: Asesmen Kualitatif Dengan Menggunakan L8 Biome Sebagai Referensi

Dataset	Threshold 2.000	Threshold 3.000	Threshold 4.000	Threshold 5.000	Fmask
True Negative	29.750.350	30.187.358	30.463.824	30.655.817	30.444.433
False Positive	1.085.976	648.968	372.502	180.509	391.893
False Negative	9.330	23.941	60.501	119.883	111.876
True Positive	2.746.664	2.732.053	2.695.493	2.636.111	2.644.118
Akurasi	0,967394	0,979968	0,98711	0,991058	0,985003
Kappa Koefisien	0,816216	0,879455	0,918617	0,94122	0,904839

Sesuai dengan perkiraan bahwa terdapat *trade-off* dari penggunaan *threshold*. Semakin rendah *threshold*

yang digunakan, maka semakin tinggi *commission error* yang dihasilkan meski *omission error* rendah. Sebaliknya,

semakin tinggi *threshold* yang digunakan semakin tinggi *omission error* yang dihasilkan meski *commission error* lebih rendah (Gambar 3-3). Meski demikian, pada banyak aplikasi, ketepatan pendekripsi awan lebih diutamakan dan *commission error* lebih dimaklumi (Goodwin et al., 2013; Zhu & Woodcock, 2012).



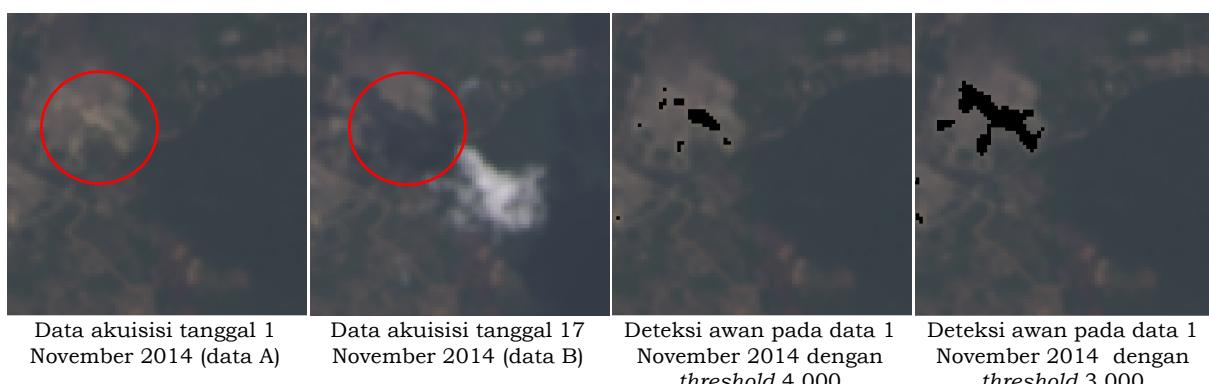
Gambar 3-3: Perbandingan *commission* dan *omission error*

Meskipun secara umum hasil penelitian ini menunjukkan hasil yang cukup baik, namun masih terdapat kekurangan yang perlu diperbaiki untuk penelitian lebih lanjut. Pada metode yang diajukan dalam penelitian ini, deteksi

awan pada piksel X sangat bergantung pada nilai piksel yang berurutan. Dengan demikian metode ini akan mengalami kegagalan jika nilai piksel yang berurutan tidak sesuai dengan yang diharapkan.

Gambar 3-4 menunjukkan salah satu contoh kasus. Lingkaran merah menunjukkan daerah dimana nilai piksel antara data yang diakuisisi pada tanggal 1 November 2014 (data A) dengan piksel dari data yang diakuisisi pada tanggal 17 November 2014 (data B) memiliki selisih yang signifikan. Hal ini disebabkan nilai piksel pada data B sangat rendah karena merupakan bayangan awan. Karena selisih yang signifikan tersebut melebihi *threshold*, maka piksel pada data A teridentifikasi sebagai awan.

Selain itu, kasus lain yang dapat mengakibatkan kegagalan pada metode ini adalah jika piksel pada lokasi tertentu pada keseluruhan data merupakan awan. Jika hal ini terjadi maka tidak ada selisih antara nilai piksel X dengan piksel X-1 atau X+1 yang akan melebihi *threshold*. Dengan demikian seluruh piksel akan dianggap sebagai non-awan



Gambar 3-4: Contoh kasus ketika metode yang diusulkan gagal mengkategorikan awan

4 KESIMPULAN

Dalam makalah ini dilakukan ujicoba metode identifikasi awan dengan menggunakan data *time-series* multitemporal dari Landsat-8. Metode ini dilakukan dengan membandingkan beberapa *threshold* dengan perbedaan nilai piksel dari data yang diakuisisi pada tanggal yang berbeda. Dengan menggunakan empat *threshold* yang berbeda dihasilkan empat dataset. Dataset yang dihasilkan menunjukkan

trade-off pada *omission* dan *commission error* dari penggunaan *threshold*. Semakin rendah *threshold* yang digunakan, akan menghasilkan *commission error* yang semakin tinggi dan *omission error* yang semakin rendah.

Pada asesmen kuantitatif maupun kualitatif, dapat dilihat bahwa metode yang diajukan (pada *threshold* 4.000) memiliki hasil yang lebih baik dibanding metode FMask. Meskipun perlu menjadi catatan bahwa L8 Biome sebagai

referensi pun memiliki kesalahan akurasi.

Beberapa hal dapat dilakukan untuk melanjutkan penelitian ini. Metode yang diajukan dapat diterapkan pada data lain yang hanya memiliki kanal visual. Selain itu, metode ini dapat dikembangkan untuk memperbaiki beberapa kasus yang dapat menyebabkan kegagalan dalam mengidentifikasi awan.

UCAPAN TERIMAKASIH

Terima kasih pada Pustekdata LAPAN yang telah menyediakan data Landsat-8 dan fasilitas demi terselenggaranya penelitian ini.

DAFTAR RUJUKAN

- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. 1976. *A land use and land cover classification system for use with remote sensor data*. U.S. Geological Survey Professional Paper, No. 964. USGS, Washington, D.C.
- Goodwin, N. R., Collett, L. J., Denham, R. J., Flood, N., & Tindall, D. 2013. *Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time series*. Remote Sensing of Environment, 134, 50–65. <https://doi.org/10.1016/j.rse.2013.02.019>
- Hagolle, O., Huc, M., Pascual, D. V., & Dedieu, G. 2010. *A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images*. Remote Sensing of Environment, 114(8), 1747–1755. <https://doi.org/10.1016/j.rse.2010.03.002>
- Huang, C., Thomas, N., Goward, S. N., Masek, J. G., Zhu, Z., Townshend, J. R. G., & Vogelmann, J. E. 2010. *Automated masking of cloud and cloud shadow for forest change analysis using Landsat images*. International Journal of Remote Sensing, 31(20), 5449–5464. <https://doi.org/10.1080/01431160903369642>
- Hughes, M. J., & Hayes, D. J. 2014. *Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing*. Remote Sensing, 6(6), 4907–4926. <https://doi.org/10.3390/rs6064907>
- Irish, R. R., Barker, J. L., Goward, S. N., & Arvidson, T. 2006. *Characterization of the landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm*. Photogrammetric Engineering and Remote Sensing, 72(10), 1179–1188. <https://doi.org/10.14358/PERS.72.10.1179>
- Jin, S., Homer, C., Yang, L., Xian, G., Fry, J., Danielson, P., & Townsend, P. A. 2013. *Automated cloud and shadow detection and filling using two-date Landsat imagery in the USA*. International Journal of Remote Sensing, 34(5), 1540–1560. <https://doi.org/10.1080/01431161.2012.720045>
- Li, Z., Shen, H., Li, H., Xia, G., Gamba, P., & Zhang, L. 2017. *Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery*. Remote Sensing of Environment, 191(April 2013), 342–358. <https://doi.org/10.1016/j.rse.2017.01.026>
- Sun, L., Liu, X., Yang, Y., Chen, T. T., Wang, Q., & Zhou, X. 2018. *A cloud shadow detection method combined with cloud height iteration and spectral analysis for Landsat 8 OLI data*. ISPRS Journal of Photogrammetry and Remote Sensing, 138, 193–207. <https://doi.org/10.1016/j.isprsjprs.2018.02.016>
- Zhou, G., Zhou, X., Yue, T., & Liu, Y. 2016. *An optional threshold with SVM cloud detection algorithm and DSP implementation*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 41(July), 771–777. <https://doi.org/10.5194/isprsaarchives-XLI-B8-771-2016>
- Zhu, Z., Qiu, S., He, B., & Deng, C. 2019. *Cloud and Cloud Shadow Detection for Landsat Images: The Fundamental Basis for Analyzing Landsat Time Series*. Remote Sensing Time Series Image Processing, May, 3–23. <https://doi.org/10.1201/9781315166636-1>
- Zhu, Z., & Woodcock, C. E. 2012. *Object-based cloud and cloud shadow detection in Landsat imagery*. Remote Sensing of Environment, 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>
- U.S. Geological Survey. 2016. *L8 Biome Cloud Validation Masks*. U.S. Geological Survey, data release. doi:10.5066/F7251GDH. Diakses 2 Februari 2020. <https://landsat.usgs.gov/landsat-8-cloud-cover-assessment-validation-data>

