The validation of water quality parameter algorithm using Landsat 8 and Sentinel-2 image in Palabuhanratu Bay

M K Mukhtar, Supriatna, and M D M Manessa
Department of Geography, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: ysupri@sci.ui.ac.id

Abstract. Remote sensing and geographic information systems can be applied to extract coastal and marine parameters related to the identification of possible data types, approaches and algorithms as a quick solution in water quality assessment. The purpose of this research are to find the suitable algorithms of salinity and total suspended solid for Palabuhanratu Bay and see the performance of Sentinel-2 satellite image in implementing algorithms based on Landsat satellite image. This study applies several algorithms to extract the estimated salinity and total solid suspended values from the Landsat 8 and Sentinel-2 satellite image using Google Earth Engine. The results of the analysis show algorithms that are suitable for implementation in the waters of Palabuhanratu Bay are the Cilamaya algorithm for estimating salinity values, and the Budhiman algorithm for estimating the total suspended solid value. Sentinel-2 satellite image has a good performance for implementing algorithms that built on Landsat image. So, the algorithm that build on Landsat image can be used to detect salinity and TSS in Sentinel-2 image.

Keywords: Algorithm, landsat, salinity, sentinel, TSS, validation

1. Introduction

The observation of oceanographic conditions based on conventional terrestrial methods in a large area will take time and money, then the combination of remote sensing technology and Geographical Information Systems (GIS) can be a solution to reduce costs, time and energy in terrestrial surveys [1]. In the marine sector, satellite image can be used to observe and analyse water conditions [2]. One of the methods to extracting coastal and marine parameter such as water quality assessment is using the empirical water quality algorithms [3]. Several empirical water quality algorithms for remote sensing data extraction are sea surface temperature, total suspended solid (turbidity), salinity and chlorophyll-a concentration [4].

In this study, the combination of multispectral remote sensing data namely, Sentinel-2 and Landsat 8 is used to compare the performance of each satellite images. The Sentinel-2 and Landsat 8 image have a spatial resolution of 10 m [5] and 30 m [6] respectively, which are categorize as a medium resolution images. Both images can provide better accuracy than other free satellite images.

The temporal resolution of each Sentinel-2 and Landsat 8 image is 5 days [5] and 16 days [6], respectively. So, these images can detect variations in oceanographic changes at short intervals (per month). Both images have various sensitive bands to detect the value of several oceanographic
parameters [7], so these images can be used as a tool to implement algorithms of several oceanographic parameters.

To evaluate the performance of empirical water quality algorithm, Palabuhanratu Bay is chosen as a tested site. This bay has a very high potential for marine resources, but there is not much research that has observed this. Previous studies in Indonesia generally examined the validation of an algorithm for oceanographic parameters using the same satellite image [8-10]. Therefore, this study compares two different multispectral satellite images to validate several oceanographic parameter algorithms. This algorithm can be used as an initial step in determining water quality which can then be used for waste monitoring, aquaculture, determining potential fish catch zones, and so on.

2. Methodology
The study area of this research is in Palabuhanratu Bay where is located in Sukabumi Regency, West Java Province, Indonesia at coordinates 06° 57’ – 07° 07’S and 106° 22’ – 106° 33’E. The purposive sampling method was used in this study to determine the survey sample points (figure 1). The purposive sampling method is a technique of taking samples based on considerations that focus on a specific goal by not thinking about random, regional, or strata collection [11]. This technique is used considering that the Palabuhanratu Bay area is quite wide, which is about 140 km$^2$, so that some of the sample points are considered to represent the entire research area.

Satellite images used in determining the algorithm are Sentinel-2 and Landsat 8 satellite images. Both satellite images are already in surface reflectance format. The algorithm used to estimate the salinity value is the Cimandiri [12], Cilamaya [13] and Wouthuyzen in [14] algorithm and the in-situ validation values are obtained from field surveys using a salinometer. In total suspended solid (TSS) estimation, the algorithm used is the Parwati in [15], Budhiman [16] and Lestari [17] algorithm, while the in-situ value is obtained through laboratory tests of water samples. These algorithms are generally built based on Landsat satellite image (table 1). However, in this study, the performance on the Sentinel-2 imagery is also tested.

![Figure 1. Survey sample points](image-url)
We are implementing the algorithms and extracting sample point values using Google Earth Engine application by uploading the sample point shapefile, calling the image with the image collection function, setting the date of the image with the date filter function, then creating a script according to the calculation of each algorithm. After that, we are extracting the value of each sample point with the export to table function.

The use of the algorithm in this study was taken based on previous research which had a similar characteristic area. From the various algorithms that can be used, it is necessary to see the suitability of the algorithm if implemented in the study area by looking at the correlation and accuracy of the in-situ value to the image value by calculating the correlation coefficient (R) and the error value (NMAE).

\[
\text{NMAE} \, (\%) = \frac{1}{N} \times \sum \frac{|x_{\text{estimated}} - x_{\text{measured}}|}{x_{\text{measured}}} \times 100
\]  

where, 
- \( \text{NMAE} \) : Normalized Mean Absolute Error 
- \( N \) : Amount of data 
- \( x_{\text{estimated}} \) : Value of processing yield 
- \( x_{\text{measured}} \) : The value of field measurements that are considered correct 

\[
R = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}[n(\Sigma y^2) - (\Sigma y)^2]}
\]  

where, 
- \( Y \) : Response variable or resultant variable (dependent) 
- \( X \) : Variable predictor or variable causal factor (independent) 
- \( N \) : Amount of data 

We used NMAE because NMAE can be used in different study areas to compare models, unlike other statistical test indicators which are highly dependent on local conditions [18]. The minimum NMAE value is < 30 % to be able to see the error value in the extraction of seawater parameter data from remote sensing data [19]. Algorithms that have correlation coefficient in a sufficient to strong range can be used as an estimator of the value of the distribution of oceanographic parameters. Interpretation of the strength of the relationship or correlation between two variables can refer to the table 2.

### Table 1. Salinity and TSS algorithm

| Parameter | Algorithm                                                                 | Equation                                                                 | Year |
|----------|---------------------------------------------------------------------------|--------------------------------------------------------------------------|------|
| Salinity | Cimandiri                                                                | 29.983+(165.047*Blue)-(260.227*Green)+ (2.609*Red)                        | 2016 |
|          | Wouthuyzen                                                               | -142.72* ((-61.182x^3) + (70.129x^2)- (34.022x)+4.885) +32.702           | 2008 |
|          | Cilamaya                                                                 | 139.556970+(86.21318*LN Blue)-(24.62518* LN Red)                         | 2019 |
|          | TSS                                                                       | Parwati                                                                  | 2006 |
|          |                                                                            | 7.9038*exp (23.942*Red)                                                  | 2004 |
|          |                                                                            | 3.3238*exp (34.099*Red)                                                  | 2006 |
|          |                                                                            | 24197x^3-22050x^2+6813x +664.98                                          | 2009 |

x = Blue chromaticity

LN= Natural logarithmic
3. Results and discussion

3.1. Salinity

The salinity value was obtained from field measurements on January 15, 2020. Estimation of the salinity value used Sentinel-2 image on January 15, 2020 (figure 2). Estimation of salinity values using Landsat 8 image on the adjacent date, namely 17 January 2020 (figure 3).

In algorithm testing, three kinds of algorithms are used, namely the Cimandiri algorithm [12], Wouthuyzen [14], and Cilamaya [13]. The Wouthuyzen algorithm has the lowest correlation coefficient (R) value on Sentinel-2 images and Landsat 8 images, respectively -0.12 and 0.004, which shows that the in-situ value with the image value has a weak relationship. The NMAE value of the two images also has the highest yield on the Sentinel-2 image and the Landsat 8 image, respectively 175.21 % and 230.58 %. This algorithm cannot be applied to the study area because it has an NMAE > 30 %.

The Cimandiri algorithm has a correlation coefficient (R) on each Sentinel-2 image and the Landsat 8 image are -0.43 and 0.30, which shows that the in-situ value with the image value has a sufficient relationship. The NMAE value in the Sentinel-2 image and Landsat 8 images were 19.63 % and 16.59 %, respectively. This algorithm can be applied in the study area because it has an NMAE value < 30 %, but in the Cilamaya algorithm, the correlation coefficient (R) in the Sentinel-2 image and Landsat 8 image is higher than the Cimandiri algorithm, respectively 0.53 and 0.40 which shows that the in-situ value with the image value each has a strong and sufficient relationship. The NMAE values in Sentinel-2 image and Landsat 8 images were 11.16 % and 27.16 %, respectively. Because the Cilamaya algorithm has a higher correlation coefficient and the NMAE value is still tolerable, the Cilamaya algorithm will be used to estimate the salinity distribution in Palabuhanratu Bay.

| Correlation coefficient (±) | Explanation          |
|-----------------------------|----------------------|
| 0                           | No correlation       |
| 0–0.25                      | Very weak correlation|
| 0.25–0.5                    | Sufficient correlation|
| 0.5–0.75                    | Strong correlation    |
| 0.75–0.99                   | Very strong correlation|
| 1                           | Perfect correlation   |

Table 2. Correlation coefficient [20]

![Figure 2. Salinity value from Sentinel-2.](image-url)
3.2. Total Suspended Solid (TSS)

The TSS value was obtained from field measurements on January 15, 2020. The TSS value estimation used Sentinel-2 image on January 15, 2020 (figure 4). Estimation of TSS values using Landsat 8 image on the adjacent date, namely January 17, 2020 (figure 5).

In algorithm testing, three kinds of algorithms are used, namely the Parwati [15], Budhiman [16], and Lestari [17] algorithm. The Lestari algorithm has the lowest correlation coefficient (R) on Sentinel-2 images and Landsat 8 images, respectively 0.29 and 0.09, which shows that the in-situ value and image value have a weak relationship. The NMAE value of the two images also had the highest yields on the Sentinel-2 image and the Landsat 8 image at 57.57 % and 1053.14 %, respectively. This algorithm cannot be applied to the study area because it has an NMAE > 30 %.

The Parwati algorithm has a correlation coefficient (R) on the Sentinel-2 image and the Landsat 8 image of -0.45 and -0.20, respectively, which shows that the in-situ value with the image value respectively has a strong until weak relationship. The NMAE value in Sentinel-2 image and Landsat 8 images were 2.99 % and 6.93 %, respectively. This algorithm can be applied in the study area because it has an NMAE value of < 30 %, but the Budhiman algorithm has a higher correlation coefficient (R) and lower NMAE than Parwati’s algorithm.

The Budhiman algorithm has a correlation coefficient (R) on the Sentinel-2 image and the Landsat 8 image, respectively 0.61 and -0.34, indicating that the in-situ value with the image value respectively has a strong and sufficient relationship. The NMAE values in the Sentinel-2 image and the Landsat 8 images are 25.37 % and 6.69 %, respectively, which enter the NMAE tolerance value, so the Budhiman algorithm will be used in estimating the TSS distribution in Palabuhanratu Bay.

The determination of the suitable algorithm for the study area is determined by the value of the correlation coefficient (R) and NMAE (table 3). Sentinel-2 and Landsat 8 image is used to compare how the performance of each image is based on the error value. In the salinity and TSS algorithms, the correlation coefficient (R) on Sentinel-2 is higher, and the error value (NMAE) is lower than the Landsat 8. This can be caused by taking different Landsat 8 image for several days with different field measurements and image processing at each service provider.

Sentinel-2 satellite image has a high correlation coefficient value and low NMAE compared to Landsat 8 satellite image in the salinity and TSS algorithms. So, Sentinel-2 satellite image has good performance for implementing the algorithms that built on Landsat image, so that algorithms can be used to detect water quality values such as salinity and TSS in Sentinel-2 satellite image.

However, we detected a poor relationship between variables that was not more than 0.6. This occurs due to the lack of sample point data due to cloud cover in the image, so that
the cloud-covered sample points cannot be used. In addition, the waters of Palabuhanratu Bay are quite large, which makes sampling points not available in one day, even though satellite images only record on the same day. Thus, we assume the available sample points are representative of the entire area.
Based on the above calculations by considering the NMAE and R values, the suitable algorithms to be implemented in the study area are the Cilamaya algorithm [13] to estimate the salinity value and the Budhiman algorithm [16] to estimate the total suspended solid value in Palabuhanratu Bay.

3.3. Distribution of salinity and TSS

The salinity values on the Sentinel-2 satellite image is in the range 5–15 ‰. The salinity value on the coast is indicated by class 10–15 ‰, while in the offshore it is in the class 5–10 ‰. The average value of the salinity concentration in Palabuhanratu Bay is 9.7 ‰. Based on the Venice System [21] salinity values in the 5–18 ‰ range belonging to the meso-mesohaline (brackish) category. The spatial distribution of salinity in the Sentinel-2 satellite image has a greater concentration in the coastal area and decreases towards the offshore. According to Garisson, the effect of river water makes salinity variations in coastal waters greater than in offshore [22], this can be caused by the rainy season so that a lot of river water enters the sea (figure 6).

The salinity value in the Landsat 8 satellite image is in the range of 5–10 ‰. The average salinity concentration in Palabuhanratu Bay is 6.5 ‰ which is included in the meso-mesohaline (brackish) category based on the Venice System [21]. Palabuhanratu Bay has brackish salinity which can be caused by the large amount of river water that empties into the sea during the rainy season. The spatial distribution of salinity in Landsat 8 satellite image on the coast and in the high seas does not have a significant variation.

TSS values on Sentinel-2 satellite image are in the range of 10-30 mg/L. The TSS value on the coast is in the class 15–30 mg/L, while in the offshore it is in the class 10–15 mg/L. The average value of TSS concentration in Palabuhanratu Bay is 13.7 mg/L. According to Permatasari et al., the TSS value < 25 mg/L means that TSS does not have a significant effect in fisheries [23]. The TSS spatial distribution in the Sentinel-2 satellite imagery has a greater concentration in the coastal area and the lower it is towards the offshore. This is because during the rainy season, water from the river brings more sediment and empties into the sea (figure 7).

![Figure 6. Salinity distribution in Palabuhanratu Bay.](image-url)
The TSS value on the Landsat 8 satellite image is in the range of 0–50 mg/L. The average value of the TSS concentration in Palabuhanratu Bay is 3 mg/L which has no significant effect on fisheries [23]. The spatial distribution of TSS on Landsat 8 around Cimandiri Estuary has a higher concentration and decreases toward the offshore. This can be caused by a buildup of sediment carried by the river flow.

From both images we can see that salinity and TSS distribution on Sentinel-2 satellite image has more accurate to the field data, this can be happened because Sentinel-2 satellite image was recorded on the same day as the field survey time, unlike the Landsat 8 satellite image which was recorded two days later. Both salinity and TSS has higher value in coastal zone than offshore, this can be happened during the rainy season so lots of freshwater mixed with sea water.

4. Conclusion
Sentinel-2 imagery has a good performance for implementing algorithms that built on Landsat imagery because it has a high correlation coefficient value compared to the extraction value from Landsat 8 satellite imagery. So, the algorithm that build on Landsat imagery can detect water quality parameter values such as salinity and TSS in Sentinel-2 imagery also.

Algorithms that are suitable for implementation in the study area are the Cilamaya algorithm [13] to estimate salinity values and the Budhiman algorithm [16] to estimate the total suspended solid value in Palabuhanratu Bay. In future studies, we suggest taking a large number of sample points during the rainy and dry seasons in order to get a comparison of the distribution of oceanographic conditions in these seasons and increase the accuracy of the research.

Acknowledgments
This work was financially supported by Universitas Indonesia under research grant PUTI Q2 2020 with grant contact number NKB-1662/UN2.RST/HKP.05.00/2020.
References

[1] Yusuf D and Rijal S A S 2017 Buku Ajr Penginderaan Jauh (Gorontalo: UNG Press)
[2] Andres L, Boateng K, Borja-Vega C and Thomas E 2018 Water 10 756
[3] Topp S N, Pavelsky T M, Jensen D, Simard M and Ross M R V 2020 Water 12 169
[4] Sarah N F Z, Rivani A, Puspitasari B, Ikhwansyah F, Maulidyah F, Dwi R D, Eka S P and Widyatmanti W 2017 Aquacultura Indonesiana 17 60-8
[5] ESA 2012 Sentinel - 2: ESA’s Optical High-Resolution Mission for GMES Operational Services (Leiden: ESA Communications)
[6] USGS 2013 Using the USGS Landsat Level-1 Data Product Landsat Mission available at https://www.usgs.gov/core-science-systems/nli/landsat/using-usgs-landsat-level-1-data-product
[7] Pahevan N, Chittimalli S K, Balasubramanian S V and Vellucci V 2019 Remote Sens. Environ. 220 19-29
[8] Supriatna S, Pin T G and Kalipaksi R 2018 AIP Conf. Proc. 2023 020181
[9] Kurniadin N and Maria E 2020 Jurnal Geodesi dan Geomatika 03 64-70
[10] Jaelani M L, Sulistyah U D and Winarso G 2016 Journal of Geodesy and Geomatics 12 1-9
[11] Arikunto S 2006 Prosedur Penelitian Suatu Pendekatan Praktek (Jakarta: Rineka Cipta)
[12] Supriatna L, Supriatna J, Koetsoer R H and Takarina N D 2016 AIP Conf. Proc. 1729 020079
[13] Kaffiah S, Supriatna and Damayanti A 2020 IOP Conf. Ser. Earth Environ. Sci. 481 012071
[14] Nafizah L M, Jaelani and Winarso G 2016 Jurnal Teknik ITS 5

[15] Mubarok I D, Rifardi and A Tanjung 2019 Jurnal Perikanan dan Kelautan 24 119-29
[16] Budhiman S, Hobma T W and Vekerdy Z 2004 Mapping TSM Concentrations from Multisensor Satellite Images in Turbid Tropical Coastal Waters of Mahakan Delta, Indonesia, International Institute for Geo-Information Science and Earth Observation
[17] Lestari I B 2009 Pendugaan Konsentrasi Total Suspended Solid (TSS) dan Transparansi Perairan Teluk Jakarta dengan Citra Satelit Landsat (Bogor: IPB University) available at https://repository.ipb.ac.id/handle/123456789/12523
[18] Isa M M 2016 Analisis Kualitas Air untuk Budidaya Rumput Laut Eucheuma Cottoni dengan Citra Landsat 8 STUDI Kasus: Laut Selatan Pulau Lombok, Nusa Tenggara Barat Bachelor project (Surabaya: Institut Teknologi Sepuluh Nopember)
[19] Jaelani L M, Setiawan F and Matsushita B 2015 Uji Akurasi Produk Reflektan-Permukaan Landsat Menggunakan Data In situ di Danau Kasumigaura, Jepang Proc. Pertemuan Ilmiah Tahunan Masyarakat Ahli Penginderaan Jauh Indonesia, Bogor 2015 464-70
[20] Sarwono J 2006 Metode Penelitian Kuantitatif dan Kualitatif (Yogyakarta: Graha Ilmu)
[21] The Venice System for the Classification of Marine Waters According to Salinity 1958 available at https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.4319/lo.1958.3.3.0346
[22] Garisson T 2004 Essentials of Oceanography 5th edition (Australia: Brooks/Cole Cengage Learning) vol 22 pp 238-9
[23] Permatasari Y D, Priyantari N, Cahyono B E, Fisika J and Jember U 2019 Jurnal Geografi Gea 19 55-66