Empirical vs Semi-Analytic Model for Total Suspended Solid Detection

Pingkan Mayestika Afgatiani[1,*] Syifa Wismayati Adawiah[2] Syarif Budhiman[1]

[1]*Research Centre for Remote Sensing, Research Organization for Aeronautics and Space, National Research and Innovation Agency (BRIN) Indonesia*
[2]*Directorate of Environment, Maritime, Natural Resources, and Nuclear Policy, Deputy for Development Policy, National Research and Innovation Agency (BRIN) Indonesia*
[*]*Corresponding author. Email: pingkan.mayestika.afgatiani@brin.go.id*

ABSTRACT

Many parameters are used to determine the quality of waters, one of which is Total Suspended Solids (TSS). In addition to direct data measurement and field sampling, TSS concentrations can also be estimated from satellite data by developing models based on the reflectance of the light received by the sensor. This article aims to compare two available empirical and semi-analytic models that can estimate TSS concentrations in Bekasi coastal waters using Landsat 8 and Sentinel 2 satellite data. The empirical model was made in 2018 in Bekasi coastal waters, while the semi-analytic model was created in 2004 in the Mahakam coastal waters. As a validation using field data from the Bekasi coastal waters, taken in 2019. The results showed that the semi-analytic model has a smaller error value than the empirical model, with RMSE 51.4 mg/l and 585777.2 mg/l respectively. The result indicates that the semi-analytic model can better estimate TSS even though it is applied at different times and locations. In contrast, the empirical model shows a very high error even though it is used in the same area when the empirical model was created. For further study, we tried to apply the semi-analytic model to the Sentinel-2 image, and it was found that the semi-analytic model also has good capabilities with a lower RMSE value of 44.1 mg/l. In conclusion, the semi-analytic model is better for extracting TSS information than the empirical model.

Keywords: *water quality, suspended matter, water pollution, Landsat-8 satellite*

1. INTRODUCTION

Total Suspended Solid (TSS) is one of the physical parameters that indicate the quality of the waters. Commonly, in situ and laboratory measurement methods are used to determine the TSS concentration. But these common methods are both time consuming and costly. Other method that can be done to overcome those issues is remote sensing technology. Some physical parameters, such as total suspended solids, can be identified using remote sensing, based on the reflection of light from the water column received by satellite sensors.

The ability of satellite sensors to detect objects using spectral reflections is very interesting to study. Generally, existing research uses Landsat imagery as the primary material in the analysis and is combined with models that have been built previously, as in the [1], [2], [3], [4].

Many models have been developed to determine the TSS content in waters [5]–[9] [10] explained that three types of models are commonly used in analyzing TSS: empirical, theoretical, and semi-analytic. The empirical model is the simplest method. Generally, the empirical model is site specific. Empirical model that was developed in one area cannot be used in different area. Empirical models can be used to determine the analytical process's initial value to improve the process by narrowing the range of variation [11]. The theoretical model is based on radiation transfer theory for spectrum simulation at the Top of the Atmosphere (TOA) with different TSS and atmospheric conditions [10]. A semi-analytic model for TSS analysis using in situ Inherent Optical Properties (IOP) information. IOP is an inherent property of water constituents that occurs when they interact with light, such as absorption, reflection and transmission of light. The specific IOP used as the model

A. D. Saputro et al. (Eds.): ICOSEAT 2022, ABSR 26, pp. 711–716, 2023.
https://doi.org/10.2991/978-94-6463-086-2_94

parameter has the potential to produce the best TSS value [12], because Specific IOP is related to the individual scattering and absorption of TSS. This study aims to detect the distribution of TSS using Landsat 8 and Sentinel 2 satellite imagery with two available empirical and semi-analytic models.

2. METHOD

2.1. Study Area

This research is conducted in the northern coastal waters of Bekasi Regency, West Java (Figure 1). This location might be affected by fishery activities in the form of fish ponds, mangrove tourism, fishing activities, and industries. This location is close to the mouth of the Citarum River where the river mouth forms a delta due to sedimentation.

Figure 1 Study Area

2.2. Processing Satellite Data and Validation

In general, this study covers the process of image processing using Google Earth Engine (GEE) using both empirical and semi-analytic models to estimate TSS concentration and continue with validation using in situ data (Figure 2).

The image data used are Landsat 8 and Sentinel 2. We used Landsat 8 date of acquisition are 1st to 30th August 2019, while Sentinel 2 date of acquisition is 23rd August 2019. In general, the study was using surface reflectance data available in GEE, delineating water and non-water areas, and using available model to determine the TSS concentration. The data used is Landsat 8 surface reflectance data for estimating TSS concentration, and Sentinel 2 surface reflectance data for further study and verification of the best model between empirical and semi-analytic. This data set is atmospherically corrected surface reflection from the Landsat 8 OLI/TIRS sensor. Furthermore, the Normalized Difference Water Index (NDWI) analysis was performed to delineate water and non-water areas. Pixels with a value of 0 in masking are excluded from the calculation. Next, convert the image to RGB 8-bit for display.

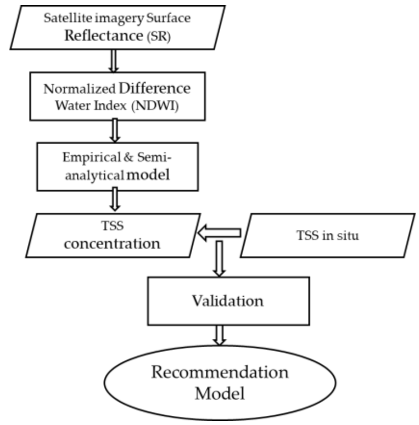

Figure 2 Flow Chart

The TSS concentration was measured using the gravimetric method carried out in the laboratory [13]. Moreover, TSS analysis uses two models, namely the semi-analytic model built by [14] and the empirical model built by [15]. The semi-analytic model (2004) was built in the Mahakam Delta coastal waters, and the empirical (2018) was built in the Bekasi coastal waters. Both of these models were verified with field data. The semi-analytic model is known to have an R^2 value of 0.94, while the empirical (2018) model has an R^2 value of 0.824 with respective field data.

Semi-analytical (2004) model as follows Equation (1):

$$MPT\left(\frac{mg}{L}\right) = 8.1429(\exp(^{23.704\ x\ A})) \tag{1}$$

Meanwhile, empirical (2018) model as follows Equation (2):

$$MPT\left(\frac{mg}{L}\right) = 155.28B^3 - 2740.4B^2 + 15912B - 30261 \tag{2}$$

Note:
A = red band
B = $\frac{Red\ Band}{(Green\ Band)(Green\ Band)}$

The next step is testing the validation between the results of observations and estimates. Validation testing uses Root Mean Square Error (RMSE), as follows Equation (3):

$$RMSE = \sqrt{\frac{\Sigma(X\ obs - X\ est)^2}{N}} \quad (3)$$

Where X est is the value of the image processing results, X obs is the value of the field measurement results. This validation test will help determine which model approximates the TSS concentration measured in the field.

Furthermore, the selected model was verified by applying it on the Sentinel-2 data on August 23rd, 2019, to see the model's ability to extract TSS concentration on different satellite data. Sentinel-2 imagery is processed using GEE and using Surface Reflectance data. The steps taken are the same as processing with Landsat 8, including the RMSE calculation.

3. RESULT AND DISCUSSION

3.1. Result

The resulting data processing is shown in three different tables. Table 1 shows information on TSS concentration with three methods: in situ, empirical, and semi-analytic. Table 2 shows information on validation tests between empirical and semi-analytic models. In contrast, table 3 shows information on the estimation results and the difference between in situ and Landsat 8 and Sentinel 2 image processing.

The results show a very significant differences in the estimated concentration of the empirical model compared to the semi-analytic model (Table 1). The TSS concentration calculated using the empirical model has a closer value to the field measurements and is stable, although overall the error is still high (more than 50 mg/l).

A validation test was conducted to determine the extent to which the two models can estimate TSS concentration (Table 2). In the empirical model, the correlation coefficient (R^2) value is lower compared to the semi-analytic. On the other hand, the RMSE value is very high compared to semi-analytic.

Based on the results of the validation test, it seems that semi-analytical models have a better ability to estimate TSS concentrations. Furthermore, the model is tested on Sentinel 2 images to find out whether the model has the same capabilities. The calculation results show that the semi-analytic model extraction on Sentinel-2 have varying concentrations, but the RMSE is better than Landsat-8.

Table 1. TSS concentration (mg/L) pada three method: in situ data, empirical, and semi-analytical models

TSS concentration (mg/L)		
In situ	Empirical	Semi-analytical
80.7	639.9	70.9
36.0	9100.8	12.9
63.3	14365.8	12.5
52.0	186278.6	15.7
80.3	82573.7	18.6
67.0	18999.3	17.4
93.0	47731.5	16.1
71.0	38.5	12.2
61.0	14426.4	11.7
101.3	209.4	19.7
79.3	193.1	25.3
77.7	95.8	58.2
113.0	140.1	155.8

Table 2. Validation Test

Empirical		Semi-analytical	
R^2	RMSE	R^2	RMSE
0.1	58577.2	0.3	51.4

Table 3. Estimation TSS concentration (mg/L) using semi-analytic model

In situ	Landsat 8		Sentinel 2	
	Est.	Diff.	Est.	Diff.
80.7	70.9	9.8	80.6	0.1
36.0	12.9	23.0	34.6	1.4
63.3	12.5	50.9	33.8	29.6
52.0	15.7	36.3	20.0	32.0
80.3	18.6	61.7	24.8	55.6
67.0	17.4	49.6	23.8	43.2
93.0	16.1	76.9	25.4	67.6
71.0	12.2	58.8	23.1	47.9
61.0	11.7	49.3	21.4	39.6
101.3	19.7	81.6	18.9	82.4
79.3	25.3	54.0	35.7	43.6
77.7	58.2	19.4	48.3	29.4
113.0	155.8	-42.8	132.9	-19.9
			RMSE Sentinel 2	44.1

3.2. Discussion

It seems that the TSS concentration between empirical and semi-analytic models is very much different. Empirical models tend to produce very high TSS concentrations. The research [16] also show that TSS estimation with an empirical model will create a much larger value than in situ. The development of the model took samples of Bekasi coastal waters. In contrast to the semi-analytic model, which was developed from different location data, it gives higher accuracy. This model has been used as a reference in other studies [9], resulting in that this method had a very high R^2 value reaching 0.982, and [1] also state that this method produces a minor error. It can happen because of the time differences when taking image data with in situ data. In addition, dynamic water conditions can significantly affect these results. In addition to the time of image data collection, there are other factors, such as the quality of image data that is not cloud-free resulting in the surface reflectance value having more errors as well as image data recorded on only one date, while TSS measurements can be carried out in several days. The Empirical (2018) model is based on the observed relationship between optical properties and TSS concentration in the field. The empirical relationship is geographically specific, so this cannot be applied in other areas [10]. This study also shows that the empirical model is time-specific even though it is carried out at the same location. It also indicates that it is necessary to take samples under different water conditions from time to time.

Meanwhile, the semi-analytical (2004) model is built using a semi-analytic model. This model uses the relationship between seawater IOP and TSS concentration. The study of [10] explained that the semi-analytic model has a higher inversion precision and universality than the empirical model. The same phenomenon can also be seen in the study results, which show that the TSS concentration generated by semi-analytic have a better correlation than the empirical model. Another study [17][18] also use a semi-analytic model to estimate TSS in inland waters and get a lower error of 30%.

Validation testing is carried out in the next step. This validation test was conducted to determine the extent of the similarity between the image processing results and the in situ results. The validation test used Landsat 8 images acquired from August 1 to August 30, 2019, while in situ data retrieval was carried out on August 23-25, 2019. The data used is a mosaic image data in GEE. The results of the validation test are shown in Table 2. The correlation coefficient determines the magnitude of the relationship between two or more variables. The RMSE value is the standard deviation of the residual or prediction error. RMSE provides information about the magnitude of the error in the concentration value between the predicted value and the actual value at the same concentration unit [19]. The value of R^2 and the correlation coefficient in both models is low. The resulting RMSE value also turns out to be very large for the empirical model, which is 58,577.2 mg/l, while the semi-analytic model is 51.4 mg/l. It certainly shows that the semi-analytic model is more suitable for the analysis of the distribution of TSS in Bekasi coastal waters. The difference in sediment content within water column (i.e.: clear water, moderately turbid water, highly turbid water, extremely turbid water) can also affect the calculation of the estimated TSS concentration. Each of previous type of water, based on sediment content, has a different wavelength absorption ability, so the models used are different [8]. Empirical models require further processing to get more accurate results [5].

The results in Table 3 show that TSS estimated concentration using Sentinel-2, are higher and closer to the field value than Landsat 8. It has an RMSE value of around 44.1 mg/l, while the RMSE of Landsat 8 is 51.4 mg/l. It shows semi-analytic models retrieving TSS concentrations for Landsat 8 and Sentinel 2. It is possible because the difference between the TSS estimation concentration between Landsat 8 and Sentinel 2 might be due to the difference in wavelength between the bands.

The semi-analytic model provides a stable value in that the overall concentrations tends to be homogeneous, and the TSS estimation is low and does not differ much. At the same time, the TSS results in the field are quite varied, from the lowest concentrations (10-20 mg/L) to moderate (>20-80 mg/L), and higher concentration (80 mg/L).. Additionally, [20] stated that in the east monsoon (June-August), only the coastal areas have high TSS content [21]. However, this value still contains the error of the reflected value by the water surface. To minimize error, future research needs to be done more about the data used, which is the value of water-leaving radiance [22].

4. CONCLUSION

The empirical model produces a very high TSS concentration than the semi-analytic model. The results of the validation test with in situ data show that the RMSE value is lower than the empirical model, which means that the semi-analytic model is more suitable for further analysis. The study's results also show that the empirical model not only depends on the location of the model preparation but also on the timing of the use of the empirical model. Although the empirical model was built in the same area, it will give a high error because it considers in situ concentrations more as input in the development of the model, if the concentration range when developing the empirical model does not represent the range of concentrations in the field.

AUTHORS' CONTRIBUTIONS

All authors have different contributions and responsibilities. Pingkan Mayestika Afgatiani as main contributor (concepting, data processing, writing, and editing), while Syifa Wismayati Adawiah and Syarif Budhiman as member contributors (writing).

ACKNOWLEDGMENTS

The author would like to thanks the Research Center for Remote Sensing, National Research and Innovation Agency (BRIN) which supports for this research.

REFERENCES

[1] L. Indeswari, T. Hariyanto, and C. Bekti Pribadi, Pemetaan Sebaran Total Suspended Solid (TSS) Menggunakan Citra Landsat Multitemporal dan Data In Situ (Studi Kasus : Perairan Muara Sungai Porong, Sidoarjo), *J. Tek. ITS*, vol. 7, no. 1, 2018, doi: 10.12962/j23373539.v7i1.28698.

[2] I. Sholihah, L. M. Jaelani, and T. Salam, Analisis Sebaran Padatan Tersuspensi dan Transparansi Perairan Menggunakan Landsat 8, *J. Tek. Its*, vol. 5, no. 2, 2016, pp. 5–8, doi: 10.12962/j23373539.v5i2.17175.

[3] M. Arief, S. W. Adawiah, M. Hartuti, and E. Parwati, Algoritma Dua Dimensi Untuk Estimasi Muatan Padatan Tersuspensi Menggunakan Data Satelit Landsat-8, Studi Kasus: Teluk Lampung, *J. Penginderaan Jauh dan Pengolah. Data Citra Digit.*, vol. 13, no. 2, 2017, pp. 109–120, doi: 10.30536/j.pjpdcd.2016.v13.a2517.

[4] E. Parwati and A. D. Purwanto, Time Series Analysis of Total Suspended Solid (Tss) Using Landsat Data in Berau Coastal Area, Indonesia, *Int. J. Remote Sens. Earth Sci.*, vol. 14, no. 1, 2017, p. 61, doi: 10.30536/j.ijreses.2017.v14.a2676.

[5] S. W. Adawiah, K. T. Setiawan, E. Parwati, and R. Faristyawan, Development of Empirical Model of Total Suspended Solid (TSS) by using Landsat 8 on the Coast of Bekasi Regency, *IOP Conf. Ser. Earth Environ. Sci.*, vol. 750, no. 1, 2021, doi: 10.1088/1755-1315/750/1/012039.

[6] J. Zhao, Remote Sensing Evaluation of Total Suspended Solids Dynamic with Markov Model : A Case Study of, *Sens. MDPI*, vol. 20, 2020, p. 6991.

[7] C. Wang, S. Chen, D. Li, D. Wang, W. Liu, and J. Yang, A Landsat-based model for retrieving total suspended solids concentration of estuaries and coasts in China, *Geosci. Model Dev.*, vol. 10, no. 12, 2017, pp. 4347–4365, 2017, doi: 10.5194/gmd-10-4347-.

[8] D. Jiang *et al.*, Remotely estimating total suspended solids concentration in clear to extremely turbid waters using a novel semi-analytical method, *Remote Sens. Environ.*, vol. 258, no. October 2020, 2021, doi: 10.1016/j.rse.2021.112386.

[9] T. Hariyanto, T. C. Krisna, Khomsin, C. B. Pribadi, and N. Anwar, Development of total suspended sediment model using landsat-8 OLI and In-situ data at the surabaya coast, East Java, Indonesia, *Indones. J. Geogr.*, vol. 49, no. 1, 2017, pp. 73–79, doi: 10.22146/ijg.12010.

[10] J. L. Kong *et al.*, A semi-analytical model for remote sensing retrieval of suspended sediment concentration in the Gulf of Bohai, China, *Remote Sens.*, vol. 7, no. 5, 2015, pp. 5373–5397, doi: 10.3390/rs70505373.

[11] M. Ligi *et al.*, Testing the performance of empirical remote sensing algorithms in the Baltic Sea waters with modelled and in situ reflectance data, *Oceanologia*, vol. 59, no. 1, 2017, pp. 57–68, doi: 10.1016/j.oceano.2016.08.002.

[12] P. Dorji, P. Fearns, and M. Broomhall, A semi-analytic model for estimating total suspended sediment concentration in turbid coastal waters of northern Western Australia using MODIS-Aqua 250 m data, *Remote Sens.*, vol. 8, no. 7, 2016, pp. 33–38, doi: 10.3390/rs8070556.

[13] R. B. Baird, A. D. Eaton, and E. W. Rice, *Standard Methods for the Examination of Water and Wastewater*. 2017.

[14] S. Budhiman, T. W. Hobma, and Z. Vekerdy, Remote sensing for Mapping TSM concentration in Mahakam Delta: an analytical approach, *13th Omi. Work. Valid. Appl. Satell. Data Mar. Resour. Conserv.*, no. January, 2004, pp. 5-1-5–14.

[15] M. A. Adawiah, Syifa Wismayati, Rizky Faristyawan, Nana Suwargana, Ety Parwati, Pengembangan Model Pemanfaatan Penginderaan Jauh untuk Ekstraksi Informasi Kualitas Perairan, *Lap. Kegiat. Litbangyasa (Tidak Dipublikasikan)*, no. 8, 2018, pp. 1–59.

[16] M. D. Camiolo *et al.*, An empirical remote sensing algorithm for retrieving total suspended matter in a large estuarine region, *Sci. Mar.*, vol. 83, no. March, 2019, pp. 53–60, doi: https://doi.org/10.3989/scimar.04847.22A.

[17] N. Bernardo, A. Carmo, and E. Park, Retrieval of Suspended Particulate Matter in Inland Waters with Widely Di ff ering Optical Properties Using a Semi-Analytical Scheme, *Remote Sens.*, vol. 11, 2019, doi: https://doi.org/10.3390/rs11192283.

[18] Z. N. Ghuvita Hadi, T. Hariyanto, and N. Hayati, Estimation of Total Suspended Sediment Solid in Porong River Waters Using Multitemporal Satellite Imagery, *IOP Conf. Ser. Earth Environ. Sci.*, vol. 936, no. 1, 2021, doi: 10.1088/1755-

1315/936/1/012006.

[19] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, Comparison of training approaches for photovoltaic forecasts by means of machine learning, *Appl. Sci.*, vol. 8, no. 2, 2018, doi: 10.3390/app8020228.

[20] B. Karbela, P. M. Afgatiani, and E. Parwati, INTERSEASONAL VARIABILITY IN THE ANALYSIS OF TOTAL SUSPENDED SOLIDS (TSS) IN SURABAYA COASTAL WATERS USING LANDSAT-8 SATELLITE DATA, vol. 17, no. 2, 2020, pp. 175–188.

[21] J. Jiyah, B. Sudarsono, and A. Sukmono, STUDI DISTRIBUSI TOTAL SUSPENDED SOLID (TSS) DI PERAIRAN PANTAI KABUPATEN DEMAK MENGGUNAKAN CITRA LANDSAT, *J. Geod. Undip*, vol. 6, 2016, pp. 41–47.

[22] P. M. Afgatiani, M. Hartuti, and S. Budhiman, Deteksi Sebaran Muatan Padatan Tersuspensi Dengan Model Empiris Dan Model Semi-Analitik Di Perairan Bekasi, *J. Ilmu dan Teknol. Kelaut. Trop.*, vol. 12, no. 2, 2020, pp. 341–351, doi: 10.29244/jitkt.v12i2.28138.