Development of Empirical Model of Total Suspended Solid (TSS) by using Landsat 8 on the Coast of Bekasi Regency

S W Adawiah1,2*, K T Setiawan2, E Parwati2, R Faristyawan2

1 Department of Marine Science, Faculty of Science and Mathematics, University of Indonesia, Indonesia
2 Remote Sensing Application Center-LAPAN, Indonesia

Email: syifa.w.adawiah@gmail.com

Abstract. The Coastal area is transitional areas between terrestrial and marine ecosystems which are still affected by changes on land and in the sea. Various activities carried out on the coast make this area vulnerable to environmental changes that originate from natural phenomena and human activities. The north coast of Bekasi is prone to water quality degradation due to its proximity to Jakarta Bay and the mouth of the Citarum River to the Java Sea. The decline in the quality of Bekasi coastal waters can be measured by several parameters such as salinity, turbidity, brightness, chlorophyll concentration and Total Suspended Solid (TSS), the presence of organic or inorganic material. Landsat 8 can estimate the concentration and distribution of TSS spatially and temporally. This study aims to develop an empirical model for estimating TSS concentrations in the coastal area of Bekasi Regency by using Landsat 8 data. This method used a combination regression from five channels. The results showed there are four empirical models with a value of $R^2$ of 0.5. The calculation of the accuracy value using RMSE is still high, which is around 100, but the construction of this model is better than the previous algorithm. Another alternative is to use multiple linear regression to see the relationship between the interaction variables of each band and the TSS concentration.

1. Introduction

Coastal areas are transitional areas between terrestrial and marine ecosystems, which are affected by changes in land and sea. This area also has a high productivity and is a support zone for community economic activities such as fishery and seaweed cultivation, trading ports, and conservation areas. Human activities from land and coast have made this area vulnerable to environmental changes. Environmental changes in coastal areas such as abrasion and sedimentation that occur on the north coast of Java. One of the coastal areas that have ecological and economic value for the surrounding community is the coast in the north of Bekasi Regency, namely Muaragembong District.

The north coast of Bekasi is prone to water quality degradation due to its proximity to Jakarta Bay and the mouth of the Citarum River to the Java Sea. In addition, changes in land conditions in the Citarum watershed have resulted in erosion and sedimentation at the mouth of the estuary and along the coast. The increased land area around the river mouth was caused by the supply of sediment carried by the river currents [1]. The effect of abrasion and sedimentation can be seen with changes in the coastline along the Muaragembong coast from 2007-2015 where sedimentation occurred of the Citarum downstream [2]. The most obvious change in this land is the reduction in mangrove area, reduction in mangrove land due to conversion into residential areas and ponds [3]. Some of the causes of these cases can lead an environmental changes, especially the quality of coastal waters. The decline in the quality of Bekasi coastal waters can be measured by several parameters of water quality such as salinity, turbidity, brightness, chlorophyll concentration and Total Suspended Solid (TSS), the presence of organic or inorganic material, to the concentration of heavy metals. Several physical parameters of water quality also become a reference in determining the suitability and quality of the Gracilaria sp. in Pantai Mekar and Pantai Bahagia Villages, Muaragembong [4]. Based on the research, the carrying capacity of coastal land in the Muaragembong has
decreased fisheries productivity every year [5]. The water quality parameters that can be identified easily in the field is turbidity. Turbidity can be recognized through changes in water color clarity. Turbidity consists of particles that mix in water to form colloids and form suspended solids. Suspended Solids or known as TSS are solids that are present in solution but not dissolved, can cause the solution to become cloudy, and cannot immediately settle to the bottom of the solution. Suspended solids and other dissolved colored materials can reduce the clarity of water by making the appearance blurry, fuzzy, or muddy. Turbidity measurements are often used as an indicator of water quality based on clarity and an estimate of the total suspended solids in the water. Turbidity and sedimentation in coastal areas can be seen using remote sensing data. Remote sensing data is effective in monitoring changes that occur in coastal areas such as turbidity and estimate TSS concentration from time to time. Turbidity changes on the Muaragembong coast can be seen visually while TSS concentrations changes can be analyzed through a regression model approach using Landsat 8 image data.

Landsat 8 can estimate the concentration and distribution of TSS spatially and temporally. Estimation of TSS concentration is carried out using a simple empirical approach using linear regression methods, namely by using the reflectance value of each band or band combination which is then correlated with the TSS concentration obtained from the field survey to produce the TSS algorithm. Several studies using this method and approach have also been carried out using both single and multi-band. The TSS model using a single band or combination band is selected based on the highest correlation value such as the correlation generated using a linear regression model with green or red bands [6]. TSS model was created by connecting the TSS value derived from the image data with the actual value in the field, the TSS prediction value from the image data based on the visible band ratio, and the visible band ratio with NIR band [7]. Landsat multitemporal data can evaluate conditions in a research area where land changes can affected an increase or decrease in TSS concentrations in Berau waters [8]. The development of empirical models for estimating TSS in waters is very local in nature and only applies in specific areas so that the model that has been developed cannot be applied regionally or globally [9], [10]. This study aims to develop an empirical model for estimating TSS concentrations in the coastal area of Bekasi Regency by using Landsat 8 data.

2. Data and Methodology
The data used include the Landsat 8 satellite image data for the Bekasi Regency area on the path / row 122/064 with the acquisition date of 22 July 2018, while the TSS concentration value was obtained from a field survey that was carried out on 6-9 August 2018. The total TSS sample was 25 samples were divided into 15 samples to build empirical models, while 10 samples were used to test the accuracy of the model. The Landsat 8 data obtained is still in the form of raw data so it needs to be converted into a reflectance value using the formula 1. The image reflectance used comes from bands 1, 2, 3, 4, and 5. The numerical model is needed to describe the relationship between the reflectance value from single band or multi band on Landsat 8 and TSS value in the field. The model development uses simple regression, namely linear, exponential, logarithmic, and polynomial orders 3, 4, and 5 with single and ratio multiband, while the accuracy test is done by calculating the Root Mean Square Error (RMSE) of the developed model (formula 2). The research location shown in Figure 1.

\[
\rho_{\lambda} = (M_{\rho} \times Q_{\text{cal}} + A_{\rho})/\cos (90° - \theta SE)
\]

\[
RMSE = \sqrt{\frac{\sum_{i=0}^{n} (\text{Observation}_i - \text{Estimation}_i)^2}{n}}
\]
The first step in this research is to change the digital number value into a reflectant value. The reflectant values used in this study are band 1 (blue coastal), band 2 (blue), band 3 (green), band 4 (red), and band 5 (NIR). Next step is developing the regression model: the reflectance value used comes from the use of a single band or multi band ratio. Then a simple regression analysis was carried out where the graph formed was the correlation between the reflectance band value (X axis) and the TSS concentration in the field (Y axis). The regression model using linear, exponential, logarithmic, polynomial order 3, 4, and 5. The regression model that resulted from this single and combination band reached ± 300 combinations and the best was chosen where the value of $R^2 > 0.5$. TSS concentration obtained from the results of field measurements consisting of a total of 25 samples divided into 15 samples were used to build the model while 10 samples were used for model validation measurements using RMSE. The flow chart of this research shown in Figure 2.
3. Results and Discussions

3.1. Numerical Model for TSS

The reflectance value of Landsat 8 is correlated with field data, then an algorithm is built from single band to the use of various bands and the graphical form and correlation value are seen. Overall, there are about 300 algorithms produced with the highest correlation value $R^2$ is 0.5 where the highest correlation is given to the algorithm that was built using a multi-band combination. It is known that high TSS concentrations are very sensitive when detected using red wavelengths while low TSS concentrations are detected at green wavelengths [11]. The results of the development of the visible band combination (blue and green band) model with the NIR band and the red band give $R^2$ value greater than that using only the visible band combination. The correlation graph formed from this research shown that the combination of red band and NIR band together produces a low of $R^2$ value so that in this research the best band combination is produced by combining visible band with red band (blue, green, and red) or visible band with NIR band (blue, green, and NIR). Figure 3 shows the best empirical model algorithm.
In the picture above, it can be seen that each algorithm has almost the same pattern. Figures 3(a) and 3(b) have a graphical pattern, a polynomial regression model with order of 4, a variable constant, and the same $R^2$ value, namely 0.5385; the difference between this algorithm in the use of the band as the value for the X variable. The band used in Figure 3(a) as the X value is $B1*B4/(B3*B4)$ while in Figure 3(b) as the X is $B1*B5/(B3*B5)$. The difference between the two algorithm models lies in the use of the red band and NIR, where in Figure 3(a) the red band is used while in Figure 3(b) the NIR band is used. The use of these two bands to develop the empirical model can be assumed that the two bands are sensitive to the presence of TSS in the waters. In addition, sedimentation and erosion on the Muaragembong coast are also high due to land use which is widely used for pond and this area is the Citarum River watershed, so it is suspected that high sedimentation resulted in high TSS concentrations. Sedimentation that occurs in the estuary of the Citarum River causes new land deposits which can be identified on satellite images using red and NIR bands which is sensitive to separate land and sea in the image. Another interesting thing about the two algorithms is the use of band 1 (coastal band) in building empirical models. It is known, the use of band 1 in the waters has not been explored further, even though band 1 is able to see changes in the intensity of colors mixed in water so that it can be used to monitor the concentration of chlorophyll and suspended sediments in the water. In Figures 3(a) and 3(b) the interaction between band 1 (blue coastal) and band 3 (green) is able to detect the presence of TSS concentrations so that the use of the two visible bands makes it possible to develop TSS models from satellite images.

In Figures 3(c) and 3(d) the algorithm model formed is a polynomial with order 4. The graphic pattern formed in Figure 3 (c) begins with a decreasing pattern then increases and at the peak it decreases again, while in Figure 3 (d) the graph begins with an up and down pattern then goes up again. From the two patterns, it can be seen that the pattern formed in Figure 3(c) is an inverse pattern from Figure 3(d). Another thing that differences is the use of the X variable where in Figure 3 (c) the value of the X variable is $B2*B5/(B3*B4)$ while in Figure 3 (d) the value of the X variable is $B3*B4/(B2*B5)$ with $R^2$ value of 0.4511. The use of bands 2 (blue) and 3 (green) in the algorithm is able to record the reflectance of water particles indicating that water with different constituents has different spectral characteristics [12]; despite that, the use of the red and NIR bands to develop the TSS model together gives a lower $R^2$ value because the use of the two bands together (red and NIR) gives a lower reflectance value near to zero.

From the four graphs, it is found that the empirical model formed is polynomial with order 4, which is different from other research which show that TSS estimation forms an exponential pattern using band 6 (Shortwave infrared-SWIR) to analyze Total Dissolved Suspended (TDS) [13]. Previous research regarding the TSS algorithm model that was formed produced an exponential and linear regression model, but in this research a 4th order polynomial model was produced.
3.2. The Accuracy Model of TSS
The results of the accuracy test of the model produce an RMSE value above 100. In Table 1, it can be seen that numerical models 1 and 2 are polynomial regression models with order 4 which have the same $R^2$ value of 0.5385 and RMSE of 103.927. The different is between models 1 and 2 is the variable constant X where in the model 1 where X is $B_1*B_4/(B_3*B_4)$ while the model 2 where X is $B_1*B_5/(B_3*B_5)$. The result of accuracy test in model 3 is polynomial order 4 with $R^2$ value 0.4202 has RMSE 94.135 where the X is $B_2*B_5/(B_3*B_4)$, while in model 4 it is polynomial order 3 with $R^2$ value 0.451 and RMSE 100.291 where the X is $B_3*B_4/(B_2*B_5)$. The results of this calculation indicate that the regression model for building the TSS algorithm still has a high error. The results of the RMSE calculation can be seen in Table 1 below.

| Mode | Numerical Model | Where X | $R^2$ | RMSE |
|------|-----------------|---------|-------|-------|
| 1    | $Y= -15567X^4 + 73689X^3 -126790X^2 + 95654X - 25888$ | $B_1*B_4/(B_3*B_4)$ | 0.5385 | 103.927 |
| 2    | $Y= -15567X^4 + 73689X^3 -126790X^2 + 95654X - 25888$ | $B_1*B_5/(B_3*B_5)$ | 0.5385 | 103.927 |
| 3    | $Y= 89288X^4 - 263224X^3 + 283544X^2 -132122X + 22571$ | $B_2*B_5/(B_3*B_4)$ | 0.4202 | 94.135 |
| 4    | $Y= 1741.6X^4 - 8682.5X^3 + 14053X - 7263.3$ | $B_3*B_4/(B_2*B_5)$ | 0.451 | 100.291 |

The high RMSE value generated from each model can occur due to the quality and quantity of data, such as the lack of data needed to build the model and to test the accuracy. In addition, the distribution of data that is possible heterogeneous results in data being an error for other data. The TSS concentration value of the field survey had the lowest value range of 42 mg/L to 270 mg/L so that the data range was wide. It is necessary to classify the TSS concentration data, for example dividing the concentration class into low, medium, and high levels. However, this does not guarantee that the RMSE score will get better.

3.3. Comparison Model of TSS with Previous Research
The development of an empirical model in Bekasi waters has been done before [14] by using the red band ratio and the green band, a linear regression model has been generated by the resulting algorithm, but the accuracy of the model has not been tested. The algorithm of the model, namely:

$$TSS = 155.28X^3 - 2740.4X^2 + 15912X - 30261 \quad (3)$$

Where X is: reflectance of red band / (reflectance of green band)$^2$ from Landsat 8

The accuracy test of this model using RMSE has been carried out by comparing the results of the empirical model with the semi-analytic model [15]. In that research [15] semi-analytic models provide a better RMSE value of 51.4 with $R^2$ of 0.3 while the RMSE of the empirical model worth 58577.2 with $R^2$ is worth 0.1. The results of the current research shows that the use of the polynomial model of order 4 is better than the linear regression model from previous research where the $R^2$ value from 0.1 to 0.5385 and the resulting RMSE value from 58577.2 to 103.927.

The development of empirical models often faces several obstacles, including the quality of the image data used. In this study, the quality of the Landsat 8 data used was dominated by haze on the western side of the coast so that some of the TSS samples obtained could not be used to build a model so that the amount of data used was reduced. In addition, when the field survey was to be predicted according to the date Landsat 8 passed, it turned out that the data at that time was of a worse quality.
than the previous image data. However, the empirical model approach is still an alternative to see the distribution and concentration of TSS spatially and temporally quickly.

An alternative model besides developing an empirical model is through a semi-analytic approach [16]. Semi-analytical model has higher inversion precision and universality than empirical model [17] so that the resulting error is lower. In addition, the use of empirical models can be developed using multiple regression models. For this study, two algorithms can be selected, namely $Y = -15567X^4 + 73689X^3 - 127690X^2 + 95654X - 25888$ with $X_1 = B1*B4/(B3*B4)$ and $Y = -15567X^4 + 73689X^3 - 127690X^2 + 95654X - 25888$ with $X_2 = B1*B5/(B3*B5)$ as input for the development of multiple linear regression models. Multiple linear regression analysis aims to determine whether or not there is an effect of the independent variable (X) on the dependent variable (Y), in this case the TSS concentration. The two algorithms become a variable X (independent variable) which can be seen the effect of all synthetic bands made with TSS data together.

4. Conclusion

Empirical models can describe the distribution and estimation of TSS concentrations spatially and temporally in coastal areas. This model is still far from the actual condition so that it needs the development of other models. The most likely thing is to develop a semi-analytic model and multiple linear regression. Multiple linear regression can be used as a reference to analyze the interactions that occur from each variable reflectance band to TSS concentration.

Acknowledgments

The Author would like to thanks the Director of Remote Sensing Application Center Dr. Rokhis Khomaruddin to support this research and Saintek program from Ministry of Research, Technology, and Higher Education for funding scholarship so that this article can be realized.

References

[1] Paryono, Damar A, Susilo S B, Dahuri R, and Suseno H, 2017 Sedimentasi Delta Sungai Citarum Kecamatan Muara Gembong, Kabupaten Bekasi J. Penelitian dan Pengelolaan Daerah Aliran Sungai, 1, p. 15-26.
[2] Paryono, Sukoharjo S S, Damar A, Susilo S B, Dahuri R, and Suseno H, 2016 Analisis Dinamika Spasial dan Temporal Penggunaan Lahan dan Implikasinya Terhadap Sedimentasi di Wilayah Pesisir DAS Citra J. Kelautan Nasional, 11, 3 p. 189-200.
[3] Putra H, Prasetyo L B, and Santoso N, 2016 Monitoring Perubahan Garis Pantai dengan Citra Satelit di Muaragembong Bekasi J. Pengelolaan Sumber Daya Alam dan Lingkungan, 6, 2 p. 178-186.
[4] Deswati R H and Luhur E S, 2014 Profil Budidaya dan Kelembagaan Pemasaran Rumput Laut (Gracillaria sp) di Kecamatan Muara Gembong, Kabupaten Bekasi, Jawa Barat. Bul. Mitra Riset Sosok Kelautan dan Perikanan, 9, 1 p. 31-34.
[5] Oktaviani S, Yuniviter, and Imran Z, 2019 Daya Dukung Optimum Berbasis Pola Tata Guna Lahan Pesisir di Muara Gembong, Kabupaten Bekasi J. Ilmu dan Teknologi Kelautan Tropis, 11, 1 p. 75-87.
[6] Arief M, Adawiah S W, Hartuti M, and Parwati E, 2017 Algoritma Dua Dimensi untuk Estimasi Muatan Padatan Tersuspensi Menggunakan Data Satelit Landsat-8, Studi Kasus: Teluk Lampung J. Penginderaan Jauh dan Pengolahan Data Citra Digital 13, 2 p. 109-120.
[7] Parwati E, Trisakti B, Carolita I, Kartika T, Harini, S, Dewanti R, 2006 Analisis Hubungan Penutup/Penggunaan Lahan dengan Total Suspended Matter (TSM) kawasan pesisir Segara Anakan Menggunakan Data Ideraja. J. Penginderaan Jauh. 1, 1 p. 87-97.
[8] Parwati E and Purwanto A D, 2017 Time Series Analysis of Total Suspended Solid (TSS) Using Landsat data in Berau Coastal Area, Indonesia J.of Remote Sensing and Earth Sciences (IReSES)14, 1 p. 61-70.
[9] Laili N, Arafah F, Jaelani L M, Subehi L, Pamungkas A, Koenhardono E S, and Sulisetyono A,
2015 Development of Water Quality Parameter Retrieval Algorithms for Estimating Total suspended Solid and Chlorophyll-a Concentration Using Landsat 8 Imagery at Poteran Island Water. International Geoinformation Conference, October p. 55-62.

[10] Wibisana H, Soekotjo, B M, and Lasmito U, 2018 Preliminary Study of Total Suspended Solid Distribution in Coastal Ujung Pangkah Gresik Based Reflectance Value of Landsat satellite imagery Indonesian Journal of Geography, 50, 1 p. 42-48.

[11] Hariyanto T, Krisna T C, Khomsin, Pribadi C B, and Anwar N, 2017 Development of Total Suspended Sediment Model Using Landsat 8 OLI and Insitu Data at the Surabaya Coast East Java Indonesia Indonesian Journal of Geography 49, 1 p. 73-79.

[12] Le C, Li Y M, Zha Y, Sun D, Huang C, Zhang H, 2011 Remote Estimation of Chlorophyll a in Optically Complex Water Based on Optical Classification Remote Sensing of Environment 115, 2 p. 725-737.

[13] Sukwika, T and Putra H. 2018. Analisis Sedimentasi dan Konsentrasi Atmosfer Pada Zona Mangrove di Muaragembong, Bekasi. J Pengembangan Kota, 6, 2 p. 186-195.

[14] Adawiah S W, Faristyawan R, Suwargana N, Parwati E, and Arief M, 2018 Pengembangan Model Pemanfaatan Penginderaan Jauh untuk Ekstraksi Informasi Kualitas Perairan Laporan Kegiatan Litbangyasa (not published). LAPAN p. 59.

[15] Afgatiani P M, Hartuti M, and Budhiman S, 2020 Deteksi Sebaran Muatan Padatan Tersuspensi Dengan Model Empiris dan Model Semi Analitik di Perairan Bekasi J. Ilmu Teknologi Kelautan Tropis, 12, 2, p. 341-351.

[16] Budhiman S, Hobma T W, and Vekerdy Z, 2004 Remote Sensing for Mapping TSM Concentration in Mahakam Delta: an Analytical Approach. Thirteen OMISSAR Workshop Validation and Application of Satellite Data for Marine Resources Conservation, October p. 1-15.

[16] Kong J L, Sun X M, Wong D W, Chen Y, Yang J, Yan Y, and Wang L X, 2015 A Semi Analytical Model for Remote Sensing Retrieval of Suspended Sediment Concentration in the Gulf of Bohai, China Remote Sensing, 7, 5 p. 5373-5397.