# AN IMPROVED METHOD FOR CALIBRATION AND VALIDATION OF MODIS CHLOROPHYLL-A IN THE NORTH MALACCA STRAITS

Mohd Nadzri Md Reba[1*], Nor Zafirah Ab Lah[1], Eko Siswanto[3], Katsuhisa Tanaka[4]
[1]Institute of Geospatial Science and Technology (INSTeG), Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor Darul Takzim, Malaysia,
Email: nadzri@utm.my, nzafirah89@gmail.com
[2]Japan Agency for Marine-Earth Science and Technology (JAMSTEC) 2-15, Natsushima-cho, Yokosuka-city, Kanagawa, 237-0061, Japan,
Email: ekosiswanto@jamstec.go.jp
[3]Japan International Research Centre for Agricultural Sciences (JIRCAS) 1-1, Ohwashi,Tsukuba, Ibaraki, 305-8686, Japan,
Email: katuhi@affrc.go.jp

*Corresponding author: nadzri@utm.my

**KEYWORDS:** Chlorophyll-a, Malacca Straits, MODIS, ocean color, spatial bias.

**ABSTRACT:** Estimation of Chlorophyll-a (Chl-a) for optically complex water from satellite is challenging as the ocean colour satellite has a coarse spatial resolution led to low pixel-to-point correlation. In data extraction, the spatial variation from pixel based and point based may lead to bias estimate and the scale effects can induce to systematic error of MODIS standard Chl-a algorithm, OC3M. This paper presents the aim to reduce the spatial bias in locally-tuned OC3M algorithm by employing spatial weight function onto the remote sensing reflectance, $R_{rs}$ extracted from the satellite data. This method can reduce the scale effect when using larger windows kernel for extracting $R_{rs}$. Spatial weight function on OC3M for kernel size of 7x7 provided the the absolute percentage difference about 50% and RMSE of 0.3. Yet, result of locally tuned OC3M with weight function enhanced the results by APD of 34% and RMSE of 0.17. This method offers better treatment on spatial induced bias and proved as alternative in calibration and validation of satellite derived Chl-a with higher quality of regression and more match-up samples. The spatial weight function in locally tuned OC3M algorithm is the simple and effective method in reducing the spatial induced bias of Chl-a estimation.

## 1. INTRODUCTION

Ocean colour remote sensing has been demonstrated as a useful tool to study marine biological production. In the past decades until present, many ocean colour bio-optical models, satellite retrieved chlorophyll algorithms and applications in ocean-colour remote sensing have been developed globally and locally. Moderate Resolution Imaging Spectroradiometer (MODIS) data commonly used in ocean colour remote sensing study as it is free to access and provided fairly good resolution at 1 km. However, in the case of satellite chlorophyll retrieval, overestimation of chlorophyll concentration (hereafter Chl-a) to the in-situ chlorophyll-a was reported particularly in the coastal areas (Darecki & Stramski, 2004; Volpe et al., 2007; Werdell et al., 2009). Therefore, there is a need to perform local calibration and validation (CAL/VAL) exercise of the satellite derived Chl-a.

Estimation of Chl-a for optically complex water from satellite is challenging as the ocean colour satellite has a coarse spatial resolution led to low pixel-to-point correlation. Ladner et al., (2007) has demonstrated that finer resolution is required for validation of coastal products in order to improve match-up samples between in situ data and the high spatial variability of satellite properties in coastal regions. The spatial variability induced by pixel-to-point based comparison of both acquisition may lead to the bias of Chl-a estimation due to the fact that the in-situ Chl-a concentration is homogeneous in the 0.1 water column that is used to compare with the Chl-a concentration retrieved from MODIS at a pixel size with the average concentration of ~1 $km^2$ water area. This scale effect can induce systematic error of standard MODIS derived Chl-a estimation model (OC3M) (Chen, Yi, & Wen, 2013).

This paper demonstrates an improved CAL/VAL exercise of MODIS Chla estimation in which pixel-to-point spatial variation is reduced by applying weight function on the scale effect. The CAL/VAL exercise was performed over the Malacca Straits in order to find the impact of coastal Chl-a as this regional sea is closely surrounded by Peninsular Malaysia and Sumatra Indonesia (Figure 1). Spatial difference on data acquisition are used to model the weight function and later, model is applied to the three-band MODIS OC3M algorithm. This paper presents the application of spatial weight function on OC3M algorithm to reduce spatial bias in locally-tuned OC3M algorithm.

## 2. MATERIALS AND METHODS

### 2.1 Data Acquisition

The primary remote sensing data is the MODIS-Aqua (MODISA) Level 2 reprocessed and published in version R2013, which is retrieved from the NASA Ocean Colour Website. The data contains Chl-a product derived by global OC3M algorithm and the remote sensing reflectance ($R_{rs}$) at ten band-centre wavelengths. SeaWiFS Data Analysis System (SeaDAS) version 6.3 with modified atmospheric correction tool (Gordon & Wang, 1994) is used to derived the daily MODISA Level 2 $R_{rs}$ that has been acquired from October 2011 to August 2012 to establish match-up MODISA sample corresponding to the in-situ attribute. In-situ data were taken at Paya Island station located in the northern part of Malacca Straits with irregular depth around 17 to 55 meter (Thia-Eng et al., 2000). Phytoplankton fluorescence data were measured continuously by using fluorometer INFINITY-CLW installed at 1 meter depth. From the fluorescence profiles measured by fluorometer, the Chl-a concentration was measured by implicating method proposed by Suzuki & Ishimaru (1990). Figure 1 shows map of location of in-situ stations, the northern station, ST1 (Paya Island) and the southern station (ST2), and yet this paper only discusses on the finding for ST1.

### 2.2 Locally tuned ocean color algorithm

Basically, the average of the $R_{rs}$ at 443, 488 and 547 nm (main wavelength as a function in Equation 2) within the 3 by 3 pixels is computed by excluding the negative value (i.e., null value) and positioning the centre of the kernel window at location where the in situ Chl-a were collected (e.g. Siswanto et al., 2011; Werdell et al., 2009) as shown in Figure 1.

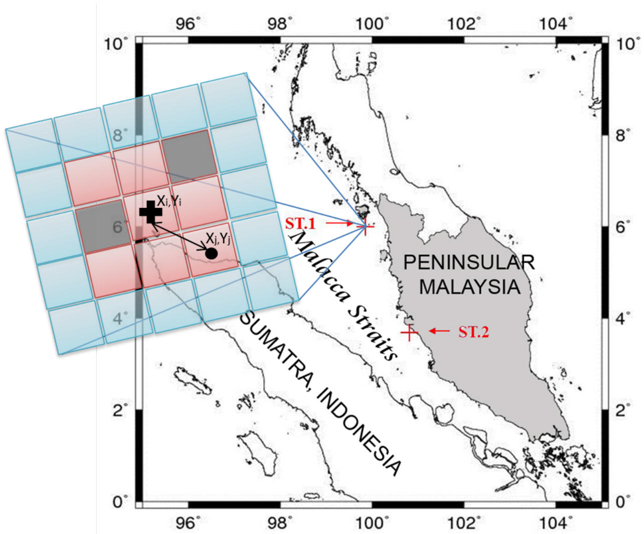



Figure 1. Map of study area and illustration of pixel-to-point based correlation. An example of kernel window that consists of pixels within different kernel size. In-situ point depicted in "plus" sign is always at the centre of the window and the pixels surrounding in "dot" is used to determine the correlation. The black pixel is contaminated by cloud features, the red pixels and blue pixels are the valuable pixel for kernel 3x3 and kernel 5x5 respectively.

The use of different kernel size of 3x3 and 7x7 is also anticipated for assessing the quality of CAL/VAL and the impact of kernel size in this study. Equation (1) and (2) shows the OC3M algorithm and the maximum band ratio (MBR), known as $R_{3M}$, respectively.

$$C_a = 10(c_0 - c_1 R_{3M} + c_2 R_{3M}^2 + c_3 R_{3M}^3 - c_4 R_{3M}^4) \tag{1}$$

$$R_{3M} = \log_{10}(R_{547}^{443} > R_{547}^{488}) \tag{2}$$

where $c_0$, $c_1$, $c_2$ and $c_3$ is the coefficent function at first, second, third and fourth polynomial order respectively, and $R$ is the $R_{rs}$ whose the superscript and the subscript term is respectively refered to the $R_{rs}$ at wavelength in the denominator and the numerator of the band ratio. Some studies suggest that by using a smaller kernel size is reasonable in reducing the geophysical variability impact, however the chances of getting the meaningful result particularly within 3 by 3 pixels is very limited in this study due to heavy cloud cover registered in the pixel (i.e., void pixel in MODIS image). It is a trade-off by increasing the kernel window size should introduce bias in estimating the average of match-up Chl-a as the assigned convolution pixels are not closely collocated at the position of the in-situ station (more than 3 km away from the in-situ station). Therefore, in this paper, we have carried out an application of weight function that compensates the impact of spatial variation in the bigger kernel size so that more reliable and trustworthy match-up samples could be retrieved with low bias of Chl-a estimates.

### 2.3 Application of spatial weight function

Spatial average weight function basically requires the distance between the pixel at the center of kernel window, which is normally collocated at the in-situ station, and the surrounding pixels within the kernel size window. In this study, the spatial distance, *d*, is determined by using the Great Circle Distances formulation that uses two points on the surface of the earth and for this case, the latitude and longitude of the center and surrounding pixels are used respectively. This is easy to be computed by applying distance function in Matlab™. Finally, the average of $R_{rs}$ within different kernel size and at different wavelength, $\lambda$, is calculated by normalizing it with the spatial distance as shown in equation (3).

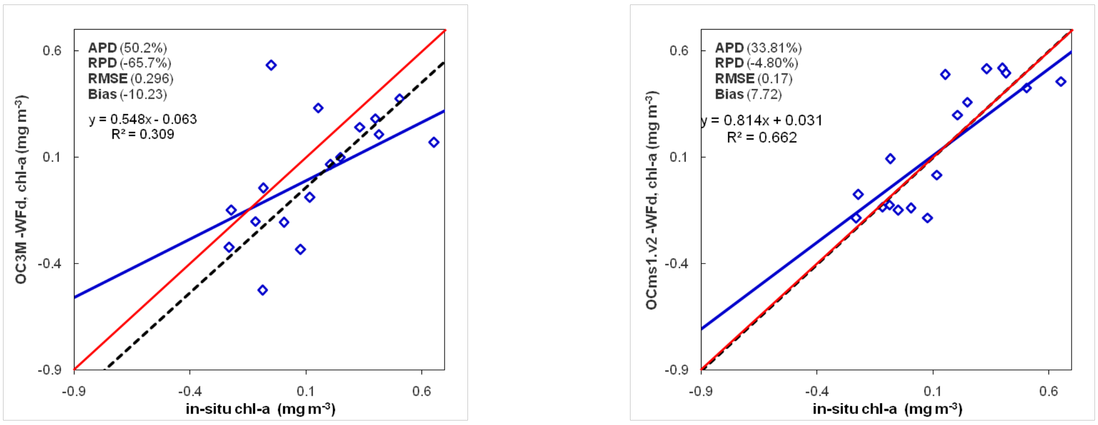
$$\hat{R}rs(\lambda) = \frac{\sum_{i=1}^{N} d^{-1} Rrs(\lambda)_i}{\sum_{i=1}^{N} d^{-1}} \tag{3}$$

Later, the average of $R_{rs}$ with the spatial and temporal weight estimate is applied in the equation (2) to find the spatial bias free Chl-a that allow for utilizing bigger kernel size and thus, increasing the number of match-up samples. Non-linear regression between satellite derived Chl-a and in-situ Chl-a is applied and to help in improving this CAL/VAL exercise - the iterative regression is designed by taking into account the slope and intercept of 1 and 0 respectively as its objective function and finally retrieving the coefficient functions in equation (1) that are best suit for the lowest spatial bias. To assess the quality of CAL/VAL exercise, statistical tools like the absolute percentage difference (APD), the relative percentage difference (RPD), the root mean squares error (RMSE), bias and $R^2$ were determined on the Chl-a retrieved by several modes of OC3M algorithm; (a) global OC3M, (b) OC3M with the weight function (WFd), (c) locally-tuned OC3M (OCms), and (d) OCms with WFd.

## 3. RESULTS AND DISCUSSION

Table 1 tabulates results of coefficient functions and statistics analysis of the Chl-a regression in the different mode of OC3M algorithms and at different kernel size. The OC3M algorithm gives the lowest correlation ($R^2$) with in-situ and this implies that the global OC3M is less reliable for Chl-a retrieval in the Malacca Straits. Although the number of match-up samples is increased from 9 to 18 by applying bigger kernel window (7x7), the correlation remains low, the RMSE is significance and the APD exceeds 35% which is the limit to determine acceptability of the algorithm. The study assumes that the overestimation bias and spatial variability of MODISA Chl-a pixels are the sources of these discripencies. To reduce the impact of spatial variability on the OC3M derived Chl-a, the study applied the weight function on spatial difference and hence significant increase in R2, RMSE and APD are existed. Yet, its bias is higher in both kernel sizes that has most likely been induced by the overestimation effect. With this regard, locally tuned OC3M is used by applying iterative regression between Chl-a OC3M and insitu and this provides a promising estimation of the MODIS derived Chl-a. Figure 2 shows the iterative regression for Chl-a retrieval using weighted OC3M and weighted locally tuned OC3M in 7x7 window. Figure 2(a) shows the Type-II regression model on the weighted OC3M that is hardly to converge with the unity line and this is not a case for locally tuned OC3M, Figure 2(b), that has completely converged at the end of iteration. Besides, locally tuned OC3M produced the acceptable APD of 33.8% and $R^2$ of 0.662.

Larger window size of 7x7 provides more match-up samples than 5x5 window size and this gives advantage of reducing the total bias. However, the impact of window size does not influence so much to the quality of regression and this has proved in the Table 1 in which window size of 3x3 produces better results. The impact of scale effect is more obvious in the case of Chl-a in Malacca Straits as the spatial weight function completely boast up the quality of CAL/VAL by almost of the R2 is more than 0.5, the RMSE is lower than 0.2 and APD is not exceeding 35%. By mapping the locally tuned OC3M Chl-a, shown in Figure 3, clearer indication of Chl-a along the Straits of Malacca is evident and this may give better understanding on the geophysical meaning of Chl-a.

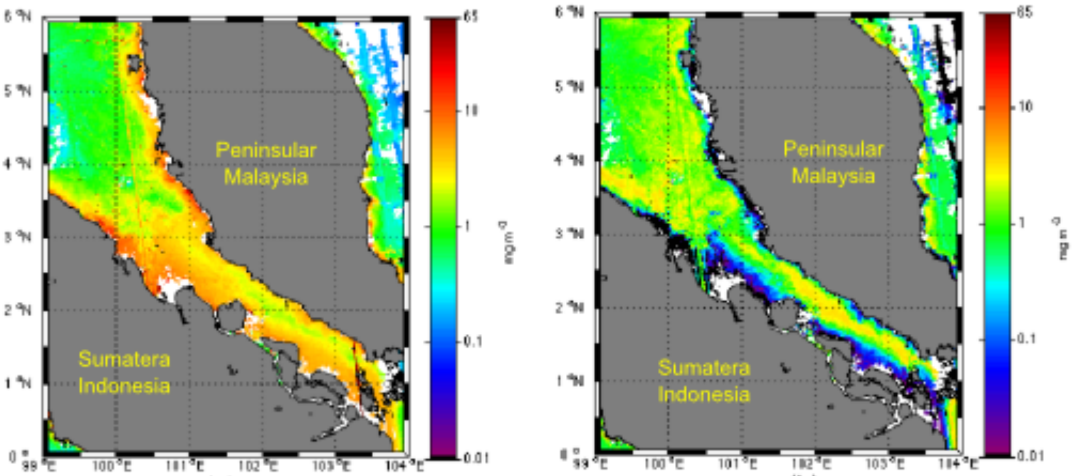



(a) (b)

Figure 2. Plot of Chl-a estimate using the spatial weight average of $R_{rs}$ versus the in-situ Chl-a from 7x7 window extraction. The dotted, blue and red line is the Type-II regression, general regression and unity line, respectively. The R2 is of the general regression line. (a) Chl-a retrieved by OC3M with WFd before local-tuned, and (b) Chl-a with WFd after local-tuned.

Table 1: Statistics of CAL/VAL exercise in different mode of ocean color retrieval and in 3x3 and 7x7 window respectively.

| Kernel size | CAL/VAL type | c0 | c1 | c2 | c3 | c4 | $R^2$ | RMSE | Bias | RPD | APD | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3x3 | OC3M | 0.28 | -2.75 | 1.46 | 0.66 | -1.40 | 0.001 | 0.60 | -61.06 | -177.97 | 61.1 | 9 |
| | OC3M + WFd | 0.28 | -2.75 | 1.46 | 0.66 | -1.40 | 0.62 | 0.19 | -15.39 | -89.20 | 28.5 | 9 |
| | OCms | 0.27 | 100.52 | -1844.58 | 10259.81 | -17441.10 | 0.68 | 0.13 | 4.49 | 1.02 | 26.3 | 9 |
| | OCms + WFd | 0.49 | -0.65 | -43.92 | -54.76 | 1037.47 | 0.56 | 0.02 | 6.67 | -7.57 | 28.8 | 9 |
| 7x7 | OC3M | 0.28 | -2.75 | 1.46 | 0.66 | -1.40 | 0.25 | 0.31 | -6.60 | -62.42 | 53.8 | 18 |
| | OC3M + WFd | 0.28 | -2.75 | 1.46 | 0.66 | -1.40 | 0.31 | 0.30 | -10.23 | -65.678 | 50.2 | 18 |
| | OCms | 0.51 | -0.03 | -38.82 | 143.17 | -136.28 | 0.58 | 0.19 | 9.75 | -5.05 | 35.9 | 18 |
| | OCms + WFd | 0.52 | 0.57 | -57.98 | 256.54 | -315.50 | 0.66 | 0.17 | 7.72 | -4.80 | 33.8 | 18 |




(a) (b)

Figure 3: Map of time-averaged MODIS derived Chl-a from October 2011 to January 2012 (during the North East Monsoon). (a) Chl-a derived from global OC3M and (b) Chl-a derived from locally tuned OC3M with spatial weight function.

## 4. CONCLUSION

The locally-tuned OC3M algorithm with spatial weight function has shown promising improvement in estimating the MODIS derived Chl-a, and it has potential to compensate spatial variability when applying kernel as large as 7x7 to get more match-up samples in the cloud-prone area of Straits Malacca. This method proved the aim of this study to reduce as much as possible the bias, RMSE and APD without deteriorate the quality of correlation between the satellite and in-situ attribute. The study envisages to provide locally tuned OC3M of both ST1 and ST2 acquisitions for better insights in the Chl-a retrieval in the Straits of Malacca. Besides, temporal variability could be taken into consideration in the spatio-temporal weight function. As a conclusion, the spatial weight function on the locally tuned OC3M proved as the alternative of CAL/VAL exercise in Malacca Straits.

## APPENDIX

For the purpose of algorithms' validation, 5 statistical parameters were chosen. These parameters are:

1. correlation coefficient ($R^2$)
2. root mean square error (RMSE)
3. mean relative percentage difference (RPD)
4. mean absolute percentage difference (APD)

The $R^2$ coefficient from the correlation analysis indicates the covariance between in-situ Chl-a ($Chl_{is}$) and algorithms-derived chl-a ($Chl_{ret}$). RMSE indicates the spread of data as compared to the best agreement and was computed as:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(Chl_{ret} - Chl_{is})\right]$$

RPD is the mean percentage difference between $Chl_{ret}$ and $Chl_{is}$ weighted on $Chl_{is}$ values; RPD gives the systematic error or direction of bias, whether it is overestimation or underestimation with respect to the in-situ Chl-a values, it also can be thought as a relative bias;

$$RPD = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Chl_{ret}-Chl_{is}}{Chl_{is}}\right)100$$

APD is slightly different than RPD as it does not give any information about the direction of discrepancy. It gives the uncertainty percentage of the retrieved Chl-a values with respect to in-situ Chl-a value and it represents a sort of relative RMS; it was computed as

$$ADP = \frac{1}{N}\sum_{i=1}^{N}\left(\left|\frac{Chl_{ret}-Chl_{is}}{Chl_{is}}\right|100\right)$$

## ACKNOWLEDGEMENT

Financial support for this study was provided by a grant from the Ministry of Higher Education flagship with Universiti Teknologi Malaysia under project vote number Q.J130000.2527.03H21. The authors wish to thank Dr. Katsuhisa Tanaka from JIRCAS for providing the in-situ data.

## REFERENCES

Chen, J., Yi, C., & Wen, Z. (2013). Systematic Underestimation of MODIS Global Chlorophyll-a Concentration Estimation Algorithm Associating With Scale Effect. *IEEE Sensors Journal*, *13*(5), 1656–1661. doi:10.1109/JSEN.2013.2239638

Darecki, M., & Stramski, D. (2004). An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment*, *89*(3), 326–350. doi:10.1016/j.rse.2003.10.012

Ladner, S. D., Sandidge, J. C., Lyon, P. E., Arnone, R. a., Gould, R. W., Lee, Z. P., & Martinolich, P. M. (2007). Development of finer spatial resolution optical properties from MODIS. *Proceedings of SPIE*, *6680*, 668013–668013–9. doi:10.1117/12.736845

Siswanto, E., Tang, J., Yamaguchi, H., Ahn, Y.-H., Ishizaka, J., Yoo, S., … Kawamura, H. (2011). Empirical ocean-color algorithms to retrieve chlorophyll-a, total suspended matter, and colored dissolved organic matter absorption coefficient in the Yellow and East China Seas. *Journal of Oceanography*. doi:10.1007/s10872-011-0062-z

Suzuki, R., & Ishimaru, T. (1990). An Improved Method for the Determination of Phytoplankton Chlorophyll using N , N-Dimethylformamide *. *Journal of the Oceanographical Society of Japan*, *46*(July), 190–194.

Thia-Eng, C., Gorre, I. R. ., Adrian Ross, S., Bernad, S. R., Gervacio, B., & Corazon Ebarvia, M. (2000). The Malacca Straits. *Marine Pollution Bulletin*, *41*(1-6), 160–178. doi:10.1016/S0025-326X(00)00108-9

Volpe, G., Santoleri, R., Vellucci, V., Ribera d'Alcalà, M., Marullo, S., & D'Ortenzio, F. (2007). The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. *Remote Sensing of Environment*, *107*(4), 625–638. doi:10.1016/j.rse.2006.10.017

Werdell, P. J., Bailey, S. W., Franz, B. a., Harding Jr., L. W., Feldman, G. C., & McClain, C. R. (2009). Regional and seasonal variability of chlorophyll-a in Chesapeake Bay as observed by SeaWiFS and MODIS-Aqua. *Remote Sensing of Environment*, *113*(6), 1319–1330. doi:10.1016/j.rse.2009.02.012