# Classification of Chlorophyll Concentrations in Coastal Water Using Linear Regression with THEOS Imagery

Ahmed Asal Kzar[1], Hayder Saad Abdulbaqi[2], M. Z. MatJafri[#3], H. S. Lim[#4]

[1]*Physics Department, Faculty of Science, Kufa University, Najaf, Iraq.*
[2]*Department of Physics, College of Education, University of Al-Qadisiyah, Iraq.*
[#]*School of Physics, Universiti Sains Malaysia (USM),11800 Penang, Malaysia*

**Abstract:**

The increasing of water pollutants in the coastline has caused environment problem by the widely usage of the region's resources for their life. The waters surrounding areas is one of the most effected from these areas especially by chlorophyll. In this study, using of the linear regression for classification then mapping of chlorophyll concentrations through finding the relationship between the chlorophyll concentration and the digital number of an individual band represented by red or green or blue of the adopted satellite (THEOS) for Penang strait, Penang, Malaysia. The validation data were measured from the study area simultaneously with the satellite image acquisition. The results were by selecting one of the three relationships of the three bands. As result, the red band had the best relationship through given the highest coefficient of determination (0.967) that calculated between the two groups of the chlorophyll concentrations: measured (validation data) and calculated by the linear regression. In addition, we tested the new algorithm with noise, it has proved its steadfastness with noisy images over a range of noise levels. Therefore, chlorophyll mapping of water pollution in the adopted study area can be achieved by the linear regression with the based THEOS satellite image.

**Keywords:**

Chlorophyll; environment; THEOS; image classification; water quality mapping.

## I. Introduction

There is an influence by Marine waters in general and coastal areas in especial on global environment via different pollutants. The main two causes are human activities and natural influences. Therefore, the water pollution has attracted the potential of the scientists, researchers, and decision makers in several field in this world, because the problem is directly and indirectly related with the sustainability development studies. The techniques of remote sensing are widely used for different applications especially in environment application. Remote sensing is useful, effective, and advanced technique for water quality mapping.

Calculation of pollutants concentrations of water quality by regression algorithm technique, has been applied by many researchers [1-3]. A multivariate model for coastal water quality mapping by using satellite images has been applied by [4]. Monitoring of surface water quality in coastal area of Penang [5]. Mapping spatial distribution of total suspended solids (TSS) by using a statistical model depending THEOS satellite imagery over Penang island, Malaysia [6]. A feed-forward Hopfield neural network algorithm (FHNNA) with a color satellite image for water quality mapping [7]. A method of Detection the dynamic linkage between landscape characteristics and water quality in a subtropical coastal watershed of the southeast of china [8]. Water quality mapping using of Hopfield neural network algorithm (MHNNA) with ALOS imagery in the west coastal waters of Langkawi Island, Malaysia [9]. Development and application of a remote sensing-based Chlorophyll-a concentration prediction model for complex coastal waters of Hong Kong [10]. A comparison of multiple linear regression models and machine learning using remote sensing variables for high resolution mapping of soil properties in south-Western Burkina Faso . Optical classification of chlorophyll concentration and suspended sediment in the Coastal Waters of the Northern Indian Ocean [12]. Using landsat-8 and sentinel-2a satellite for chlorophyll estimation of lake water and coastal water [13]. Using MERIS data for A soft-classification-based chlorophyll-a estimation method in the highly turbid and eutrophic Taihu Lake [14]. Introduced an optical hybrid Chlorophyll-a Algorithm for Assessing Trophic States of a Tropical Brazilian Reservoir depend on MSI/Sentinel-2 Data [15]. Using of Linear Regression Model and Remote Sensing techniques for Prediction of Soil Salinity [16].

In this study, we adopted linear regression for finding the linear relationship between the chlorophyll concentration and digital number (DN). Linear regression method considered easier, faster, and more efficient than a lot of calculation methods. The finding of the relationship for the three bands: red, green, and blue, then selecting of the relationship (best band) of the adopted satellite image, where this relationship cannot be found in a lot of satellite images but in this study we exploit this type of satellite image which is THEOS (Thailand Earth Observation System), it gives us the required relationship for application of water quality to give reliable results.

## II. Study Area

The investigated area in this study is the Straits of Penang Island. It is northwest part of Malaysia, it located in longitudes 100° 09' E to 100° 20 E and latitudes 5° 12'N to 5° 30' N and zone 85. where Penang is the second contaminated city in this country, with an estimated population more than 720,000. Penang Island is approximately 295 sq. km. Penang Straits is part of Indian Ocean and it is between Penang Island and the mainland Peninsular Malaysia. Fig. 1 gives the study area.

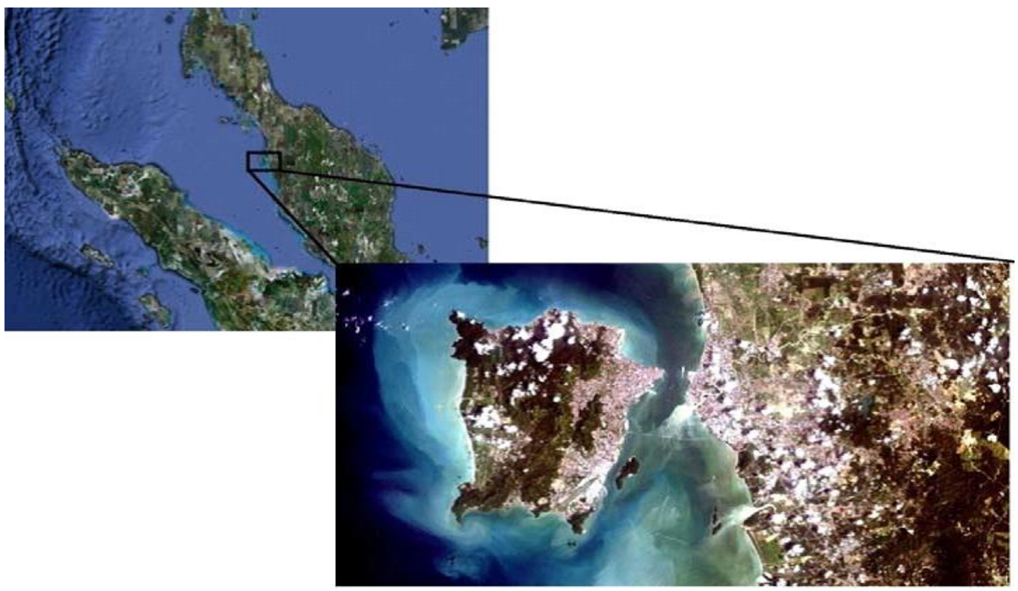

**Fig. 1**: The study area.

## III. The Satellite Bands and Data Collection

The collected samples from adopted study area are 13 samples of different polluted water with the concentrations levels of chlorophyll, these samples collected simultaneously with the acquiring of THEOS satellite image on 12 November 2009. The THEOS bands and their wavelengths as spectral characteristics demonstrated in Table 1.

**Table 1:** The spectral characteristics of the used satellite.

| Band | Wavelength |
|---|---|
| Band1 (blue) | 0.45-0.52 µm |
| Band 2 (green) | 0.53-0.60 µm |
| Band 3 (red) | 0.62-0.69 µm |
| Band 4 (near infrared) | 0.77-0.90 µm |

The used bands here in this work are only the visible band (red, green, and blue). The spatial resolution of the adopted THEOS image is 15 meters. Fig. 2 shows the raw satellite image (THEOS) and the distributed collected samples of validation. Where The values of collected samples and their positions are in the Table 2.

**Fig. 2.** The raw satellite image and the distribution of collected samples (validation).

**Table 2:** The Values of collected samples and their positions.

| *Long* | *Lat* | *Tss(mg/l)* |
|---|---|---|
| 100.3366028 | 5.398308333 | 8.5 |
| 100.3303611 | 5.390994444 | 3.6 |
| 100.325475 | 5.385580556 | 3.3 |
| 100.3189667 | 5.379897222 | 3.1 |
| 100.329225 | 5.367391667 | 2.5 |
| 100.3438389 | 5.364375 | 4.3 |
| 100.3525028 | 5.365713889 | 5.2 |
| 100.356575 | 5.370316667 | 9.2 |
| 100.3611861 | 5.375191667 | 10.2 |
| 100.3633667 | 5.380883333 | 12.3 |
| 100.3587694 | 5.383336111 | 11.3 |
| 100.3614917 | 5.390111111 | 4.9 |
| 100.3436194 | 5.388794444 | 2.8 |

## IV. Methodology

There are many methods have been used to calculate the pollutants concentrations with the remote sensing techniques. here in this work we use linear regression because it is useful, fast, simple, and common model, it has been used in several applications in scientific studies [17-20]. Also, THEOS satellite sensor in this application has linear relationship between its digital numbers of the individual visible band and the chlorophyll parameter, linear regression model used to represent the main technique to get best results for water quality mapping. The investigated relationship between the important value which is represented by the chlorophyll concentration value with the digital number of a visible band of THEOS image bands. As result, we will use the band that gives the best result represented by the highest coefficient of determination ($R^2$). Where the linear regression does not always adopted with THEOS image for water quality mapping [6]. Also, it does not always adopted with other satellites [21], but a relationship with high coefficient of determination such as that has been exploited in this work yes.

## V. Work Algorithm

The used algorithm of our work is represented by the following steps:

**Step 1-** Adopt validation data (in-situ data) through collecting samples from the study area simultaneously with the satellite image acquisition, fixing samples positions by a handheld global positioning system (GPS).

**Step 2-** Find the pollutant concentration values (chlorophyll) of the measured samples in the lab.

**Step 3-** Ignore the land region by giving its pixels zeros value (black) for focusing on water area only in the satellite image..

**Step 4-** Extract the digital numbers of the three bands values (DN) that are: red, green, and blue from the satellite image pixels that have the same positions of collected samples and set group (13 values) for each band.

**Step 5-** Find the relationship of linear regression between chlorophyll concentration values and red band values (DN), where this relationship evaluated by coefficient of determination ($R^2$).

**Step 6-** Repeat (Step 5) for green and blue bands.

**Step 7-** Pick the best relationship of the three bands relationships, which has the highest ($R^2$).

**Step 8-** Apply the best relationship to find chlorophyll concentration values in each pixel of water pixels in the satellite image through extracting the (DN) of the best band from the pixel.

**Step 9**- Classify the chlorophyll concentration values by grouping them through giving a class for each group of values.

**Step 10**- Draw map for chlorophyll concentration values by color coded that gives color for each class.

## VI. Testing Algorithm with Noise

After success of linear regression algorithm for classification water pollutants, we prove the steadfastness of this algorithm through test it with the biggest problem which faces the satellite imagery, it is noise; noise is the undesirable signal mixed with the real image signal, it leads to reducing the image information in the homogenous and the edge regions of the image [21], so it affects the accuracy of the applied process. Gaussian noise is the adopted noise type in this test. This type of noise has been adopted for test many remote sensing images [22-24]; Where the noise was added to the adopted satellite image and then applying the new algorithm for testing it. Then, this procedure repeated to extra quantities of Gaussian noise. The number of noise allocations that are added to the used image is 10 allocations, from 0.001 to 0.01 of variance in the Gaussian equation, and the interval was via 0.001 at each time. After each time of addition of noise, we apply the new algorithm and measure the accuracy through the coefficient of determination ($R^2$) of each relationship of the three bands in the satellite image.

## VII. Results and Discussion

The results obtained by applying the proposed algorithm to the adopted satellite image are demonstrated in models, tables and figures. Where the main and final result represented in the map of the chlorophyll concentrations. the models of the three relationships between chlorophyll and each band are as the followings:

***Chlorophyll*** = *0.0943* ***R*** *- 2.4741* ………(1)

***Chlorophyll*** = *0.181**G** - 9.3187* ………(2)

***Chlorophyll*** = *0.2122* ***B****- 21.876* ……… (3)

Where ***Chlorophyll*** is the concentration of chlorophyll in (mg/m$^3$) unit. ***R,G,*** *and* ***B*** the three bands red, green, and blue respectively. The accuracy that has been given by these relationships was by the coefficient of determination ($R^2$) between Chlorophyll and the digital number of the band as in the Table 3.

**Table 3:** The accuracy between chlorophyll and the digital number of each band.

| *Band* | $R^2$ |
|---|---|
| *Red* | 0.967 |
| *Green* | 0.944 |
| *Blue* | 0.904 |

Clear from the accuracy table that red band gives the best results and we use it to calculate the chlorophyll concentrations for drawing a map of chlorophyll of the study area.

Table 4 and Fig. 3 represent the measured chlorophyll concentration values in the validation data and the calculated concentration values by using the best model (red band model) that gave high accuracy.

**Table 4:** The accuracy between chlorophyll and the digital number of red band model.

| *Sample No.* | *Measured Chlorophyll (mg/m$^3$)* | *Calculated Chlorophyll (mg/m$^3$)* |
|---|---|---|
| 1 | 8.5 | 7.43 |
| 2 | 3.6 | 3.84 |
| 3 | 3.3 | 3.37 |
| 4 | 3.1 | 3.37 |
| 5 | 2.5 | 2.43 |
| 6 | 4.3 | 4.22 |
| 7 | 5.2 | 6.39 |
| 8 | 9.2 | 8.65 |
| 9 | 10.2 | 10.16 |
| 10 | 12.3 | 13.27 |
| 11 | 11.3 | 10.35 |
| 12 | 4.9 | 5.16 |
| 13 | 2.8 | 2.52 |

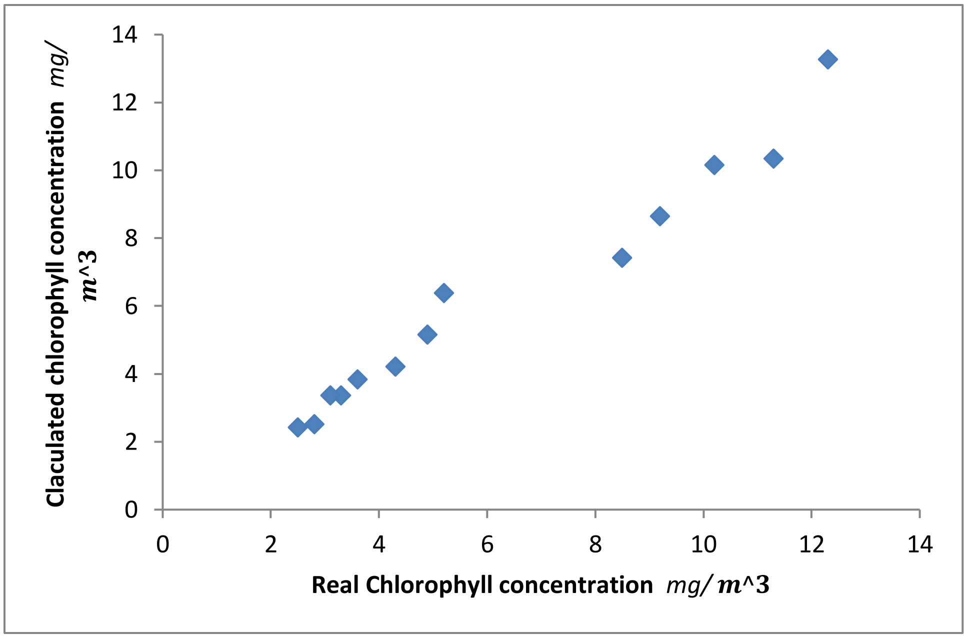


**Fig. 3:**The high correlation between the measured and calculated of chlorophyll by using red band model. .

The most important result in this work is the produced map of chlorophyll pollutant by the new algorithm of using linear regression, where the result as we mention with the red band model of the adopted satellite image THEOS. Fig. 4 demonstrates chlorophyll map in Penang straight, Penang, Malaysia.

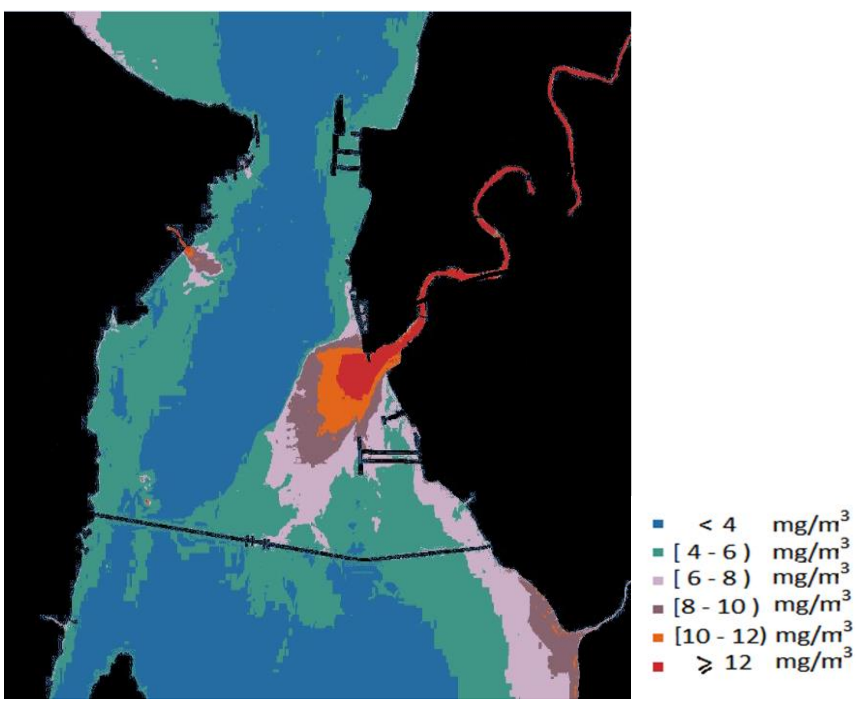


**Fig. 4.** Chlorophyll map.

The produced chlorophyll map a good demonstration of this pollutant concentrations in the investigated study area, this map gives a fact that this region is full with this pollutant in different levels of its concentrations beginning from the most dangerous concentration that represented by red color until the least concentration that represented by blue color, where this sequence denotes on the mitigation of this pollutant through mixing it with the unpolluted water, this mitigation started from the estuary and be reduced grade by grade until reaching the least concentration. These grade has been color coded (classes) after giving the range of each class [24]. Fig. 5 gives the sizes in $km^2$ of chlorophyll classes in investigated study area.

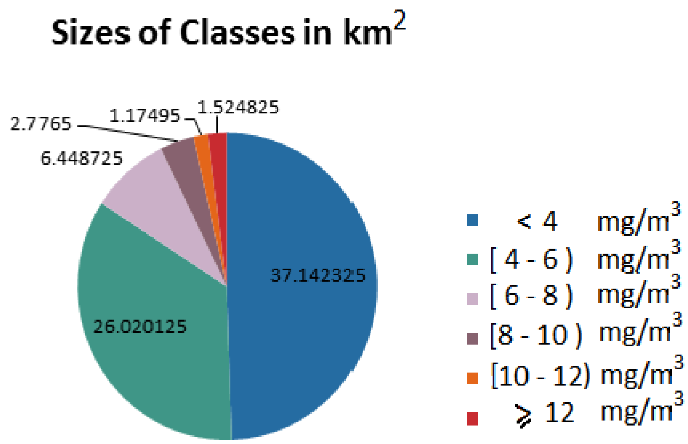


**Fig. 5.** Sizes of chlorophyll classes in $km^2$ of investigated study area.

**Fig. 6** illustrates the results of noise test of the new algorithm of linear regression for the three bands models.

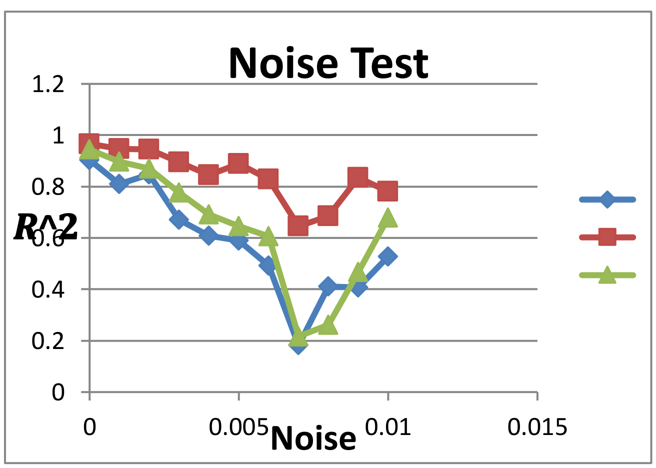


**Fig. 6.** Test of linear regression method with noise for the three bands models.

Fig. 6 gives the steadfastness of the new algorithm against noise which is the biggest problem faces the images in generally and the remote sensing images in particularly, but the red band proves its success comparing with other bands (green and blue). However,
this means the new algorithm of linear regression can be used with amount of noise with the adopted THEOS satellite image.

## VIII. Conclusions

The algorithm of the linear regression approved it success for water quality mapping. Through calculating and classify chlorophyll concentrations Where the red band model exhibited a higher coefficient of determination (0.967). Also, the testing noise of the now algorithm for the three bands, we found the red band proved its steadfastness with noisy images over a range of noise levels, comparing with the green and blue bands. In addition to, the resulted map gives dilution of chlorophyll concentration in water through the finding differences between the digital numbers of the band in water pixels of the used THEOS image. This study provides a brief overview of chlorophyll mapping at the Penang Straight, Penang, Malaysia. The imagery by the remote sensing can be used to provide information for effective planning management. Via linear regression, we get extra information about the pollutant concentrations related with areas of the chlorophyll concentrations values. Applying the linear regression to THEOS data for chlorophyll mapping in the study area provided reliable and high quality results. Our results showed successful evidence of applying the new algorithm to the water quality mapping of color satellite image. Concerning the environmental health, Penang strait poses a bad case in water quality.

## IX. Acknowledgments

This study was supported by the University Sains Malaysia. We would like to express our thanks and appreciation the technical staff who contributed in this study. Special thanks to the Geo-Informatics and Space Technology Development Agency (GISTDA) that provided us the THEOS data.

## X. References

[1] Dekker, A.G., Vos, R., and Peters, S.: '*Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data'*, International journal of remote sensing, 2002, 23, (1), pp. 15-35

[2] Doxaran, D., Froidefond, J.-M., Lavender, S., and Castaing, P.: '*Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations',* Remote sensing of Environment, 2002, 81, (1), pp. 149-161

[3] Tassan, S.: '*An improved in-water algorithm for the determination of chlorophyll and suspended sediment concentration from Thematic Mapper data in coastal waters*', International Journal of Remote Sensing, 1993, 14, (6), pp. 1221-1229

[4] Su, Y.-F., Liou, J.-J., Hou, J.-C., Hung, W.-C., Hsu, S.-M., Lien, Y.-T., Su, M.-D., Cheng, K.-S., and Wang, Y.-F.: '*A multivariate model for coastal water quality mapping using satellite remote sensing images*', Sensors, 2008, 8, (10), pp. 6321-6339

[5] Syahreza, S., San, L.H., Jafri, M.Z.M., and Mustapha, M.R.: *'Monitoring surface water quality in coastal area of Penang*', International Proceedings of Chemical, Biological and Environmental Engineering, 2011, 8, pp. 297-300

[6] Asadpour, R., San, L.H., Alashloo, M.M., and Moussavi, S.Y.: '*A statistical model for mapping spatial distribution of total suspended solid from THEOS satellite imagery over Penang Island, Malaysia*', Journal of Applied Sciences Research, 2012, 8, (1), pp. 271-276

[7] Kzar, A.A., Jafri, M.M., San, L.H., Al-Zuky, A.A., Mutter, K.N., and Al-Saleh, A.H.: '*A feed-forward Hopfield neural network algorithm (FHNNA) with a colour satellite image for water quality mapping', in Editor (Ed.)^(Eds.): 'Book A feed-forward Hopfield neural network algorithm (FHNNA) with a colour satellite image for water quality mapping*' (IOP Publishing, 2016, edn.), pp. 012075

[8] Huang, J., Li, Q., Pontius, R.G., Klemas, V., and Hong, H.: '*Detecting the dynamic linkage between landscape characteristics and water quality in a subtropical coastal watershed, Southeast China*', Environmental management, 2013, 51, (1), pp. 32-44

[9] Kzar, A.A., Jafri, M., Zubir, M., Mutter, K.N., and Syahreza, S.: '*A Modified Hopfield Neural Network Algorithm (MHNNA) Using ALOS Image for Water Quality Mapping*', International journal of environmental research and public health, 2016, 13, (1), pp. 92

[10] Nazeer, M., and Nichol, J.E.: '*Development and application of a remote sensing-based Chlorophyll-a concentration prediction model for complex coastal waters of Hong Kong*', Journal of Hydrology, 2016, 532, pp. 80-89

[11] Forkuor, G., Hounkpatin, O.K., Welp, G., and Thiel, M.: '*High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models',* PloS one, 2017, 12, (1)

[12] Monolisha, S., Platt, T., Sathyendranath, S., Jayasankar, J., George, G., and Jackson, T.: 'Optical classification of the coastal waters of the Northern Indian Ocean', Frontiers in Marine Science, 2018, 5, pp. 87

[13] Yadav, S., Yamashiki, Y., Susaki, J., Yamashita, Y., and Ishikawa, K.: '*CHLOROPHYLL ESTIMATION OF LAKE WATER AND COASTAL WATER USING LANDSAT-8 AND SENTINEL-2A SATELLITE*', International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2019

[14] Zhang, F., Li, J., Shen, Q., Zhang, B., Tian, L., Ye, H., Wang, S., and Lu, Z.: '*A soft-classification-based chlorophyll-a estimation method using MERIS data in the highly turbid and eutrophic Taihu Lake*', International Journal of Applied Earth Observation and Geoinformation, 2019, 74, pp. 138-149

[15] Cairo, C., Barbosa, C., Lobo, F., Novo, E., Carlos, F., Maciel, D., Flores Júnior, R., Silva, E., and Curtarelli, V.: '*Hybrid Chlorophyll-a Algorithm for Assessing Trophic States of a Tropical Brazilian Reservoir Based on*

*MSI/Sentinel-2 Data'*, Remote Sensing, 2020, 12, (1), pp. 40
[16] Hihi, S., Rabah, Z.B., Bouaziz, M., Chtourou, M.Y., and Bouaziz, S.: *'Prediction of Soil Salinity Using Remote Sensing Tools and Linear Regression Model',* Advances in Remote Sensing, 2019, 8, (3), pp. 77-88
[17] Rodbard, D., and Frazier, G.: '[1] Statistical analysis of radioligand assay data': 'Methods in enzymology' (Elsevier, 1975), pp. 3-22
[18] Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W.: '*Applied linear statistical models',* 1996
[19] Shabalin, A.A.: '*Matrix eQTL: ultra fast eQTL analysis via large matrix operations*', Bioinformatics, 2012, 28, (10), pp. 1353-1358
[20] van Erp, T.G., Hibar, D.P., Rasmussen, J.M., Glahn, D.C., Pearlson, G.D., Andreassen, O.A., Agartz, I., Westlye, L.T., Haukvik, U.K., and Dale, A.M.: '*Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium*', Molecular psychiatry, 2016, 21, (4), pp. 547-553
[21] Al-Zuky, A.A.D., and Kzar, A.A.: '*Hit noise reduction in some x-ray images*', Al-Nahrain Journal of Science, 2009, 12, (4), pp. 78-86
[22] Saleh Al-amri, S., Kalyankar, N., and Khamitkar, S.: '*A Comparative Study of Removal Noise from Remote Sensing Image'*, arXiv preprint arXiv:1002.1148, 2010
[23] Lee, J., and Hoppel, K.: '*Noise modeling and estimation of remotely-sensed images', in Editor (Ed.)^(Eds.): 'Book Noise modeling and estimation of remotely-sensed images*' (IEEE, 1989, edn.), pp. 1005-1008
[24] Zhu, X.X., and Bamler, R.: *'Tomographic SAR inversion by $ L_ {1} $-norm regularization—The compressive sensing approach*', IEEE transactions on Geoscience and Remote Sensing, 2010, 48, (10), pp. 3839-3846