Utilising Landsat-8 OLI in determining Soil pH using single and combination band method for paddy fields area in Malaysia.

Pauziyah Mohammad Salim¹, Noramirah Samsuddin¹, Ummi Atiqah Alias¹, Muhammad Zamir Abdul Rasid², Hasliana Kamaruddin²

¹Faculty of Architecture Planning and Surveying, UiTM Shah Alam, Selangor Malaysia. ²Malaysia Agricultural Research and Development Institute (MARDI), 43400 Serdang, Selangor, Malaysia.

E-mail: pauzi732@uitm.edu.my

Abstract. The pH of a soil is a measure of its acidity or alkalinity. The pH of the soil is important in agricultural activities because it has an impact on crop yield. Remote sensing, Geographic Information Systems (GIS), and digital soil maps are becoming more appreciated in soil science studies. The research thus focuses on creating a predicted soil pH map of the study area using Landsat-8 satellite imagery and GIS. The single and combination of spectral bands were used to generate three models using simple linear regression. The results suggested that the approach is not sensitive enough for prediction of soil pH in the study area. R² value obtained are 0.049 from Model 1, 0.016 from Model 3, and 0.0003 for Model 2. All the models indicate that the soil pH is in the acid situation but the full range of observed pH is not matched by any predicted model. In the validation process, Model 1 has an RMSE value of 0.397, whereas both Model 2 and 3 have RMSE value of 0.405. To obtain a more promising pH result, it is suggested to use indices such as vegetable indices (VI), salinity index (SI), and a combination of band ratios.

1. Introduction

In agriculture activities, the pH of the soil is critical. Because it influences a range of chemical and biological processes, soil pH is a control variable in soils. It is a determination of the acidity or alkalinity of a soil. Soil salinity, alkalinity, and acidity all have an impact on soil and crop production. [1] It also reduces agriculture yield while limiting pastureland grass species variety. [2] Soil pH is useful to analyse in agriculture since it is an indication of soil acidification and overfertilization concerns. It regulates the amount and quantity of soil minerals that plants require to thrive. [3] Farmers can detect soil fertility in its initial stages and intervene locally by monitoring the surface dynamics of soil pH.

The most common technique of assessing soil pH in Malaysia is by ground sample. This method has shown to be very efficient, but it is also very costly in terms of manual labour and the lack of spatial coverage at any given time. Remote sensing technology is one of the modern approaches of analysing and interpreting natural environment, including soil, as it assess the electromagnetic energy falling from
the sun on the targets and communicates with them as well to absorb, transmit, reflect or emit. [4] Many studies have been carried out using remote sensing techniques to identify the soil pH condition and track patterns in soil pH changes. [5] This technique has also aided in the development of environmentally friendly agriculture and food production. [6]

2. Problem Statement
Malaysia’s climate is hot and humid all year, with an average global temperature of 27°C (80.6°F) and little seasonal changes. This region has tropical soil, most of which is considered as a problem soil, such as peat, sandy soil, acid sulphate soil, and highly weathered soil which need special management strategies to optimise yields usually enhanced by applying some amendments.

In acid sulphate soil, adding compost as a soil amendment can raise soil pH and improve rice productive efficiency. [7] While liming an acidic soil enables for better soil absorption of nutrients by the plant throughout its roots. [8]

Remotely sensed data, Geographic Information Systems (GIS), and digital soil maps are becoming more appreciated in soil science studies. It has been demonstrated that remote sensing is a cost-effective, instant, and repeatable method of providing quantitative and spatially distributed data on soil characteristics. (Irons et al., 1989) With these purposes in mind, this study uses remotely sensed Landsat-8 OLI/TIRS data and GIS techniques to estimate soil pH values for paddy fields in Malaysia.

3. Materials and Methods

3.1 Study Area
The chosen site for the study is the South Seberang Perai District in Penang, Malaysia. Pulau Pinang, as well recognised as Penang state (Figure 1), is one of Malaysia’s 13 states, area of West Malaysia at an estimated coordinate of 5.4141° N, 100.3288° E. It has a capacity of 162,905 people and occupies an area of 242 square kilometres.

South Seberang Perai, also referred as Seberang Perai Selatan (SPS), is a granary region that produces agricultural products. The paddy-growing region has made a significant contribution to the national rice supply. (IADA Pulau Pinang, 2021)
3.2 Methodology
This section outlines the recommended structure for achieving the study’s aims and outcomes. Figure 2 depicts the proposed research technique.

![Diagram of Methodology and Ground Sampling Data]

**Figure 2.** Flow of Methodology and the Ground Sampling Data.

The imagery was obtained from the USGS websites of Landsat-8 collection. Surface reflectance values, as measured at ground surface in the absence of atmospheric effects, are contained in the Level-2 data, which is the more latest Landsat Level-2 data products. The data is publicly available on the website (https://earthexplorer.usgs.gov/). The dataset is also adjusted to its Top of Atmosphere (ToA) correction. Malaysia having a tropical climate, obtaining a clear cloud cover imagery is very difficult. As the sampling collection is being done in stages throughout the year 2017, it is also difficult for us to select date that corresponding the sampling date thus the selection of the date used is based on free cloud cover criteria.

3.2.1 Ground soil pH data.
The topsoil level of the ground soil sample was taken under dry conditions (0-20cm). GPS equipment were used to identify the position (x,y) of each soil samples collected. The pH of all samples was then determined using Meter pH at a soil chemical laboratory. Seberang Perai Selatan (SPS) has 355 sampling points shown in Figure 2. For model validation purposes, 40 sampling points representing 10% of the ground soil sample data have been set aside.

3.2.2 Soil pH prediction models
The ground soil pH values are individually correlated with every band from the satellite's visible band (VIS) 2, 3, 4, Near Infra-Red (NIR) band 5, Shortwave Infra-Red (SWIR) band 6 and 7 surface reflectance using Excel spreadsheet and Spatial Analyst Tools in ArcMap 10.8 software as shown in Figure 3. With adaptation from [9], single and combination band method is used based on the coefficient of correlation R values from each band. Six scatter plots were made with satellite’s spectral band for the Y-axis and Ground Soil pH as an X-axis. According to the plots Figure 3, all the VIS, NIR and SWIR have a positive correlation but insignificant correlation coefficient R. The result is used for the soil pH prediction using linear and multiple regression.
Based in each band's $R^2$ values they are classified into three categories either strong, moderate, or weak. Table 1 displays the band values along with their category. Band 2 is the strongest band, the combination of band 3 and 4 is the moderate, while the band 5, 6 and 7 is the weakest. The finding of band 2 of having the best correlation agreement with research done by [10]. The soil pH prediction model based on regression is shown in Table 2.

![Figure 3. Correlation of Landsat-8 Bands against Ground pH.](image-url)
With the hypothesis, the higher R-value gives a better prediction of soil pH in term of distribution and accuracy. So, with lower R-value, the capability of the model to predict soil pH is less accurate. The accuracy of the predicted soil map is calculated by using the RMSE. The lower RMSE value should indicate the best model for soil pH prediction. Maps of soil pH prediction are produced by applying the models in the Raster Calculator in Spatial Analyst Tools using ArcMap 10.8 software. Validation process is done to determine the accuracy of the map produced by calculating the regression $R^2$ and RMSE.

| Bands     | $R^2$  | Model category |
|-----------|--------|----------------|
| B2 – Blue | 0.0095 | Strong         |
| B3 – Green| 0.0016 | Moderate       |
| B4 – Red  | 0.0032 | Moderate       |
| B5 – NIR  | 0.0001 | Weak           |
| B6 – SWIR | 0.0010 | Weak           |
| B7 – SWIR | 0.0012 | Weak           |

Table 2. Linear regression models for predicting soil pH parameters from satellite bands reflectance.

| Model | Model | $R^2$ |
|-------|-------|-------|
| 1     | $Y = 8.099 + 20.719 \times B2$ | 0.10  |
| 2     | $Y = 5.546 - 36.108 \times B3 + 30.8548 \times B4$ | 0.08  |
| 3     | $Y = 5.879 - 3.268 \times B5 + 10.567 \times B6 - 7.951 \times B7$ | 0.04  |

4. Results and Discussion

4.1 Predicted soil pH Map
The predicted soil pH map is shown in the Figure 4. The map (a) is predicted soil pH calculated using Model 1. The pH range calculated is from 5.9 – 6.1. The map (b) is calculated by using model 2, with the pH range from 5.9 – 6.3 and the map (c) is produced by using model 3, range from 5.9 – 6.0. While the map (d) is the ground soil pH map, range from 4.4 – 7.7.

All the models indicate that the soil pH is in the acid situation, where the mean for model 1 is 5.957, model 2 is 5.964, model 3 is 5.946 and 5.948 for ground soil pH map. As for the soil pH standard deviation, model 1 result is 0.066, model 2 is 0.078, model 3 is 0.027 and 0.624 for the ground soil pH. The descriptive statistics from the model and the ground soil map is presented in Table 3.

| Model  | Model 2 | Model 3 | Ground Data |
|--------|---------|---------|-------------|
| pH Range | 5.9 – 6.1 | 5.9 – 6.3 | 5.9 – 6.0 | 4.4 – 7.7 |
| pH Mean | 5.957 | 5.964 | 5.946 | 5.948 |
| pH Std Dev | 0.066 | 0.078 | 0.027 | 0.624 |
| % Acidic | 100% | 100% | 100% | 93.7% |
| % Neutral | 0% | 0% | 0% | 1.6% |
| % Alkaline | 0% | 0% | 0% | 4.8% |
Figure 4. Predicted Soil pH Map using simple linear regression models (a, b, c) with Landsat-8 in comparison with (d) interpolated map using ground observations.

All of the predicted soil pH map provided a similar distribution of soil pH. From the ground sample points, it is indicated that more than a half area of paddy field is located in acid situation and all the models used in this study indicated similar result. It can be seen from the map that the model allows spectral bands to be used as proxies to soil pH that is of acidic class but not neutral and alkaline. All the model also manages to accurately map areas nearby the road as almost neutral.

4.2 Validation

Figure 5 shows the regression $R^2$ value obtained for Model 1 0.049, Model 2 is 0.0003, and Model 3 is 0.016. The RMSE values also show the models predicted soil pH values did not defer greatly from the observed values but the full range of observed pH is not matched by any predicted model. In the
validation process, Model 1 has an RMSE value of 0.397, whereas both Model 2 and 3 have RMSE value of 0.405. Based on R2 and RMSE, Model 1 is considered the best model for predicting soil pH in the area.

![Scatter plots](image)

**Figure 5.** Scatter plot between measured and predicted soil pH.

5. Conclusion
This study focuses on the use of Landsat-8 imagery in order to produce a predicted soil pH map of the study area. A linear regression method using single and combination of spectral bands were used to determine the soil pH. It is found that the methods used are not sensitive enough to predict soil pH. The surface reflectance in VIS, NIR and SWIR bands of Landsat-8 showed no significant correlation with the soil pH of the study area. Perhaps by using indices such as vegetable indices (VI) such as NDVI, salinity index (SI) as well as combination of band ratio will give more promising result.

6. Acknowledgments
The authors would like to NASA providing free Landsat-8 data and Malaysian Agricultural Research and Development Institute (MARDI) for the ground data availability.
7. References

[1] Oshunsanya, “10. Introductory Chapter_ Relevance of Soil pH to Agriculture _ IntechOpen,” 2019.

[2] L. J. Wang Ren-zhong, “[3] Dynamic Population Models of the Ecological Dominance During the Retrogressive Succession of Leymus chinensis Grassland.pdf.” pp. 170–174, 1995, [Online]. Available: https://www.plant-ecology.com/EN/abstract/abstract3979.shtml.

[3] M. F. Ghazali, K. Wikantika, A. B. Harto, and A. Kondoh, “Generating soil salinity, soil moisture, soil pH from satellite imagery and its analysis,” Inf. Process. Agric., vol. 7, no. 2, pp. 294–306, Jun. 2020, doi: 10.1016/j.inpa.2019.08.003.

[4] M. A. Kadhim, “Predict Some Soil Characteristics And Prepare Their Digital Maps From Remote Sensing Data,” 2020.

[5] Y. Zhang, B. Sui, H. Shen, and Z. Wang, “6. Estimating temporal changes in soil pH in the black soil region of Northeast China using remote sensing,” Comput. Electron. Agric., vol. 154, pp. 204–212, Nov. 2018, doi: 10.1016/j.compag.2018.09.005.

[6] L. Bai, C. Wang, S. Zang, Y. Zhang, Q. Hao, and Y. Wu, “3. Remote sensing of soil alkalinity and salinity in the Wuyu’er-Shuangyang river basin, Northeast China,” Remote Sens., vol. 8, no. 2, 2016, doi: 10.3390/rs8020163.

[7] N. S. adah Abdul Halim, R. Abdullah, S. A. Karsani, N. Osman, Q. A. Panhwar, and C. F. Ishak, “9. Influence of soil amendments on the growth and yield of rice in acidic soil,” Agronomy, vol. 8, no. 9, pp. 1–11, 2018, doi: 10.3390/agronomy8090165.

[8] H. Kamaruddin, M. Z. A.R., N. A.R, and Theeba.M, “8. Development of Soil pH Map and Lime Recommendation for Rice Cultivation Area In IADA Pulau Pinang Using GIS Application,” pp. 1–5, 2019, [Online]. Available: https://www.researchgate.net/publication/342976621_Development_of_Soil_pH_Map_and_Lime_Recommendation_for_Rice_Cultivation_Area_In_IADA_Pulau_Pinang_Using_GIS_Application.

[9] M. F. Ghazali, A. B. Harto, and K. Wikantika, “2. The Simple Method to Assess Land Quality of Paddy Field Using Spectral, Soil pH and Statistical Regression Technique (Case Study of Paddy Field in Majalaya Subdistrict, Bandung Region),” KnE Life Sci., vol. 2, no. 6, p. 194, Nov. 2017, doi: 10.18502/kl6.v2i6.1040.

[10] P. Gogumalla et al., “Detecting Soil pH from Open-Source Remote Sensing Data: A Case Study of Angul and Balangir Districts, Odisha State,” J. Indian Soc. Remote Sens., vol. 9, 2022, doi: 10.1007/s12524-022-01524-9.