DLSU RESEARCH CONGRESS 2023
MANILA, PHILIPPINES
JULY 5-7, 2023

Fostering a Humane and Green Future:
Pathways to Inclusive Societies and Sustainable Development

Estimation of Metro Manila Ground Elevation using Machine Learning Techniques

Joenel Galupino[1,*] and Jonathan Dungca[1]
[1] *Department of Civil Engineering, De La Salle University, Manila, Philippines*
**Corresponding Author: joenel.galupino@dlsu.edu.ph*

Abstract: The determination of ground elevation is an essential parameter in civil engineering projects, and Light Detection and Ranging (LiDAR) technology has been generally utilized to gather this information, however, LiDAR has its drawbacks. Another way to obtain ground elevation is through exhaustive geotechnical investigations. It is advantageous to have a general overview of the elevation of a site, however, exhaustive geotechnical investigations only provide information on the specific project's location, leaving other locations unknown. In response to this, machine learning models were utilized to estimate ground elevation data for selected locations in Metro Manila, Philippines, as a case study area. Various machine learning models were trained, including the Linear Regression Model, Quadratic Regression Model, Tree Regression Model, Boosted Trees Model, and Artificial Neural Network. Among these models, the Tree Regression Model has the highest accuracy. To validate the estimated Ground Elevation data, it was compared with the Digital Terrain Model (DTM).

Key Words: Machine Learning; Metro Manila; Ground Elevation; Geospatial Intelligence

1. INTRODUCTION

The Philippines is known for its high seismic activity, which can be attributed to its unique geographic location. Being situated along the Pacific Ring of Fire, the country is frequently affected by volcanic and tectonic activity. This continuous movement and shifting of tectonic plates have contributed to the Philippines' diverse soil types and irregular ground elevation. Modern Civil Engineers are faced with a significant challenge when it comes to planning and designing construction projects. In order to ensure that their projects are successful, they require accurate and detailed site information (Wang, 2019). However, due to the limited budget and timeline of many projects, it is often only possible to collect a limited amount of site data. This limited data collection can have significant consequences. For example, engineers may only be able to collect data at the project site itself, leaving data at other locations unknown (Anbazhagan et al., 2016). At the local level, there are numerous efforts to bridge this gap. Effort such as unification of soil parameter quantification, soil property maps, and hazard assessment maps were done. There are still limitations to these studies, one of the challenges with these studies is the use of "inverse distance weighting" or "Kriging" to perform interpolation in Geographical Information Systems (GIS). These methods are used to estimate values for areas with missing data by using the values from surrounding data points. However, if the distribution of sample data points is uneven, the quality of the interpolation result may decrease. This means that in areas where data points are more sparsely distributed, the accuracy of the interpolated results may be lower.

Understanding the ground elevation is essential for many aspects of Civil Engineering design and construction. It provides an important reference point for a variety of applications, including ground water elevation, pipe embedment depth, foundation embedment depth, and watershed delineation. LiDAR (Light Detection and Ranging) is a new technology that is being used to determine ground elevation in Civil Engineering. LiDAR is a remote sensing technique that uses laser light pulses to measure Earth distances. While LiDAR has many advantages, there are also some disadvantages to using this technology. One of the main drawbacks of LiDAR is the high operating costs associated with this technology. This makes it difficult for smaller construction projects or those with limited budgets to use LiDAR to determine ground elevation. Additionally, LiDAR can be ineffective during heavy precipitation or low cloud cover, which can limit its usefulness in certain environments. Another issue with LiDAR is that visibility can be diminished at high sun angles and reflections. This can make it difficult to obtain accurate measurements of ground elevation in certain situations. Additionally, LiDAR can be unreliable when it comes to water depth and turbulent crashing waves, which can make it challenging to use this technology for underwater mapping or other applications. LiDAR can also struggle to penetrate extremely dense forests, which can lead to errors in elevation measurements. Similarly, this technology is incapable of penetrating dense vegetation, which can limit its usefulness in certain environments.

During Geotechnical Site Investigation, Ground Elevation is determined. Locally, at each construction site, a Geotechnical Site investigation must be conducted, and a Professional Report must be submitted to evaluate in-situ soil parameters for the design and analysis of foundations, particularly for structures with two or more storeys. However, geotechnical site investigations are only conducted at the project site, leaving data at other locations unknown. To fill this gap, additional site investigations are required, which can be expensive. To address this issue, machine learning has played a significant role in the development of cost-reduction models. These models can automate the processing of ground elevation data by learning from the available data, recognizing its patterns, and making decisions with minimal human input.

The objective of this study is to maximize the utilization of collected geotechnical borehole data by applying Machine Learning Modeling to estimate ground elevation for specific locations in Metro Manila, Philippines.

2. METHODOLOGY

To outline the methodology, data collection and modeling using machine learning techniques were conducted. The developed models were then applied to the case study, the flowchart is shown in Fig. 1:

Fig. 1 Methodology of the study

2.1 Case Study

The scope of this study is focused on the Metro Manila area in the Philippines, which is also known as the National Capital Region (NCR). Covering a total area of 619.57 km^2, Metro Manila is the country's capital and one of its three metropolitan areas. It comprises sixteen cities and one municipality, including Manila, Quezon City, Caloocan, Las Pinas, Makati, Malabon, Mandaluyong, Marikina, Muntinlupa, Navotas, Paranaque, Pasay, Pasig, San Juan, Taguig, and Valenzuela shown in Fig. 2, with a total of 1,690 Barangays.

Being the cultural, economic, educational, and political hub of the Philippines, the region has a profound impact both locally and internationally, with its designation as a global power city. It has significant influence on various fields such as commerce, finance, media, art, fashion, research, technology, education, and entertainment.

2.2 Data

The borehole information parameters, particularly the Ground Elevation, were extracted and formatted to MS Excel for use in Machine Learning modelling. The data extracted from the borehole logs will represent its location through Latitude and Longitude, as well as the Ground Elevation. To complement this, the United States Geological Survey (USGS) provided the Digital Terrain Model (DTM) for elevation data of Metro Manila.

The study assessed the density of boreholes per city in Metro Manila and found that a density of one borehole per square kilometer was followed (Galupino and Dungca, 2019). However, there were cases where the number of usable boreholes was insufficient. To evaluate the sufficiency of the data, accuracy rates, coefficient of determination (R^2), and root mean square error (RMSE) were examined.

Geographic Information System (GIS) software was utilized to determine the area of each zone. If a Barangay had an area greater than 1 km^2, it was microzoned into smaller zones to ensure that the predicted properties represented an area of less than 1 km^2. The centroid of each zone was then determined using GIS software and expressed in Latitude and Longitude coordinates.

2.3 Machine Learning Models

To estimate the Ground Elevation of a target location, the study first trained Machine Learning Models. Traditional Regression Models were used because the output data is a numerical value. These models include Linear Regression Model (Rath et al, 2020), Quadratic Regression Model (Sun and Wang, 2020), Tree Regression Model (Pekel, 2020), Boosted Trees Model (Elith et al, 2008), and Artificial Neural Network (Hopfield, 1988).

The modelling process involves identifying independent and dependent variables. In this case, the independent variables are latitude and longitude, while the dependent variable is the ground elevation. The machine learning algorithms use these variables to create a model that can predict the ground elevation at a specific location based on its latitude and longitude coordinates.

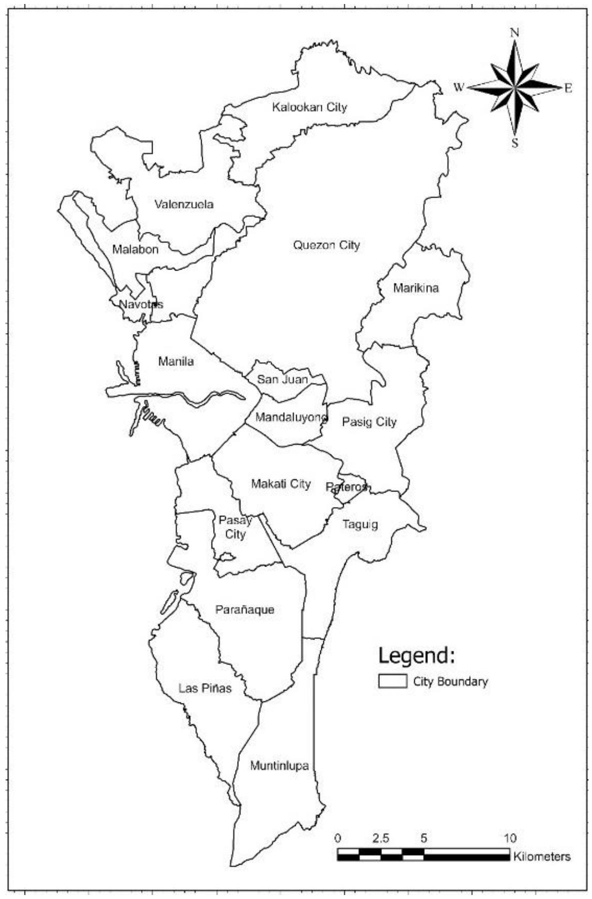

Fig. 2. Metro Manila as a case study area

The study utilized the Matlab Regression Learner program for coding, which served as the foundation for training and validating regression models. The program followed a typical procedure for training regression models, where data selection and validation were conducted first. Afterward, machine learning regression model tuning, also known as hyperparameter tuning, was performed. This was followed by machine learning regression model training, which aimed to compare the R^2 (Cheng et al,

2014) and RMSE (Wilmott and Matsuura, 2005) of competing models side-by-side to select the best one.

2.4 Maps and Validation

make coherent "Once the required data has been imported into a GIS platform, it may be integrated with other data layers to generate many unique maps such as the elevation profile of Metro Manila.

To validate the generated ground elevation data, the researchers compared it with the Digital Terrain Model (DTM) of Metro Manila provided by the United States Geological Survey (USGS). This comparison helped ensure the accuracy and reliability of the estimated parameters.

3. RESULTS AND DISCUSSION

3.1 Microzonation

A total of 1,656 geotechnical borehole data were collected both within and outside of Metro Manila, shown in Fig. 3. The Geographic Information System (GIS) software was used to calculate the area of each barangay.

For Barangays with an area greater than 1 km², they were divided into smaller zones, and the centroid of each zone was determined in Latitude/Longitude format using GIS. As a result, the original total of 1,690 Barangays (Zones) increased to 2,036 zones, shown in Fig. 4. This increase in the number of zones allowed for the modeling of elevation to further zones.

3.1 Performance of Machine Learning Models

The sufficiency of the usable borehole data was determined in the study by computing the R^2 and the RMSE, shown in Fig. 5. Among the models used, the Tree Regression Model had the highest R^2 value and the lowest RMSE value. Specifically, for the 2.67 BH/km² density, the Tree Regression Model had an R^2 value of 1 and an RMSE value of 2.73×10^{14}. Therefore, it was determined to be the best performing model among the Machine Learning Models.

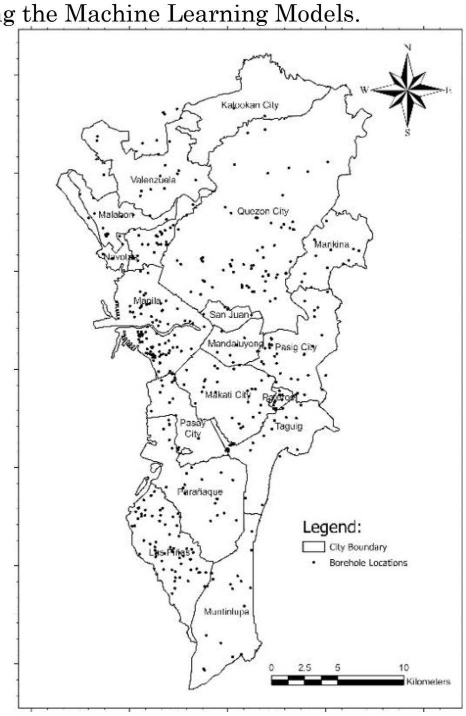

Fig. 3. Collected Data in Metro Manila

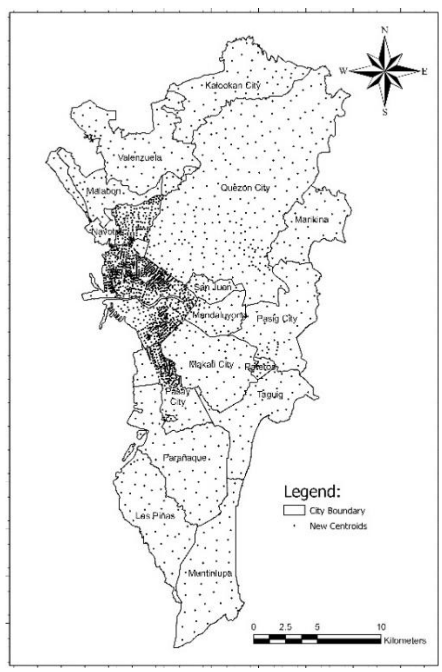

Fig. 4. Centroids of the microzoned Barangays

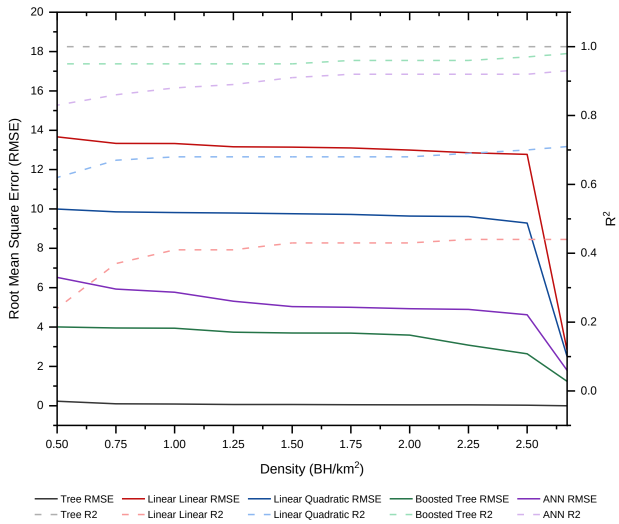

Fig. 5. Performance of Models with varying densities

3.2 Generated Map and Validation

The study generated a map using the trained machine-learning model. Estimated elevation ranges from 2 meters to 89 meters above mean sea level, shown in Fig. 6.

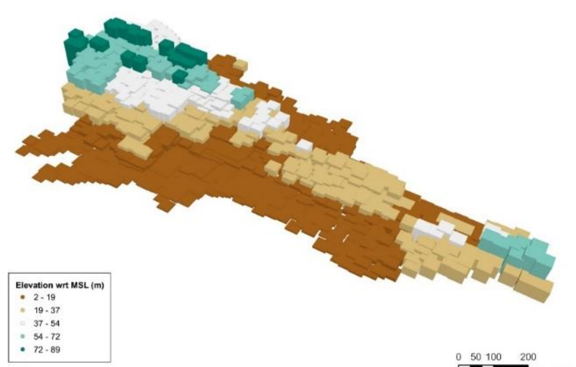

Fig. 6. Generated Map

The study area was divided into three main regions based on their elevation characteristics: the Coastal Area, the Plateau Area, and the Plains Area. The Coastal Area is located on the west side of the Plateau and faces the South China Sea, while the Plateau Area is situated in the middle of Metro Manila. The Plains Area is located between the Plateau and the Province of Rizal and is facing Laguna de Bay. The Marikina West Valley Fault serves as the boundary between the Plains and the Plateau.

The USGS provided the study with topographic maps and GIS data for elevation, which were used to extract the DTM of Metro Manila. To validate the extracted DTM, data samples from the centroids shown in Fig. 3 were extracted using a GIS software, and compared to the USGS DTM, shown in Fig. 7. The elevation from 2m to 30m showed data points near the equality line, indicating good agreement between the two models. However, for elevations from 31m to 89m, the data points were sparse, which could be due to gradual changes in elevation or the lack of sufficient training data from those locations.

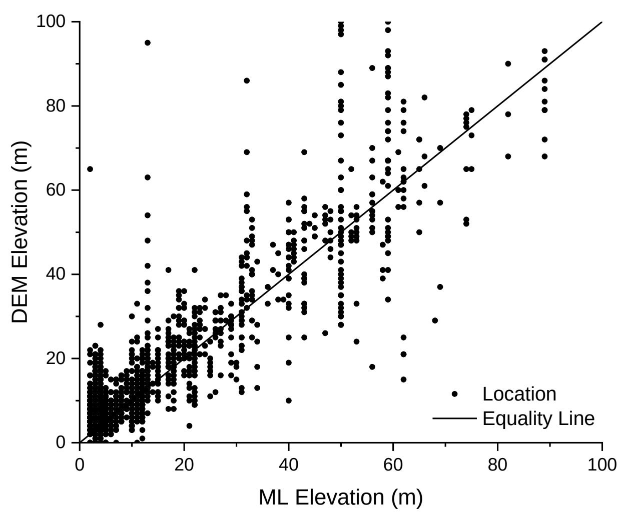

Fig. 7. ML vs. USGS

4. CONCLUSIONS

This study aims to use machine learning modeling techniques to estimate ground elevation for Metro Manila, Philippines. The study used regression models for numerical output data. Among the models, Tree Regression Model proved to be the best with an R^2 value of 1 and an RMSE of 2.73x1014 for the 2.67

BH/km² density.

The study increased the number of zones to 2,036 from the original total of 1,690 Barangays (Zones) through microzonation. The generated map revealed elevation from 2 meters to 89 meters above mean sea level, showing different regions such as the Coastal, Plateau, and Plains areas, which are facing the South China Sea, situated between the Plateau and the Province of Rizal, and facing Laguna de Bay, respectively.

The study also validated the generated data by comparing it with the DTM of Metro Manila provided by USGS.

5. REFERENCES (use APA style for citations)

Anbazhagan, P., Uday, A., Moustafa, S. S. R., & Al-Arifi, N. S. N. (2016). Correlation of densities with shear wave velocities and SPT N values. Journal of Geophysics and Engineering, 13(3), 320–341.

Cheng, C. L., Shalabh, & Garg, G. (2014). Coefficient of determination for multiple measurement error models. Journal of Multivariate Analysis, 126, 137–152.

Elith, J., Leathwick, J. R., & Hastie, T. (2008, July). A working guide to boosted regression trees. Journal of Animal Ecology.

Galupino, J. G., & Dungca, J. R. (2019). Quezon City soil profile reference. International Journal of GEOMATE, 16(58)

Hopfield, J.J., Artificial neural networks. IEEE Circuits and Devices Magazine, 1988. 4(5): p. 3-10.

Pekel, E. (2020). Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3–4), 1111–1119.

Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14(5), 1467–1474.

Sun, Y., & Wang, Q. (2020). Function-on-function quadratic regression models. Computational Statistics and Data Analysis, 142.

Wang, Z. (2019). Site Selection, Power Load, and Power Generation Procedures. Design of Solar Thermal Power Plants, 417–424.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79–82.