Improving Accuracy of Geoid Undulation Model Using Machine Learning Approaches for Sri Lanka

Jayawardana S.S.U. [1*], Jayasinghe U.D.T.M.[1] and Kekulawala K.V.S.P [2]

[1]Apprentice Surveyor, Institute of Surveying and Mapping, Sri Lanka

[2]Senior Superintendent of Surveyor, Institute of Surveying and Mapping, Sri Lanka

*shamee.suj@gmail.com (*Corresponding author's email only)

Abstract: *Surveying techniques are evolving from traditional methods to modern approaches, prominently featuring the Global Navigation Satellite System (GNSS), which enhances efficiency and cost-effectiveness. However, real-time applications of GNSS height data face challenges, particularly in achieving accurate elevation determination relative to mean sea level datum, essential for various geodetic, geophysical, and engineering tasks. In regions like Sri Lanka, surface irregularities hinder precise elevation measurements, highlighting the need for a well-defined local geoid model to obtain accurate orthometric heights. This study develops a comprehensive geoid undulation model for Sri Lanka by integrating Gravity/Orthometric and GNSS/Orthometric undulation data with machine learning techniques. Gravity/Orthometric undulation is sourced from the XGM2019e_2159 gravity model provided by the International Centre for Global Earth Models (ICGEM), while GNSS/Orthometric undulation data is derived from known orthometric and GNSS data. Utilizing a Gradient Boosting Regressor (GBR), establish the relationship between gravity and normal undulation data. The model is trained on 3,425 control points (80% of the dataset) and tested on 857 control points (20%). The results yield a root mean square error (RMSE) of ± 0.080m for the training dataset and ± 0.085m for the testing dataset, with 97% of test points within a ± 0.200m range. This study underscores the effectiveness of machine learning and gravimetric data in enhancing geoid undulation model accuracy, indicating that future improvements could be achieved with more precise gravimetric data.*

Keywords: Orthometric Height, Ellipsoidal Height, Geoid Undulation, Gradient Boosting Regressor, Machine Learning

Introduction

Surveying techniques have undergone a significant transformation in recent years, evolving from traditional methods to modern, technologically advanced approaches. The Global Navigation Satellite System (GNSS) has emerged as a pivotal tool in this transformation, offering enhanced efficiency and cost-effectiveness in geospatial data collection. Despite its many advantages, the use of GNSS in real-time applications for height determination poses challenges, particularly in obtaining accurate elevations relative to the mean sea level (MSL) datum. This is a critical requirement for various geodetic, geophysical, and engineering applications, as precise elevation data are essential for infrastructure development, land use planning, and environmental monitoring.

In regions like Sri Lanka, the determination of accurate orthometric heights is further complicated by surface irregularities and complex terrain, which can distort GNSS-based height data. The key to resolving this challenge lies in the development of a reliable local geoid model. A geoid model defines the shape of the Earth's gravitational potential and serves as a reference surface for determining orthometric heights. However, current geoid models for Sri Lanka do not provide the level of accuracy required for many high-precision applications, prompting the need for further improvement.

The motivation for this study stems from the need to enhance the accuracy of GNSS-based elevation measurements in Sri Lanka. Existing global geopotential models, such as XGM2019e_2159, provide valuable data, but they may lack the localized precision required for applications in regions with unique geophysical characteristics. The objective of this study is to develop a comprehensive geoid undulation model tailored for Sri Lanka by integrating Gravity/Orthometric and GNSS/Orthometric undulation data. This integration is achieved through the use of machine learning techniques, specifically the Gradient Boosting Regressor (GBR), to establish a relationship between gravity undulation and orthometric heights.

Literature Review

The research conducted by Mr. Vipula Abeyaratne on the "Assessment of EGM2008 over Sri Lanka" provides valuable insights into the application and effectiveness of the EGM2008 geoid model in the Sri Lankan context. The study underscores the critical role of validation and quality assessment in geoid modeling, which is essential for ensuring that the models accurately reflect the gravitational variations specific to the region.

Abeyaratne's findings suggest that while EGM2008 serves as a useful reference, there are opportunities for improvement in its accuracy when applied to Sri Lanka. The article emphasizes the necessity for ongoing research and enhanced data collection efforts to refine geopotential models. This continuous effort is vital for meeting the diverse needs of user communities, including those involved in engineering, environmental monitoring, and geospatial analysis.

The development of a hybrid geoid model for Sri Lanka utilizing global gravity field model data is a significant undertaking that combines advanced geospatial techniques with local topographical data. In this study, the research aimed to create a more accurate representation of the geoid by integrating a reliable Global Geopotential Model (GGM) with local measurements. The methodology involved a manual clustering process to account for the topographical variations across Sri Lanka. By employing 21 Fundamental Bench Mark (FBM) data points, the study ensured that the model was grounded in precise local measurements, which is crucial for enhancing the accuracy of the geoid model. To interpolate the data, the researchers utilized least squares adjustment and Inverse Distance Weighting (IDW) techniques.

The research conducted in Kuwait aimed at enhancing the accuracy of the local geoid model through the application of machine learning techniques to geoid residuals derived from GPS and levelling data by utilizing 78 GPS/levelling points. The study incorporated the effects of Global Geopotential Models (GGMs) and Residual Terrain Models (RTM) to determine geoid undulation. Three machine learning algorithms were employed for modeling the geoid residuals: Minimax Probability Machine Regression (MPMR), Gaussian Process Regression (GPR), and Multivariate Adaptive Regression Splines (MARS). The performance of the interpolation models yielded results of 1.377 m for MPMR, and 1.375 m for both GPR and MARS, indicating a high level of consistency among the machine learning approaches used. This research contributes significantly to the ongoing efforts to build a more accurate geodetic model, which is essential for various applications in geodesy, surveying, and related fields.

The German Combined Quasi Geoid GCG2016 represents a significant advancement in geospatial modeling for Germany. It has achieved a high-resolution hybrid model by integrating gravity and terrain data with a global geopotential model through spectral combination techniques. The adjustment of the gravimetric quasi-geoid to GNSS/levelling points enhances its accuracy, utilizing a correction surface derived from interpolation and collocation methods. The estimated overall accuracy of approximately 1 cm underscores the model's reliability for various applications in geodesy and related fields.

Methodology

In this research, a comprehensive methodology was employed to enhance the accuracy of geoid modelling through the use of polynomial fitting, least squares adjustment, and Gradient Boosting Regressor (GBR) techniques. The polynomial fitting and least squares adjustment were utilized to establish a foundational model based on the training data, allowing for the refinement of the geoid estimates.

The Gradient Boosting Regressor (GBR) was then applied to predict data, leveraging its ability to handle complex relationships within the dataset and improve predictive performance. The flow of this methodology is visually represented in Figure 1, which outlines the sequential steps taken in the analysis, from data preparation and model training to prediction and validation.

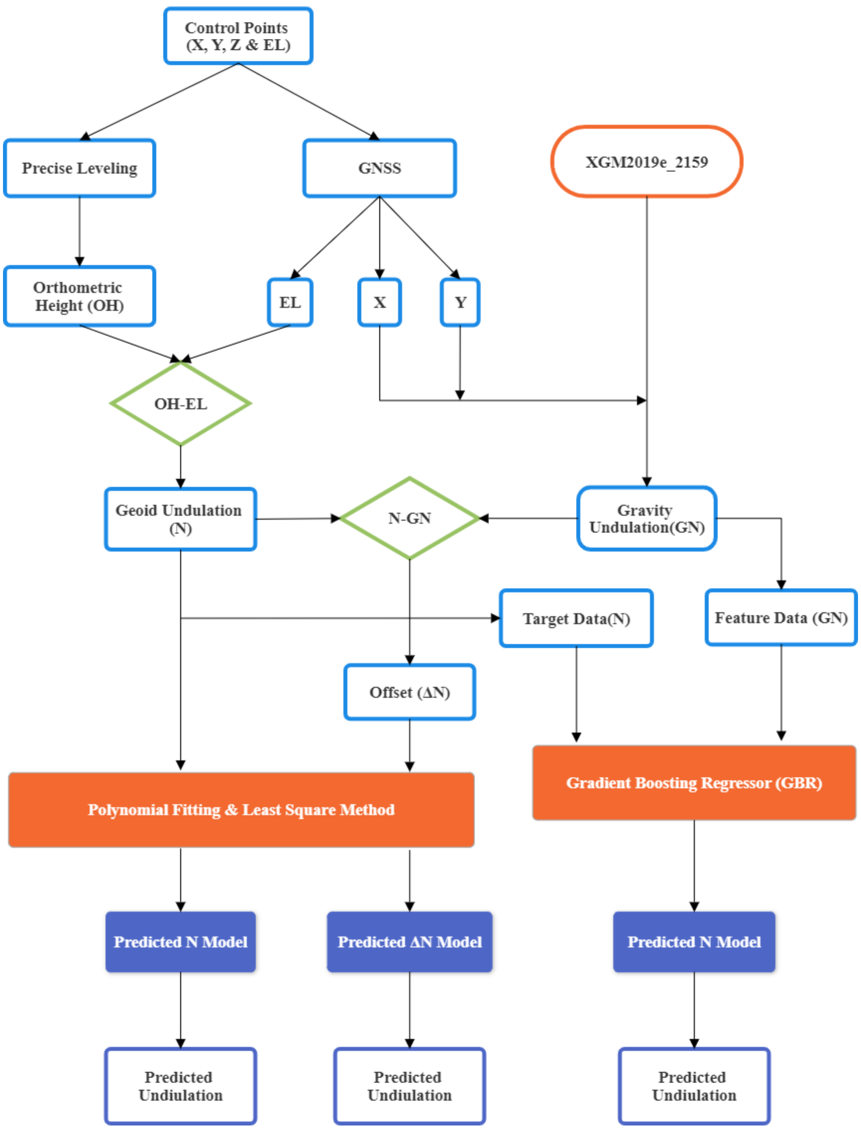

Figure 1: Methodology

a. Data Collection:

For the development of the undulation model, the following data types have been collected.

- Horizontal control points (East, North and Ellipsoid heights)
- Vertical control (primary, secondary and tertiary)
- Gravity Undulation (XGM2019e_2159)

The research involved the collection of existing main control points, from the Sri Lanka National Geodetic Control Network (SLD 99), along with First Order and Second Order Level line points. Figure 2 illustrates the distribution of all these points, providing a visual representation of their locations across the network.

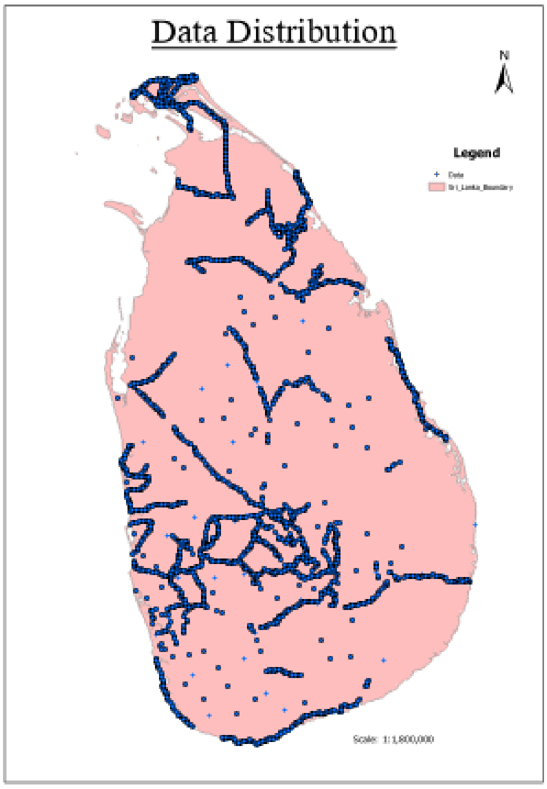

Figure 2: Data Distribution

In the preparation of the Gradient Boosting Regressor (GBR) model and the Hybrid Model, gravity undulation data is essential. To obtain this data, the International Centre for Global Models (ICGEM) was utilized. Specifically, the model XGM2019e_2159 was selected from the ICGEM website to extract gravity height anomaly data. This extraction was performed using the regular grid option, ensuring that the data aligns with the availability and requirements of the study. The gravity undulation data obtained will play a critical role in enhancing the accuracy and reliability of the geoid modelling efforts.

b. Data Preprocessing:

Due to challenges encountered during field verification, the quality of some data points was compromised. To address this, gravity undulation values were extracted from the gravity undulation surface for 4,635 control points that had GNSS/levelling data. Following this extraction, the undulation offset (ΔN) was calculated as the difference between gravity undulation and orthometric undulation (N - GN).

To facilitate a comprehensive analysis, the orthometric undulation, gravity undulation, and undulation offset (ΔN) were plotted on a graph. This visualization allows for a clear comparison of the different undulation metrics, revealing their variations and interrelationships. Figure 3 illustrates this comparison, highlighting the relationships between these metrics and providing insights into how the compromised data quality may have influenced the results.

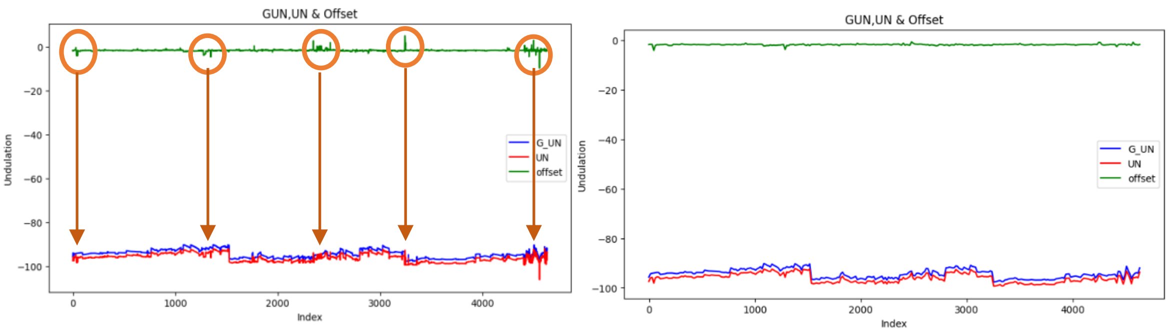

Figure 3: Outliers and Cleared Data Set

The small spikes and dips in the green line on the graph indicate irregularities where the undulation offset (ΔN) changes abruptly. These anomalies could be attributed to local geological features, measurement errors, or other factors. Notably, all level lines appear to align with the road network, suggesting minimal influence from local geological features. Therefore, the primary issues likely stem from inaccuracies in the ellipsoid height of GNSS observations at the level line points and incorrect Mean Sea Level (MSL) values at the control points from the Sri Lanka National Geodetic Control Network (SLD 99).

Therefore, the spikes identified as irregularities can be considered outliers in the dataset. To address this, statistical methods could be applied to determine the mean and variance of the dataset. By calculating these metrics and using a 95% confidence level, outliers can be effectively identified and removed from the dataset. This process ensures that the remaining data is more accurate and reliable for subsequent analysis.

After removing the outliers, the offset has been smoothed, resulting in a more stable orthometric undulation without irregular up and down variations. The figure 3 illustrates the data post-outlier removal, demonstrating a more consistent and refined representation of the orthometric undulation.

c. Methods:

Polynomial Fitting

Polynomial fitting involves finding a polynomial function that best fits a set of data points, typically through a least squares approach. The polynomial can be of various degrees (linear, quadratic, cubic, etc.) based on the complexity of the data and desired accuracy.

$$N(x, y) = \sum_{j=0}^{n}\sum_{i=0}^{n} \left[a_{ij} x^i y^j \right] : i + j \leq n$$

Hybrid Undulation Model Theory

The hybrid undulation model was generated using the difference between gravity undulation and GNSS/levelling undulation. The difference or the offset (ΔN) is considered as the undulation and inputted to the model. The resulting model demonstrated a smoother model than the normal undulation model as it exhibited minimal deviation values.

The hybrid model is obtained from gravity geoid model and offset, where gravity undulation obtains from XGM2019e model.

Hybrid Model = Gravity Geoid Model (XGM2019e) + Offset

Orthometric height can be derived from the hybrid model as follows,

Offset (ΔN) = NGeoid - NGravity

NGeoid = NGravity + Offset (ΔN)

Orthometric Height (h) = Ellipsoid Height (H) - NGeoid

Therefore,

Orthometric Height (h) = Ellipsoid Height (H) – (NGravity + Offset (ΔN))

Machine Learning - Gradient Boosting Regressor (GBR)

Predicting geoid undulation values using machine learning techniques represents an innovative approach to geospatial analysis. Machine learning offers powerful tools for modelling complex relationships between variables, which can significantly enhance the accuracy of geoid predictions.

Understanding the Relationship

The relationship between gravity undulation and orthometric undulation is inherently non-linear. Traditional methods might struggle to capture this complexity, but machine learning techniques, particularly Gradient Boosting Regressor (GBR), are well-suited to this task. GBR builds an ensemble of decision trees, with each tree aiming to correct the errors of the previous one, effectively modelling complex, non-linear relationships.

Steps in Gradient Boosting Regressor

1. Initialize the Model:

The process begins with an initial model, typically a simple constant value that minimizes the loss function (e.g., the mean of the target values for regression tasks).

$$Fo(x) = argmin_c \sum_{n=1}^{n} L(yi, c)$$

where L is the loss function, yi are the target values, and c is a constant.

2. Compute Residuals:

For each subsequent model, the residuals (errors) of the previous model are computed. These residuals represent the difference between the actual target values and the predictions made by the current model.

$$r_{im} = -\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}$$

where r_{im} are the residuals for the m-th iteration, (y_i) are the actual target values, and F(x_i) are the predicted values

3. Fit a New Model: A new decision tree is fitted to the residuals. This new model aims to predict the residuals (errors) of the previous model.

$$h_{m(x)=}argmin_n \sum_{n=1}^{n} (r_{im} - h(xi))^2$$

Where $h_m(x)$ is the new model fitted to the residuals.

4. Update the Model: The predictions of the new model are added to the previous model with a learning rate (shrinkage parameter) to control the contribution of each tree.

$$F_m(x) = F_{m-1}(x) + \eta \, h_m(x)$$

Where $F_m(x)$ is the updated model, $F_{m-1}(x)$ is the previous model, η is the learning rate, and $h_m(x)$ is the new model.

5. Repeat: Steps 2 to 4 are repeated for a specified number of iterations or until the model converges. Each new model incrementally improves the overall model by reducing the errors of the previous models.

d. Data Prepared to Model:

Finally, after filtering the data, 4,282 data points were selected for processing. Out of these, 3,425 points were allocated as training data for model development, while the remaining 857 points were reserved for testing and validating the models. This split allowed for a comprehensive evaluation of the models' performance. Using these datasets, three models were processed: the Normal Undulation Model, the Hybrid Undulation Model, and the Gradient Boosting Regressor (GBR) Model. The training data was used to build and calibrate these models, while the test data was employed to assess their accuracy and effectiveness in predicting undulation values. This approach ensures that the models are rigorously evaluated and optimized for reliable performance. The following figure shows the train data distribution and test data distribution.

Results and Discussion

Three models were employed for data processing: polynomial fitting and least squares adjustment were used for the orthometric GNSS undulation model and the hybrid undulation model, while the Gradient Boosting Regressor (GBR) machine learning method was applied for the third model. The results obtained for these models include the Root Mean Squared Error (RMSE) for both the training and test datasets, which evaluates the models' prediction accuracy. Additionally, residual plots were generated for both the training and test datasets to visually assess the model's performance and identify any patterns or inconsistencies in the residuals. Finally, predicted undulation plots for the 486,168 grid points were produced, providing a comprehensive visualization of the undulation predictions across the entire study area.

Gradient Boosting Regressor (GBR) Model

In this method, residual values of training and test data varied not exceeding 0.3m. It was significantly different from orthometric GNSS undulation model and the hybrid undulation model. For the training dataset, the Root Mean Square Error (RMSE) is ±0.080 meters, and for the test dataset, it was ±0.085 meters. In the training data, 98% of residuals were less than 0.2 meters, with 51% under 0.05 meters. Similarly, in the test data, 97% of residuals were below 0.2 meters, with 49% under 0.05 meters. Also, all residual values were less than 0.3m. The table below shows the residual ranges and data counts for both the training and test datasets.

Table 1: Range of training data & test data

Training Data		**Testing Data**	
Range	Data count	Range	Data count
0.0 - 05m	1754	0.0 - 05m	424
0.05 - 0.1m	945	0.05 - 0.1m	236
0.1 - 0.2m	680	0.1 - 0.2m	174
0.1 - 0.3m	45	0.2 - 0.3m	23

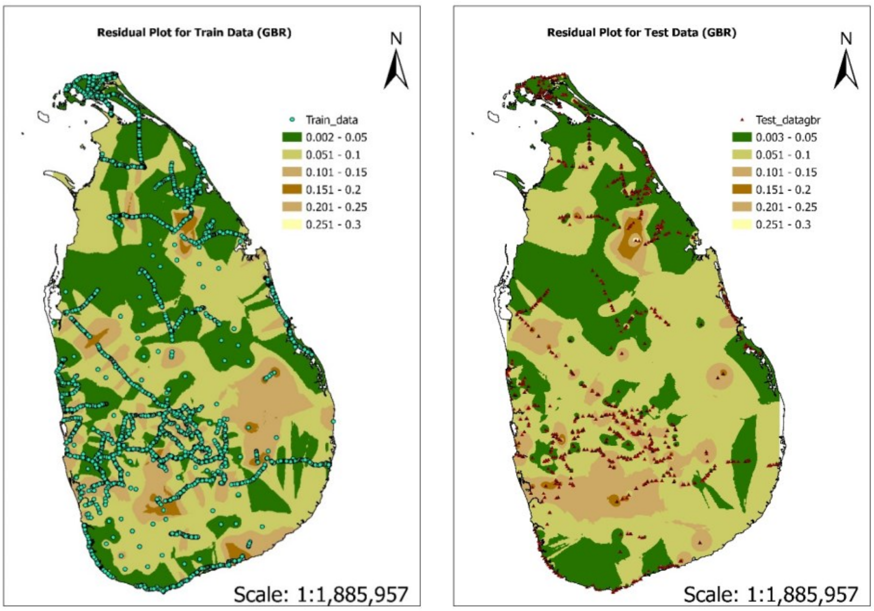

Figure 4: Train data and Test data Residual Plot

Hybrid Undulation Model

The orthometric GNSS undulation model predicted undulation values from the model compared to the actual data. They have some deviation, indicating that most predictions were close to the actual values, though some points fall outside the expected range due to reduced model accuracy.

In the hybrid undulation model, the offset (ΔN) difference between orthometric undulation and gravity undulation was used with the gravity undulation derived from the XGM2019e_2159 model. The accuracy of this model depends on the accuracy of the gravity values. Both orthometric and gravity undulation are equipotential surfaces with a non-linear relationship, which were used to build the hybrid model.

Gravity undulation, derived from global geopotential models like XGM2019e_2159, captures variations in the Earth's gravitational field with high precision. However, these models may have limitations in local accuracy due to factors like data resolution and regional anomalies.

The hybrid model accounts for the non-linear relationship between the two undulation types, using the offset (ΔN) to minimize residuals. This approach helps to reduce the impact of any inaccuracies in the gravity model by effectively 'correcting' the orthometric undulation with precise gravity-derived data. Consequently, the residuals are minimized, leading to more accurate predictions.

While the orthometric/GNSS undulation model tends to have higher residuals due to various local factors such as terrain variations and data inconsistencies, the hybrid model reduces these residuals at the same points by incorporating the gravity undulation data. In general, higher input values can result in higher residuals, but with a hybrid approach, even lower input values maintain low residuals, improving the overall accuracy and reliability of the predictions.

As a result, the residuals between the model's predictions and the actual values are smaller compared to the Orthometric/GNSS undulation model. This improvement is due to the use of offsets (ΔN) in building the model. The model managed to reduce the gap between predictions and actual values, making only minor adjustments when predicting new offsets. Therefore, the residuals for both the training and test datasets have decreased significantly. For the training dataset, the Root Mean Square Error (RMSE) was ±0.093 meters, and for the test dataset, it was ±0.091 meters, both using the 7th-order polynomial, which turned out to be the most effective. In the training data, 95% of residuals were less than 0.2 meters, with 49% under 0.05 meters. Similarly, in the test data, 96% of residuals are below 0.2 meters, with 50% under 0.05 meters. This polynomial order provided the most accurate predictions and the lowest RMSE values for both datasets, resulting in minimal residuals. The table below shows the residual ranges and data counts for both the training and test datasets.

Table 2: Range Of Training Data & Test Data for Hybrid Undulation Model

Training Data		**Testing Data**	
Range	Data count	Range	Data count
0.0 - 05m	1675	0.0 - 05m	430
0.05 - 0.1m	903	0.05 - 0.1m	222
0.1- 0.2m	684	0.1 - 0.2m	171
0.2 - 0.3m	144	0.2 - 0.3m	29
0.3 - 0.4m	18	0.3 - 0.4m	4
0.4 - 0.5m	-	0.4 - 0.5m	1

Orthometric GNSS Undulation Model

The model developed was trained exclusively using undulation values to predict model values for specific coordinates. However, the data collection points were unevenly distributed across Sri Lanka, leading to geographic coverage gaps. This lack of uniform distribution may result in significant data gaps, which can impact the model's accuracy. As a result, the residuals between the model's predictions and the check values might be higher than anticipated. This situation underscores the importance of having a well-distributed dataset to achieve more reliable and accurate undulation predictions. In the training data set, the Root Mean Square Error (RMSE) was ±0.2058 meters, and for the test data set, it was ±0.1936 meters, both at the 7th-order polynomial, which proved to be the most suitable. The train data set 77% of residual data less than 0.2m and 28.5% less than 0.05m. Also, test data set 79% of residual data was less than 0.2m and 28.5% less than 0.05m. This polynomial order provided the most accurate predictions and the lowest RMSE values for both the training and test datasets, reflecting minimal residuals. The table below illustrates the residual ranges and data counts for both the training and test datasets.

Table 3: Range Of Training Data & Test Data for Orthometric GNSS Undulation Model

Training Data		**Testing Data**	
Range	Data count	Range	Data count
0.0 - 05m	979	0.0 - 05m	245
0.05 - 0.1m	820	0.05 - 0.1m	210
0.1- 0.2m	838	0.1 - 0.2m	223
0.2 - 0.3m	419	0.2 - 0.3m	101
0.3 - 0.4m	174	0.3 - 0.4m	36
0.4 - 0.5m	52	0.4 - 0.5m	17
0.5 - 0.6m	40	0.5 - 0.6m	4
0.6 - 0.7m	56	0.6 - 0.7m	10
0.7 - 0.8m	27	0.7 - 0.8m	6
0.8m <	19	0.8m <	5

Residual plots alone do not give a complete picture of residual distribution across Sri Lanka due to the uneven spread of data points. To address this, we used three different models to predict undulation values for 486,168 regular grid points. The orthometric GNSS undulation model and the Hybrid undulation model used north and east coordinates along with gravity undulation values to make predictions. The Gradient Boosting Regressor (GBR) predicts undulation values using gravity undulation values and coordinates. Sri Lankan boundary mask was applied and used the IDW (Inverse Distance Weighting) interpolation method in ArcGIS software was to ensure that interpolation was done only within the region. Gravity values extracted according to the XGM2019e_2159 model for 486168 grid points covered by the whole Sri Lanka and the surrounding area.
The following figure shows the Orthometric GNSS undulation model's predicted undulation values for Sri Lanka, compared with the gravity undulation values developed

using the XGM2019e_2159 model. This comparison highlights the general undulation pattern and its variations with gravity undulation. However, the normal undulation pattern appears abnormal in the central hill region of the country, deviating from the expected shape. Additionally, there is another irregular pattern observed at the bottom of the island.

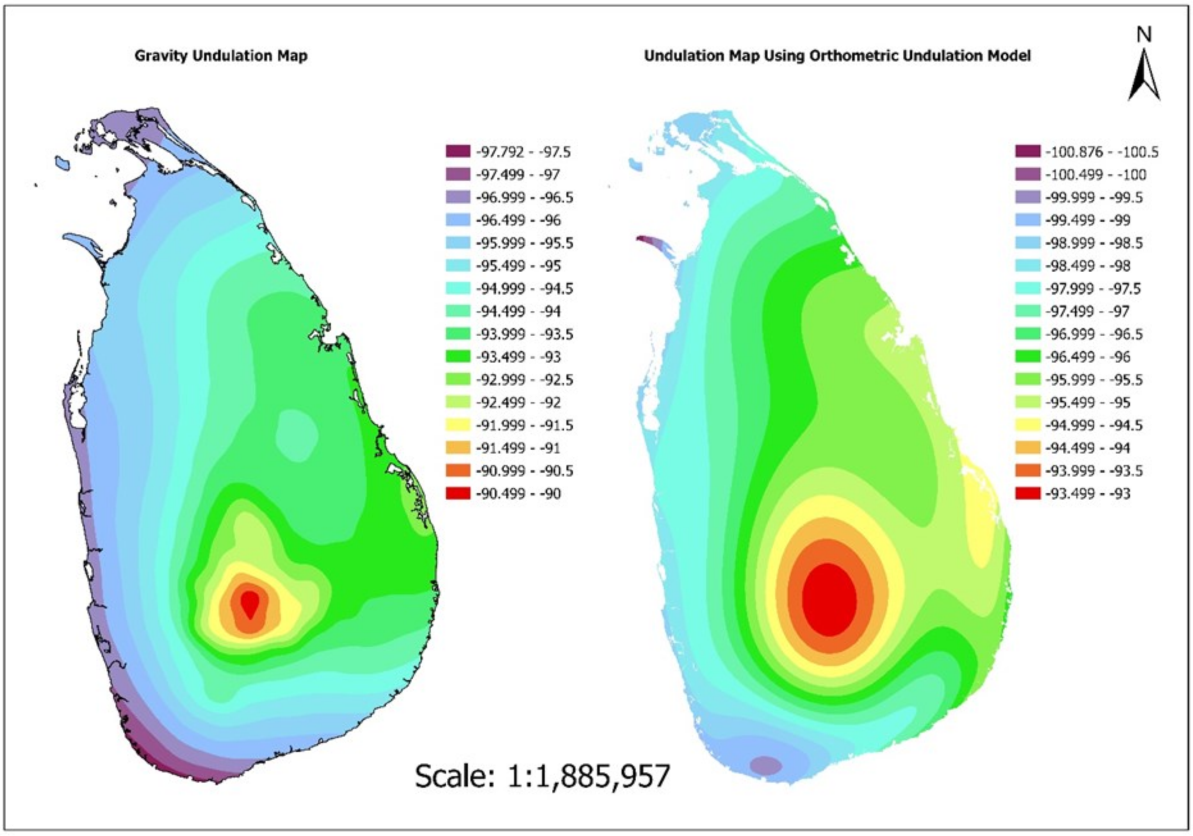

Figure 5: Gravity Model Vs Normal Undulation Model

Another map was created using a hybrid model to predict undulation values by incorporating gravity offset (ΔN) and gravity undulation. The undulation values predicted by the hybrid model are closely similar the gravity undulation values since gravity data plays an important role in constructing this hybrid model.

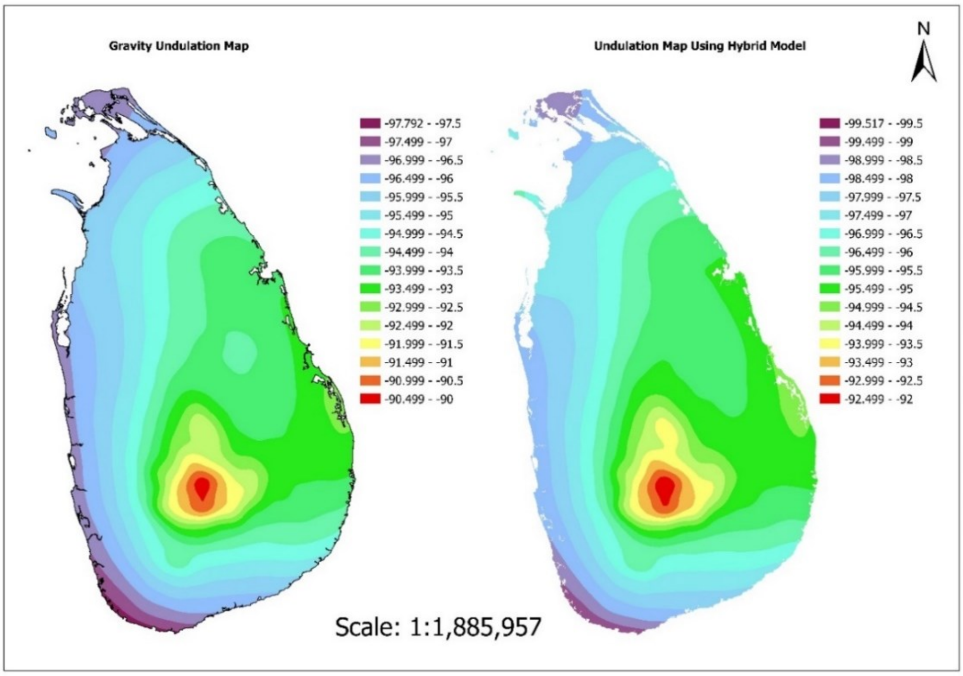

Figure 6: Gravity Model Vs Hybrid Undulation Model

The third map was created using the Gradient Boosting Regressor (GBR) model. It used gravity values from the XGM2019e_2159 model to predict undulation values. The resulting undulation surface closely resembles the gravity undulation surface since it was based on gravity values. This predicted surface was very similar to the natural undulation surface of Sri Lanka.

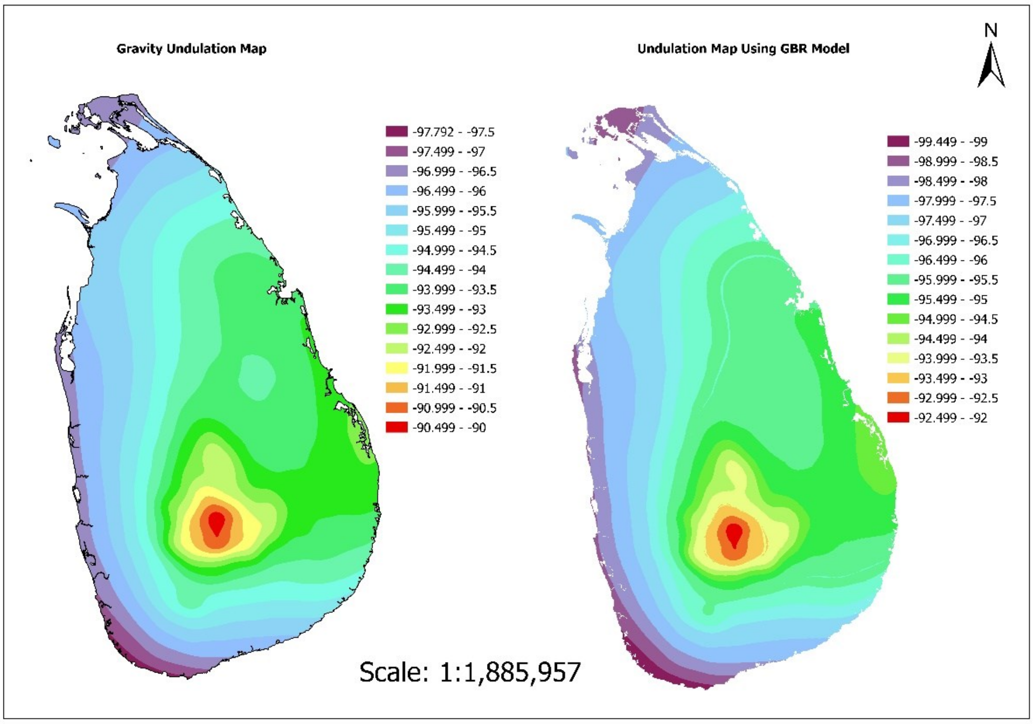

Figure 7: Gravity Model Vs GBR Undulation Model

The evaluation metrics used in the analysis included Root Mean Squared Error (RMSE), and the Orthometric GNSS Model, which relies solely on normal undulation values and showed limited accuracy. In contrast, the Hybrid Method, which incorporates gravity undulation values, demonstrated significantly improved accuracy by reducing residuals. The GBR model outperformed both by effectively using gravity data, highlighting the potential of machine learning in geoid modelling. Table 5 represents the RMSE values for three models.

Table 5: Evaluation RMSE

	Orthometric GNSS Model	**Hybrid Method**	**GBR model**
Test data RMSE	± 0.1936m	±0.0914m	± 0.0855m
Train data RMSE	± 0.2058m	±0.0926m	± 0.0803m

Table 6 presents the residuals, which are the differences between the actual undulation values and the predicted values from each model for a field-verified set of control points. These comparisons allow for an assessment of the accuracy and reliability of each model.

Table 6: Field Data Verification for Models Residuals

Point no	**North**	**East**	**Residual of Undulation values**		
			Normal	**Hybrid**	**GBR**
62PL_329_25	471728.5	506583	-0.08418	0.097737	-0.02859
62PL_329_28	471076.8	505893.2	-0.13743	0.183304	0.01423
A068	470825.7	506213.9	-0.19271	0.171336	0.062657
HBM2	471205.5	506535.7	-0.15138	0.226731	0.005738
HBM5	470264	506531.1	-0.21353	0.124167	0.109796
HBM1	471566.7	506579.9	-0.09857	0.076696	-0.00726
HBM3	471018.2	506766.3	-0.17329	0.19527	0.036849
HBM4	470663.2	506711.9	-0.20693	0.146852	0.085988
33SL-403-066	521759	516392.6	-0.49524	0.066861	-0.00076
33SL-403-068	521496.1	518185.1	-0.54011	0.03618	0.034063

71SL403-073	521610.5	522097.2	-0.62651	0.012552	-0.03191
A121	521578	518073.6	-0.53713	0.039927	0.030063
ABM01	522212.5	522058.9	-0.62538	-0.00265	-0.01922
ABM02	523124.8	522652.7	-0.73714	-0.06609	-0.04165
33PL-303-007	496912.6	499093.9	0.857924	-0.13384	0.039689
33PL-303-008	497318.4	498842.6	0.879193	-0.1174	0.021424
TBM02	497488.6	499621.9	0.917006	-0.07064	-0.02723
TBM03	497408.1	500010	0.832256	-0.17074	0.121685
TBM05	498078.2	499812.1	0.891061	-0.12197	0.064894

The lowest residuals were obtained from the GBR method, although the Hybrid Method also demonstrated low residuals compared to the Orthometric GNSS Model. Orthometric GNSS Model, where most residuals exceeded 0.50m. Thus, the Hybrid Method is more suitable than the Orthometric GNSS Model for prediction. However, when comparing all three models, the GBR method emerges as the most accurate for predicting undulation values. Model accuracy is significantly influenced by the quality of the initial stage data acquisition. To ensure reliable predictions, suitable filters were employed to identify and address outliers in the dataset. This preprocessing step helps in refining the data, which in turn enhances the accuracy and robustness of the predictive models. The accuracy of these predictions is influenced by several factors, including GNSS observation time, accuracy of ellipsoid heights, accuracy of mean sea level (MSL) heights, distribution of observation points, and the precision of the gravity undulation model.

Conclusion and Recommendation

Overall, the lowest RMSE value was obtained by the Gradient Boosting Regressor (GBR) model, indicating its superior performance. In both the orthometric GNSS undulation model and the Hybrid undulation model, the most accurate predictions were achieved using the 7th-order polynomial fitting. When comparing the Hybrid and orthometric GNSS undulation models, the Hybrid model produced the lowest residual values, highlighting its higher accuracy. The accuracy of these models depends heavily on the distribution of points, data quality, and the accuracy of the gravity model. During data processing, the same dataset was used for both training and testing across all three methods. For the GBR model, a data split of 80% for training and 20% for testing was found to be most effective, eliminating the need for manual selection of training and test data. Field observations were used to verify and correct errors in the dataset, and appropriate filters were applied to remove outliers. It was identified that both the Hybrid method and the GBR model are more suitable than the orthometric GNSS undulation model. However, the accuracy of the Hybrid method is dependent on data distribution, actual undulation value accuracy, and gravity undulation value accuracy. Conversely, the GBR model's accuracy is mainly dependent on the accuracy of gravity undulation and actual undulation value, with minimal impact from the distribution of points, as the GBR model builds a relationship between gravity and undulation within the model.

As recommendations,

- Employ advanced statistical filters to identify and correct errors in the dataset effectively.
- Use an accurate gravity model, such as one derived from airborne gravimetric methods, to enhance the precision of the gravity undulation values, as both GBR and Hybrid methods rely heavily on this accuracy.
- Ensure that the dataset is sufficiently dense by addressing gaps, especially in areas away from the road network where many points are currently concentrated.
- In this task only 136 points from the Nation Geodetic Control Network of Sri Lanka (SLD99) were used because other control points of the National Control Network were identified as outliers. However, GNSS observations are more precise of Control Network points than other control points and bench marks. Therefore, re-observation of National Control Network points for Mean Sea Level height is a most suitable approach.

References

Abeyratne, P., Featherstone, W., & Tantrigoda, D. A. (2010, January 1). Assessment of EGM2008 over Sri Lanka, an area where 'fill-in'data were used in EGM2008. ResearchGate. https://www.researchgate.net/publication/228957306_Assessment_of_EGM2008_over_Sri_Lanka_an_area_where_'fill-in'data_were_used_in_EGM2008

Albayrak, M., Özlüdemir, M. T., Aref, M. M., & Halıcıoğlu, K. (2020, May 1). Determination of Istanbul geoid using GNSS/levelling and valley cross-levelling data. Geodesy and Geodynamics. https://doi.org/10.1016/j.geog.2020.01.003

Edirisinghe, D. U., Welikanna, D. R., Periyandy, T., & Bandara, R. (2024, February 1). Development of a hybrid geoid model using a global gravity field model over Sri Lanka. Journal of Applied Geodesy, 0(0). https://doi.org/10.1515/jag-2023-0092

Hiroyuki NAKAGAWA, & Kojin WADA. (2003, March). Development of a new Japanese geoid model, "GSIGEO2000." https://www.gsi.go.jp/common/000001198.pdf

Milbert, Smith, & National Geodetic Survey, NOAA. (1996). Converting GPS Height into NAVD88 Elevation with the GEOID96 Geoid Height Model. GIS/LIS'96 Conference. In press. https://www.ngs.noaa.gov/PUBS_LIB/gislis96.html

Kaloop, M. R., Pijush, S., Rabah, M., Al-Ajami, H., Hu, J. W., & Zaki, A. (2021). Improving accuracy of local geoid model using machine learning approaches and residuals of GPS/levelling geoid height. *Survey Review*, *54*(387), 505–518. https://doi.org/10.1080/00396265.2021.1970918

Nalbant, M., Sahin, N., & Aydin, S. (1998, December). FIG GENETIC RESOURCES AT THE FIG RESEARCH INSTITUTE (AYDIN/TURKEY). Acta Horticulturae, 480, 43–48. https://doi.org/10.17660/actahortic.1998.480.5

Prasanna, H. M. I., & Tantrigoda, D. A. (2009, January 1). Determination of the geoid of central highlands in Sri Lanka. Journal of Applied Geodesy (Print). https://doi.org/10.1515/jag.2009.012

Van Westrum, D., Ahlgren, K., Hirt, C., & Guillaume, S. (2021, January 1). A Geoid Slope Validation Survey (2017) in the rugged terrain of Colorado, USA. Journal of Geodesy. https://doi.org/10.1007/s00190-020-01463-8

Kao, S. P., Chen, C. N., Huang, H. C., & Shen, Y. T. (2014). Using a least squares support vector machine to estimate a local geometric geoid model. Boletim De Ciências Geodésicas, 20(2), 427–443. https://doi.org/10.1590/s1982-21702014000200025