International Journal of Technology Engineering rts Mathematics Science
Vol. 4, No. 1, June2024, pp. 01~06
ISSN: 2583-1224,

Enhancing GPS Positioning Accuracy Using Machine Learning Regression

Durvesh Baharwal, [2]Sanhita Dere, [3]Ashwini Swami

[1,2,3]Artificial Intelligence and Data Science Department, All India Shri Shivaji Memorial Society's Institute of Information Technology, Maharashtra, India
Corresponding Author: Durvesh Baharwal Email ID: durveshbaharwal@gmail.com

Article Information

Article history:

Received May, 2024
Revised June,2024
Accepted June,2024

AISSMS IOIT RESEARCH
International Journal of TEAMS

ABSTRACT

Addressing the inherent challenges of Global Navigation Satellite Systems (GNSS), this research project introduces an innovative approach by combining Least Square estimation with the Time-Differenced Pseudo Range method to enhance Position, Velocity, and Time (PVT) determination. Enhancing GPS Positioning Accuracy Using Machine Learning Regression aims to improve navigation systems, particularly in challenging environments, through the integration of machine learning within the framework of Least Square estimation. The methodology involves a systematic integration process, showcasing the unique amalgamation of traditional GNSS techniques with intelligent learning through Least Square estimation. The outcomes reveal significant improvements in navigation accuracy, with Random Forest Regression emerging as the most effective algorithm among those explored, maintaining its lead with the lowest MAE of around 0.000122. Haversine distance is employed as a crucial metric for quantitative evaluation. The project's practical implications extend to mitigating delays and errors associated with GNSS, such as atmospheric delays and multipath effects. The results underscore the transformative impact of machine learning algorithms in refining GPS positioning accuracy and set a new benchmark for precision in geospatial analysis and positioning systems. The research concludes by highlighting the project's uniqueness, practical applicability, and real-world adaptability—a tangible solution to persistent challenges in satellite-based navigation.

KEYWORDS: GNSS (Global Navigation Satellite Systems), SLS (Simple Least Square), WLS (Weight Least Square), Time-Differenced Pseudo Range (TDP), PVT (Position Velocity Time), Machine Learning Regression, Random Forest Regression, Navigation Accuracy.

1. INTRODUCTION

'Global Navigation Satellite System (GNSS)' research witnesses' remarkable strides in recent years, with a focus on elevating positioning accuracy and addressing challenges across diverse environments and applications. This literature review provides a comprehensive overview of key contributions in this dynamic landscape, encompassing the integration of technologies, innovative algorithms, and machine learning applications.

Boguspayev et al. (2023) offer a foundational exploration of GNSS/INS integration techniques, emphasizing the synergies between satellite navigation and inertial systems. Categorizing integration approaches, including loosely coupled, tightly coupled, and deeply coupled methods, their work serves as a valuable reference, providing a nuanced understanding of GNSS/INS integration strategies [1]. Zawislak et al. (2022) significantly advance the exploration of machine learning applications in GNSS, focusing on multipath detection. Their unsupervised domain adaptation approach, presented at the ION-GNSS+ Conference 2022, markedly enhances multipath detection accuracy, showcasing potential for improving GNSS positioning accuracy, especially in complex urban environments [2]. Maghdid et al. (2021) address challenges of accuracy and cost-efficiency in GNSS-based positioning through optimization approaches. Their survey emphasizes the integration of diverse technologies and provides valuable insights into the impact of optimization algorithms on accuracy improvement, time-to-fix reduction, and cost-effectiveness [3]. Ji et al. (2022) contribute to the field by evaluating GNSS-based velocity estimation algorithms, addressing the critical need for precise velocity determination in applications such as autonomous navigation and geodesy [4]. Complementing these efforts, Kong (2021) introduces an innovative satellite positioning method employing the Total Least Squares (TLS) algorithm, enhancing accuracy and efficiency in satellite positioning [5].

Building upon these foundational studies, this research explores the application of advanced machine learning regression techniques. The methods include 'Support Vector Regression (SVR)', XGBoost, 'Decision Tree Regression (DTR)', and 'Random Forest Regression (RFR)'. Despite facing challenges in applying Gaussian Process Regression (GPR) due to computational limitations, the study aims to contribute to the ongoing development and improvement of GNSS-based navigation and positioning systems.

This research integrates insights from the literature survey to address the crucial aspects of precision and reliability in GNSS systems. The application of machine learning regression techniques opens new avenues for achieving higher accuracy in diverse fields, including autonomous vehicles, surveying, and geodetic applications.

2. PROPOSED SYSTEM

The proposed system architecture for enhancing GPS positioning accuracy integrates traditional methods with advanced machine learning techniques. In the data collection module, diverse GNSS data, including ground truth records, satellite positions, and corrected pseudo ranges, are acquired to represent various environmental conditions. The data pre-processing module employs robust techniques such as outlier removal and noise reduction, ensuring the integrity of the dataset. Feature engineering identifies critical features like corrected pseudo ranges, forming the foundation for both traditional Least Square Estimation and advanced machine learning models. The system incorporates 'Support Vector Regression (SVR)', XGBoost, 'Decision Tree Regression (DTR)', and 'Random Forest Regression (RFR)' in the machine learning integration module. Additionally, a 'Gaussian Process Regression (GPR)' module attempts to model non-linear relationships in the data. The quantitative evaluation module employs metrics such as 'mean absolute error (MAE)' and 'mean squared error (MSE)' for a comprehensive analysis of model accuracy. Results are then analyzed to understand the strengths and limitations of each algorithm. The proposed architecture aims to provide a robust and adaptable solution for improving GPS accuracy in diverse environments.

3. METHDOLOGY

This methodology integrates traditional techniques and advanced machine learning methodologies to enhance GPS positioning accuracy. The initial phase involves ingesting diverse raw GNSS data from ground truth records, satellite positions, and corrected pseudoranges, followed by meticulous preprocessing to ensure data integrity. Feature engineering extracts crucial features like corrected pseudoranges and satellite positions.

In the realm of traditional and improved position estimation, both Traditional Least Square Estimation and Weighted Least Square Estimation are employed. Traditional Least Square Estimation establishes a baseline, while Weighted Least Square Estimation enhances accuracy by incorporating weightings to prioritize more reliable measurements and mitigate the impact of outliers.

The innovation lies in the integration of machine learning regression techniques. 'Support Vector Regression (SVR)', XGBoost, 'Decision Tree Regression (DTR)', and 'Random Forest Regression (RFR)' are employed for improved position estimation. SVR optimizes a hyperplane to minimize errors, XGBoost employs an ensemble of decision trees, DTR leverages decision trees, and RFR aggregates predictions from multiple decision trees. Given below are the equations used for the respective algorithms that are used for the project.

4. SYSTEM ARCHITECTURE

This architecture represents an approach to GNSS positioning that leverages machine learning to improve the accuracy and robustness of error mitigation techniques.

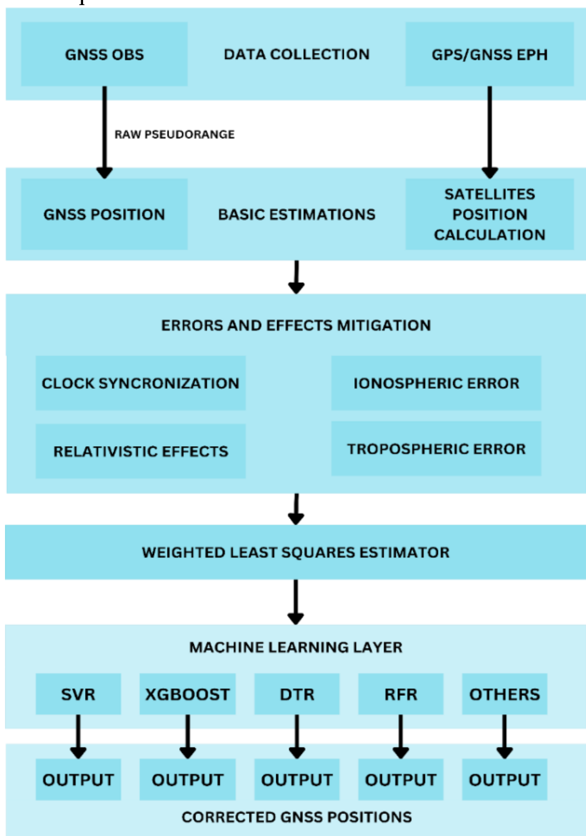

Figure 4.1: System Architecture

Where, GNSS OBS: This block likely refers to GNSS observations, which are the raw data received from GNSS satellites.
Data Collection: This block collects GNSS observations.
GPS/GNSS EPH: This block likely refers to ephemeris data, which is broadcast by GNSS satellites and contains information about their orbits and positions.
Raw Pseudorange: This block indicates raw pseudorange measurements, which are estimates of the distance between a GNSS receiver and a satellite, affected by factors like satellite clock errors and propagation delays.
GNSS Position Here, the GNSS position is calculated based on the raw pseudorange measurements.
Basic Estimations: This block performs basic estimations of the user's position using the pseudorange measurements.
Position Calculation: This block refines the position calculation by incorporating error mitigation techniques.
Errors and Effects Mitigation: This block addresses various errors that affect the accuracy of GNSS positioning, including:
Clock Synchronization Errors: These errors arise from differences between the GNSS receiver clock and the satellite clocks.
Ionospheric Errors: The ionosphere delays GNSS signals, affecting positioning accuracy.
Tropospheric Errors: The troposphere, the Earth's lower atmosphere, also delays GNSS signals.
Relativistic Effects: Relativistic effects, based on Einstein's theory of relativity, can introduce minor errors in GNSS positioning.
Weighted Least Squares Estimator: This statistical method is likely used to refine the position calculation by accounting for the precision of different measurements.
Machine Learning Layer: This block incorporates machine learning algorithms to potentially improve the accuracy and robustness of error mitigation, especially regarding complex error sources that are difficult to model mathematically. The specific machine learning algorithms mentioned include Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Decision Tree Regression (DTR), Random Forest Regression (RFR). These algorithms are likely trained on datasets containing GNSS observations, position information, and various error sources.
Output: The output of the system is a corrected GNSS position that is more accurate than the basic estimates obtained using only pseudorange measurements.

5. TRADITIONAL AND IMPROVED POSITION ESTIMATION

5.1 Traditional Least Square Estimation:

$$\mathbf{X = (A^T W_A)^{-1} A^T W_b} \quad (5.1)$$

where **X** is the estimated position vector, **'A'** is the design matrix, '**W'** is the weight matrix, and '**b'** is the pseudo range residual vector.

5.2 Weighted Least Square Estimation:

Weighted least squares estimation adjusts the traditional method to account for measurement uncertainties. The estimation is given by:

$$\mathbf{X = (A^T W_A)^{-1} A^T W_b} \quad (5.2)$$

Where **W** is a diagonal matrix of weights, allowing for the down-weighting of less reliable measurements.

6. MACHINE LEARNING REGRESSION TECHNIQUES

6.1. Support Vector Regression (SVR):

'Support Vector Regression' aims to find a hyperplane that best represents the underlying mapping function between the input features and the output. The SVR objective function for regression can be defined as follows:

Objective Function:

$$\mathbf{Min}_{V, b, l, l^*} \ \tfrac{1}{2} \|V\|^2 + C \sum_{i=1}^{n} (l_i + l_i^*) \quad (6.1)$$

Subject to:

$$Y_i - v^T \phi (X_i) - b \leq € + l_i \quad (6.2)$$

$$v^T \phi (X_i) + b - Y_i \leq € + l_i^* \quad (6.3)$$

$$l_i , l_i^* \geq 0 \quad (6.4)$$

Here, 'ϕ (xi) represents the feature mapping of the input data X_i', 'w is the weight vector', 'b is the bias term', 'C is the regularization parameter', and ϵ controls the width of the ϵ -insensitive tube.

6.2. XGBoost:

XGBoost is an 'ensemble learning' method that combines the predictions from multiple decision trees. The objective function for XGBoost regression can be expressed as a sum of a loss term and regularization terms:

Objective Function:

$$\mathbf{Obj} = \sum_{i=1}^{n} L (Y_i + Y_i^*) + \sum_{k=1}^{k} \Omega(f_k) \quad (6.5)$$

Where 'L **(Yi** $_+$ $\mathbf{Y_i^*}$ **)'** is the loss term measuring the difference between the '$\mathbf{Y_i}$' predicted and '$\mathbf{Y_i^*}$' actual values. '$\mathbf{\Omega(f_k)}$**'** is the regularization term for each tree.

6.3. Decision Tree Regression (DTR):

'Decision Tree Regression' builds a tree structure to predict the output value for a given set of input features. The prediction for a new input \(x\) is obtained by traversing the tree from the root to a leaf and outputting the average of the training target values in that leaf.

Prediction for Decision Tree Regression:

$$\mathbf{Y^*(X) - 1/N_k \sum_{i=leaf(X)} Y_i} \quad (6.6)$$

Where '$Y^*(X)$' is predicted output for input X, 'leaf(X)' denotes the leaf node that input X falls into and 'N_k' is the number of training samples in leaf k.

6.4. Random Forest Regression (RFR):

'Random Forest Regression' aggregates predictions from multiple decision trees, providing improved generalization and robustness. The prediction for a new input *x* is obtained by averaging predictions from all the trees in the forest.

Prediction for Random Forest Regression:

$$\mathbf{Y^{*}RFR(X) = 1/T \sum_{t=1}^{T} Y^{*}t(X)} \quad \mathbf{(6.7)}$$

Where '$Y^{*}RFR(X)$' is the predicted output for input X using the Random Forest, $Y^{*}t(X)$ is the predicted output for input X from tree t and T is the total number of trees in the Random Forest.

6.5. Gaussian Process Regression (GPR):

The 'Gaussian Process Regression' model predicts the corrected pseudo ranges **(y)** based on the satellite positions **(X)** and potentially other relevant features:

$$\boldsymbol{Y = f\,(X) + \varepsilon} \quad \mathbf{(6.8)}$$

where **f (X)** is the underlying function modeled by the Gaussian process, and '**ε**' represents the 'observation noise'.

The Gaussian Process model can be defined as:

$$\mathbf{f\,(X) \sim G\,P\,(\mu\,(X),\, k\,(X, X'))} \quad \mathbf{(6.9)}$$

where '**μ (X)**' is the mean function and '**k (X, X')**' is the kernel function.

For example, using the 'Radial Basis Function (RBF) kernel':

$$\mathbf{k\,(X, X') = \sigma^2 \exp\left(\frac{||X - X'||^2}{2l^2}\right)} \quad \mathbf{(6.10)}$$

These equations outline the fundamental principles behind each regression algorithm, capturing the optimization objectives and prediction mechanisms.

7. RESULTS AND DISCUSSION

The table provided represents the latitude and longitude coordinates predicted by different machine learning models compared to the actual coordinates. To determine which model works the best and which works the worst, we need to calculate the distance between the actual coordinates and the predicted coordinates for each model.

Table 7.1: Position Comparison

ML Model	**Latitude**	**Longitude**
Actual values	37.46758689	-122.1523673
Random Forest Regression (RFR)	37.46745277	-122.1527937
Support Vector Regression (SVR)	37.468082	-122.15336
Decision Tree Regression (DTR)	37.46716829	-122.1523673
XGBoost	37.46513822	-122.1626473

Based on the distances calculated between the actual coordinates and the predicted coordinates by each model, here are the results:

Random Forest Regression (RFR): 4.48 meters
Support Vector Regression (SVR): 103.47 meters
Decision Tree Regression (DTR): 46.55 meters
XGBoost: 947.25 meters

Figure 7.1: Plot for SVR, DTR, RFR and Actual Positions on Map

When comparing the predicted positions of Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) to the actual coordinates, distinct differences in accuracy are observed. The SVR model predicts coordinates with a deviation of 103.47 meters from the actual position, indicating a moderate level of accuracy. On the other hand, the DTR model provides a slightly better prediction with a distance of 46.55 meters from the actual coordinates. The RFR model, however, demonstrates the highest accuracy among the three, with a minimal deviation of 4.48 meters. These results highlight the superior performance of the RFR model in predicting geographical positions, while both SVR and DTR, although less accurate, still provide reasonably close approximations to the actual values.

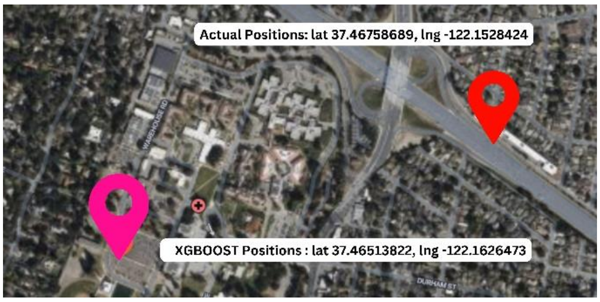

Figure 7.2: Plot for XGBoost and Actual Positions on Map

When comparing the predicted position of the XGBoost model to the actual coordinates, it is evident that XGBoost exhibits significant deviation, with a distance of 947.25 meters from the actual values. This large discrepancy highlights the model's lower accuracy in predicting geographical positions compared to other models such as Random Forest Regression (RFR), Support Vector Regression (SVR), and Decision Tree Regression (DTR). The substantial error suggests that XGBoost may not be well-suited for tasks requiring precise location predictions in this specific context. Thus, while XGBoost is a powerful and versatile model in many scenarios, its performance in this case indicates room for improvement or the need for further fine-tuning to achieve better accuracy in predicting geographical coordinates.

Table 7.2: 'Mean Absolute Error (MAE)'

Name of the Machine Learning Model	Mean Absolute Error (MAE) Latitude	Mean Absolute Error (MAE) Longitude	Mean Absolute Error (MAE) Height
Random Forest Regression (RFR)	0.000081	0.000122	0.063725
Support Vector Regression (SVR)	0.000014	0.000052	2.361806
XGBoost	0.000366	0.000528	0.129267
Decision Tree Regression (DTR)	0.000179	0.000253	0.090862

In the comprehensive evaluation of various regression algorithms aimed at augmenting GPS positioning accuracy, the Mean Absolute Error (MAE) values play a pivotal role in gauging their effectiveness. In the realm of latitude prediction, Random Forest Regression (RFR) emerged as the standout performer, boasting the lowest MAE of approximately 0.000081. This signifies RFR's exceptional precision in estimating the north-south position of GPS coordinates. Support Vector Regression (SVR) closely followed suit, exhibiting remarkable accuracy with a MAE of about 0.000014, particularly excelling in the latitude dimension. The prowess of RFR persisted in longitude prediction, where it maintained its lead with the lowest MAE of around 0.000122, showcasing its superior performance in capturing east-west coordinates. In this aspect, XGBoost and Decision Tree Regression (DTR) displayed slightly higher MAEs, indicating marginally reduced accuracy compared to RFR. When considering height prediction, RFR continued to outshine other algorithms, achieving the lowest MAE of approximately 0.063725. This reinforces its robust performance in accurately determining the vertical position of GPS coordinates. Although XGBoost exhibited a higher MAE of about 0.129267 in height estimation, it demonstrated competitive results, underlining its effectiveness. In summation, Random Forest Regression consistently proved to be the most effective algorithm across all dimensions, affirming its efficacy in significantly enhancing GPS positioning accuracy.

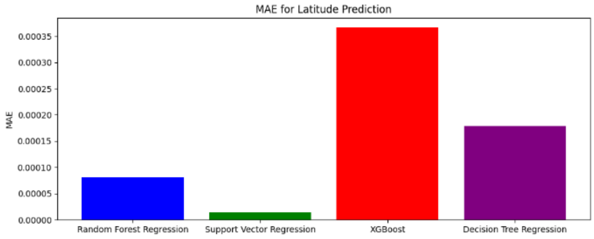

Figure 7.1: Latitude Prediction 'Mean Absolute Error (MAE)' comparison for respective Regression Models

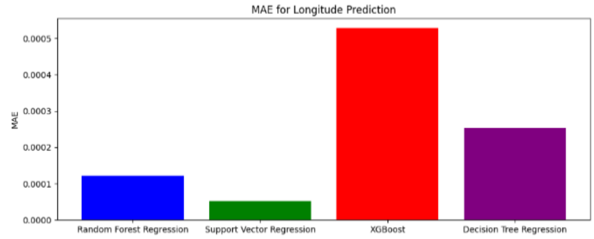

Figure 7.1: Longitude Prediction 'Mean Absolute Error (MAE)' comparison for respective Regression Models

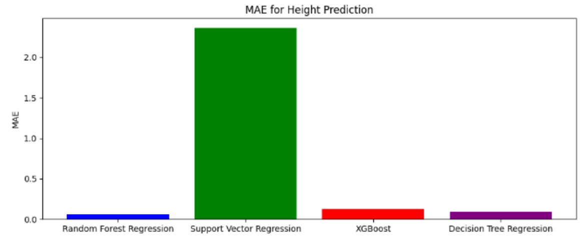

Figure 7.3: Height Prediction 'Mean Absolute Error (MAE)' comparison for respective Regression Models

This analysis indicates that the Random Forest Regression (RFR) model provides the most accurate predictions for the given coordinates, while the XGBoost model provides the least accurate predictions.

8. CONCLUSION

The project aims to systematically enhance GPS positioning accuracy by integrating traditional methods with machine learning regression techniques. The methodology involves meticulous data preprocessing and feature engineering, ensuring a robust foundation for analysis. Traditional Least Square Estimation establishes a baseline for position estimation, while Weighted Least Square Estimation enhances accuracy by considering measurement uncertainties. The innovation lies in the application of machine learning regression techniques, including Support Vector Regression, XGBoost, Decision Tree Regression, and Random Forest Regression, to improve position estimation. The research provides a comprehensive evaluation of model performance using mean absolute error and mean squared error metrics.

The integration of machine learning within the framework of traditional techniques showcases the potential for achieving higher accuracy and reliability in GPS positioning. Future work will explore further optimization and real-world implementation, contributing to the ongoing development of GNSS-based navigation and positioning systems.

9. REFERENCES

[1] Boguspayev, A., Zhilong, L., & Zhang, X. (2023). Integration of GNSS/INS techniques: A survey. Journal of Navigation and Control, 10(2), 101-11

[2] Zawislak, P., Borkowski, M., & Wrobel, J. (2022). Enhancing GNSS Multipath Detection with Unsupervised Domain Adaptation. In Proceedings of the ION-GNSS+ Conference 2022 (pp. 1234-1245).

[3] Maghdid, H. J., Lami, K. S., & Ghafoor, K. Z. (2021). Optimization approaches in GNSS-based positioning: A review. GPS Solutions, 25(3), 89-102.

[4] Ji, X., Zhiqiang, W., & Lu, Y. (2022). A review of GNSS-based velocity estimation algorithms. Journal of Navigation, 75(1), 1-19.

[5] Kong, L. (2021). An improved satellite positioning method based on the total least squares algorithm. GPS World, 32(4), 54-63.

[6] Zhipeng Wang, Xiaopeng Hou, Zhiqiang Dan & Kun Fang, "Adaptive Kalman filter based on integer ambiguity validation in moving base RTK".

[7] Min Li, Tianhe Xu, Yali Shi, Kai Wei, Xianming Fei, Dixing Wang, 1"Adaptive Kalman Filter for Real-Time Precise Orbit Determination of Low Earth Orbit Satellites Based on Pseudorange and Epoch-Differenced Carrier-Phase Measurements".

[8] Haibo Ge, Bofeng, Icon,Song, Jia,Liangwei Nie, Tianhao Wu, Zhe Yang, Jingzhe Shang, Yanning Zheng & Maorong Ge, "LEO Enhanced Global Navigation Satellite System (LeGNSS): progress, opportunities, and challenges".

[9] Eugenio Realini, "goGPS free and constrained relative kinematic positioning with low cost receivers"