Machine Learning Based Overbound Modeling of Multipath Error for Safety Critical Urban Environment

Heekwon No, Carl Milner

Machine Learning Based Overbound Modeling of Multipath Error for Safety Critical Urban Environment

Heekwon No, ENAC, Université de Toulouse, France
Carl Milner, ENAC, Université de Toulouse, France

BIOGRAPHY (IES)

Heekwon No is a postdoctoral researcher in the Ecole Nationale de l'Aviation Civile (ENAC), France. He received B.S. and Ph.D. degrees from the School of Mechanical and Aerospace Engineering at Seoul National University, South Korea. His research interests are the multi-sensor fusion based navigation, guidance and control system of the unmanned aerial or ground vehicle.

Dr Carl Milner is an Assistant Professor within the Telecom Lab at the Ecole Nationale de l'Aviation Civile. He has a masters degree in Mathematics from the University of Warwick, a PhD in Geomatics from Imperial College London and has completed the graduate trainee programme at the European Space Agency. His research interests include GNSS augmentation systems, integrity monitoring, air navigation and applied mathematics.

ABSTRACT

In the urban environment, multipath and non-line of-sight are the critical source of measurement errors and signal power loss. In urban canyons, whilst the user can still acquire the required number of satellites to obtain a position, thanks to multi-constellation GNSS, such signals may be subject to gross multipath errors and lead to a potentially unsafe position. In this paper, machine learning techniques are used to model the multipath error distributions. The set of features which have been assessed are commonly used parameters such as the elevation, S/N, and user speed. For modeling and evaluation of the model validity, a large number of hours of experimental data has been collected by driving a sensor-equipped vehicle in the urban area in Toulouse. Considering the processing of data from single-frequency type GNSS receiver, the multipath error component is extracted from measurement using appropriate techniques (measurement differential, clock bias estimation, etc.). Quantile of multipath error are modeled using neural network-based regression technique with the features. Modeling results using the proposed method are validated by an integrity assessment of the experimental data.

INTRODUCTION

The CLUG Project - meaning "Certifiable Localisation Unit with GNSS" is a 2-year project building on the use of GNSS coupled with other sensors (such as IMU and odometer) to provide a continuous and accurate train localization that could be integrated in the future European Rail Traffic Management System (ERTMS). ENAC is taking a role of contributing to local GNSS error modelling and integrity in this project. The use of GNSS as the main localization sensor for safety critical positioning of trains would substitute the costly current trackside equipment. However, the highly stringent requirements [1] for safety are seen as a major hurdle, especially for use in urban and suburban areas.

In the urban environment, multipath and non-line of-sight are the critical source of measurement errors and signal power loss. In urban canyons, whilst the user can still acquire the required number of satellites to obtain a position, thanks to multi-constellation GNSS, such signals may be subject to gross multipath errors and lead to a potentially unsafe position. There have been attempts based on the use of array antennas [2] or ray tracing technique [3] to estimate and mitigate the error. However, they are still economically and computationally expensive. If we cannot easily mitigate such errors, then a reliable and trustworthy model is needed to determine their effects on the user position error in a statistical manner and allow a complete safety analysis.

The modeling of the multipath error is not simple, due to the complex involvement of many factors including the geometrical parameter, user dynamics and environment characteristics. Moreover, its distribution commonly has a non-zero mean, large variance

and a positive skew in many situations. In order to overbound the distribution, simplified approaches have been used in the past merging data in all 'urban' environments and using often single driving parameters (S/N) [4]. Furthermore, this approach often ignores available knowledge of the environment that could reasonably be used to improve the estimation of the distribution such as the local mask.

In this paper, machine learning techniques are used to model the multipath error distributions. The machine learning techniques are used to help select the choice of features which provide the best, tightest model of the distributions. The set of features which have been assessed are commonly used parameters such as the elevation, S/N, user speed as well as features derived from topographic data.

For modeling and evaluation of the model validity, a large number of hours of experimental data has been collected by driving a sensor-equipped vehicle in the urban area in Toulouse. Processing of data from single-frequency type GNSS receiver is considered. The multipath error component is extracted from measurement using appropriate techniques (measurement differential, clock bias estimation, etc.). Quantile of multipath error are modeled using neural network-based regression technique with various features. Modeling results using the proposed method are validated by an integrity assessment of the experimental data.

MODELING METHODOLOGY

Figure 1 shows the proposed scheme of multipath modeling. In the interested area, GNSS and IMU measurements are collected and MPN (multipath and noise) is extracted from the collected data. Using the machine learning based quantile regression technique, model parameters for MPN quantile are determined. Various information can be used as features for the quantile modeling. If a topographic database is used during the modeling process, it can be stored for user-side use. After the modeling is done, user can use these stored model parameters to estimate the error quantile of live measurements and can use these quantiles in its navigation engine to enhance its functionality.

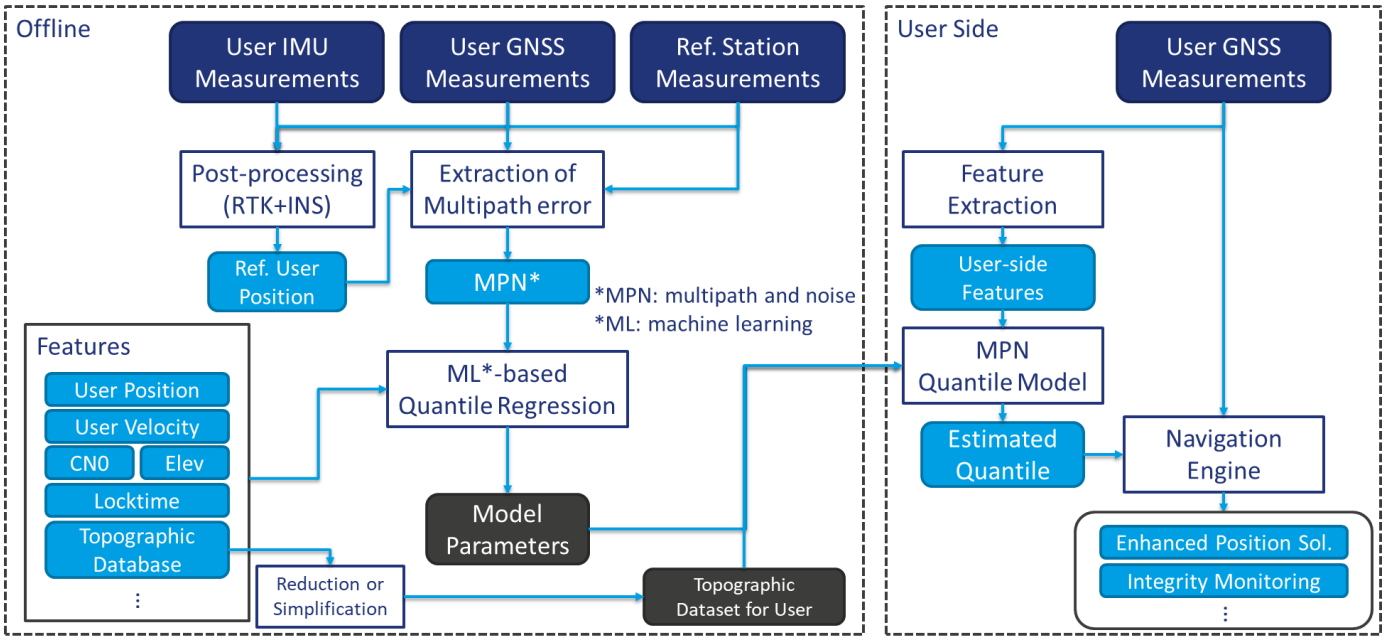


Figure 1 Proposed scheme of multipath modeling

EXTRACTION OF SINGLE-FREQUENCY MULTIPATH

Figure 2 shows the detailed process of the extraction of the single frequency multipath. The reference position is obtained by a post-processing GNSS/INS integration with sufficient accuracy to model the MPN. True range from satellite is calculated and is subtracted from the user and reference station measurements. By differencing the values from user side and reference station side, most of errors

are eliminated but user multipath and receiver clock bias remain. After removing estimated receiver clock bias, mpn is finally obtained.

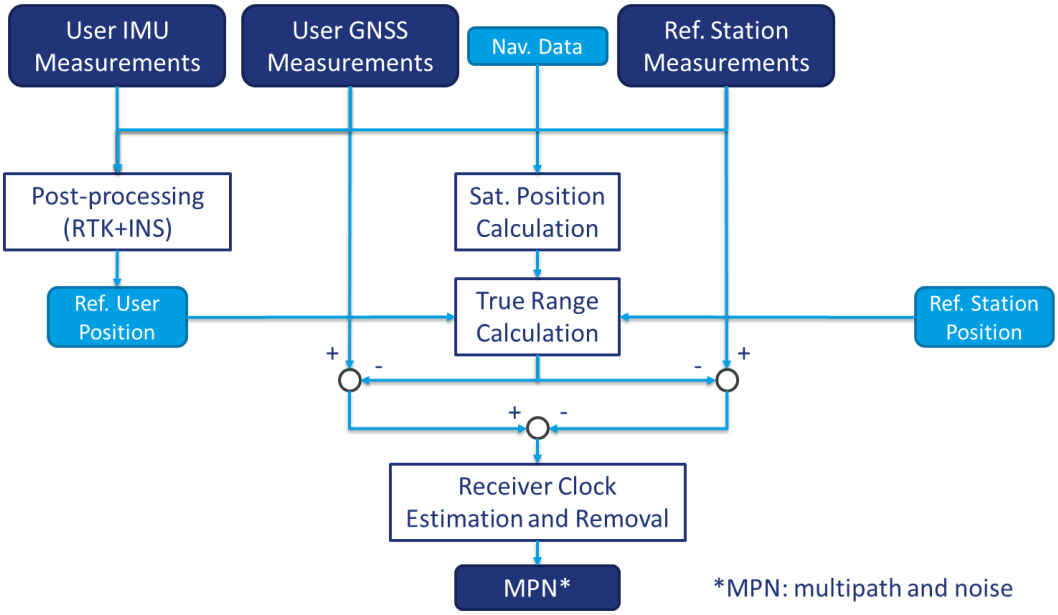


Figure 2 Process of extraction of the single frequency multipath

Detailed equations for the processing are provided here. User side pseudorange measurement is given by

$$\rho_u^i = r_u^i + B_u - b^i + I_u^i + T_u^i + \varepsilon_u^i \tag{1}$$

where the superscript i is a satellite index, the subscript u and r mean user and reference station, respectively, r_u^i is the geometrical range between user and i-th satellite, B_u is the clock error of user receiver, b^i is the clock error of i-th satellite, I_u^i is the ionospheric delay error of i-th satellite on user position, T_u^i is the tropospheric delay error of i-th satellite on user position, ε_u^i is the MPN (multipath + noise) of i-th satellite on user position and $\delta(x)$ is the estimation error of x.
With known user position obtained from the post-processing, true range from user position to i-th satellite is given by

$$\hat{r}_u^i = \left|\hat{r}^i - \hat{r}_u\right| \tag{2}$$

True ranges of user side and reference station side are subtracted from user and reference station pseudorange measurement, respectively. Then user and reference station side residual errors are obtained as

$$\rho_u^i - \hat{r}_u^i = \delta r_u^i + B_u - b^i + I_u^i + T_u^i + \varepsilon_u^i \tag{3}$$

$$\rho_r^i - \hat{r}_r^i = \delta r_r^i + B_r - b^i + I_r^i + T_r^i + \varepsilon_r^i \tag{4}$$

Differencing residual errors between user and reference station, the corrected user range error is obtained as

$$\left(\rho_u^i - \hat{r}_u^i\right) - \left(\rho_r^i - \hat{r}_r^i\right) = \left(\delta r_u^i - \delta r_r^i\right) + \left(B_u - B_r\right) + \left(I_u^i - I_r^i\right) + \left(T_u^i - T_r^i\right) + \varepsilon_u^i - \varepsilon_r^i \tag{5}$$

Assuming that the orbital, ionospheric, tropospheric error residuals, user reference position error and reference station MPN are small compared to user MPN, the corrected user range error can be written as

$$(\rho_u^i - \hat{r}_u^i) - (\rho_r^i - \hat{r}_r^i) \approx (B_u - B_r) + \varepsilon_u^i = \Delta B + \varepsilon_u^i \tag{6}$$

After the correction, the corrected range error contains user MPN and receiver clock bias. The receiver clock bias should be removed to obtain the user MPN. This receiver clock bias is estimated by a Kalman filter based on simple kinematic model. Simple kinematic model for time propagation is given by

$$\dot{x} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} x + w \text{ where } x = \begin{bmatrix} \Delta B \\ \Delta \dot{B} \end{bmatrix} \tag{7}$$

The corrected range error and doppler errors are used as measurements for the Kalman filter as follows:

$$z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x + v \text{ where } z = \begin{bmatrix} \Delta B + \varepsilon_u^i \\ \Delta \dot{B} + \dot{\varepsilon}_u^i \end{bmatrix} = \begin{bmatrix} (\rho_u^i - \hat{r}_u^i) - (\rho_r^i - \hat{r}_r^i) \\ (\dot{\rho}_u^i - \hat{\dot{r}}_u^i) - (\dot{\rho}_r^i - \hat{\dot{r}}_r^i) \end{bmatrix} \text{ for i} = 1,2, \dots, \text{m} \quad (8)$$

For satellites with good condition (CN0>40dB, lock time>5sec, etc.), multipath is negligible and noise is the dominating component of MPN (ε_u^i) and its rate ($\dot{\varepsilon}_u^i$).

Assuming the estimation error of receiver clock is small compared to user multipath, the user MPN is obtained by removing estimated receiver clock error as follows:

$$\left(\rho_u^i - \hat{r}_u^i\right) - \left(\rho_r^i - \hat{r}_r^i\right) - \Delta\hat{B} \approx \delta B + \varepsilon_u^i \approx \varepsilon_u^i \quad (9)$$

QUANTILE-BASED OVERBOUNDING

MPN error distribution of train user under open-sky is usually nicely shaped to regard it as a Gaussian distribution. In case of train user in urban area, it could be different. In faulty case or under the effect of severe multipath, the distribution of error can be biased, non-symmetric, and may have a heavy tail as show in Figure 3. To model this type of error, the Gaussian overbounding model is widely used. It models the error distribution by determining mean and variance of a Gaussian distribution which can bound the actual error up to desired probability as shown in Figure 4.

In this paper, a new modeling method, the quantile based overbounding method is proposed. This method models the error distribution by converting quantile of desired probability to variance of zero-mean Gaussian distribution as shown in Figure 5. This quantile of error is determined from the machine learning technique. The obtained quantile value can be converted to the model variance using the following relation:

$$Q_p = k_p \cdot \sigma_p \quad (10)$$

$$Q_p = F^{-1}(p) = inf\{y: F(y) \geq p\} \text{ where } F(y): \text{cumulative distribution function} \quad (11)$$

$$k_p = \overline{F^{-1}}\,(p): \text{inverse cumulative distribution function of normal distribution} \quad (12)$$

This model can naturally overbound the error distribution by the definition of quantile. Especially, it is useful for cases where the separate determination of mean and confidence bound of error is difficult. In case of multipath, it is computationally expensive to calculate the exact multipath (e.g., the ray-tracing technique). Moreover, the complex tracking loop dynamics is involved to its distribution in case of moving user. Therefore, this quantile-based modeling approach fits well for the modeling of multipath error.

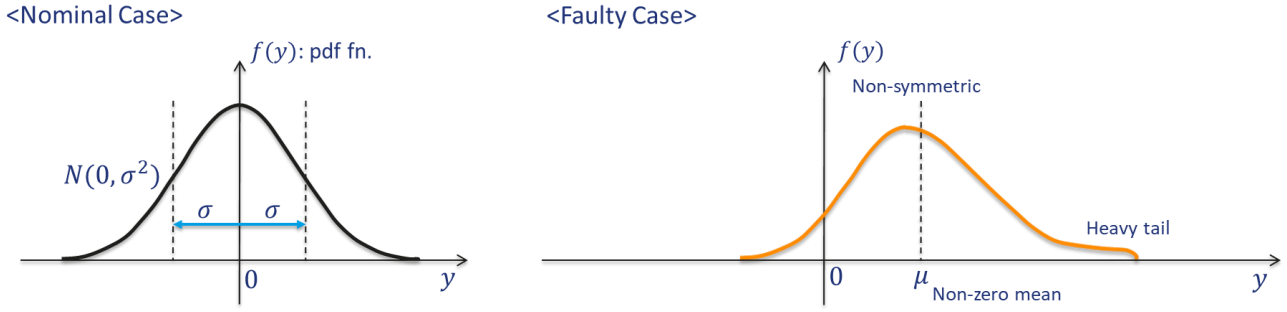


Figure 3 Distribution of error in normal (left) and faulty (right) case

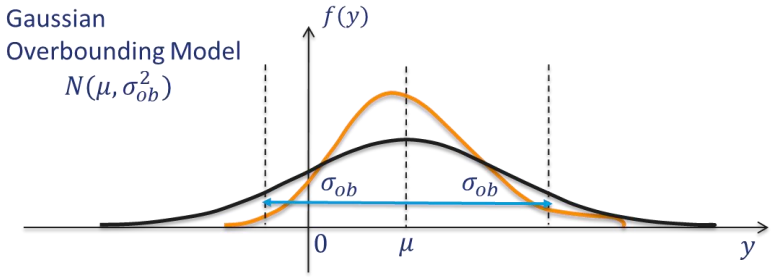


Figure 4 Gaussian overbounding model

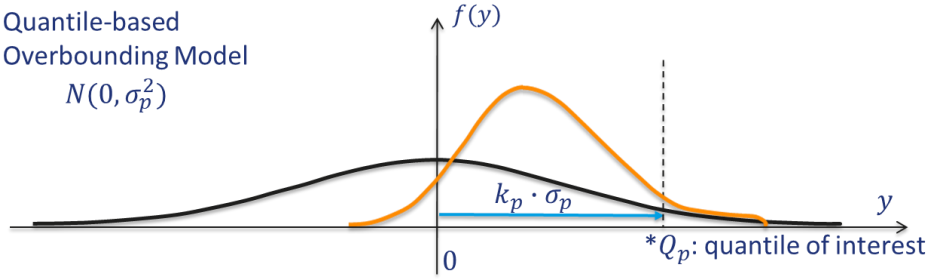


Figure 5 Quantile-based overbounding model

MACHINE LEARNING-BASED QUANTILE REGRESSION

The neural network is one of the machine learning technique. Each neuron is composed of weights and bias, a layer consists of several number of neurons, and several number of layers become a neural network containing all the weights and biases of neurons (Figure 6). These parameters are optimized to minimize cost function to achieve its goal.

< Output of a neuron >

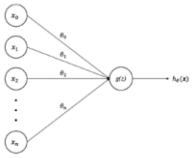
$h_\theta(x) = g(\theta^T x + b)$ where

x: Input vector

θ: Weight vector

b: Bias

g: Activation function



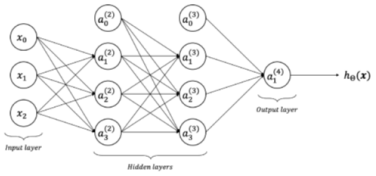
< Output of a layer, l >

$h^{(l)}(x) = g\left(\theta^{(l)} x + b^{(l)}\right)$ where

x : Input vector

$\theta^{(l)}$: Weight matrix of l-th layer

$b^{(l)}$: Bias vector of l-th layer



< Output of a NN >

In case of three layers, $h_\theta(x) = h^{(3)}\left(h^{(2)}\left(h^{(1)}(x)\right)\right)$

Figure 6 Structure of neural network

The proposed cost function for quantile regression is as follows:

$$\mathcal{L}(\boldsymbol{\xi}|q,\lambda) = \sum_{i=1}^{i=n} \mathcal{L}(\xi_i|q) + \frac{\lambda}{2n} \sum_{l=1}^{L} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} \left(\theta_{j,i}^{(l)}\right)^2 \tag{13}$$

$$\mathcal{L}(\xi_i|q) = \begin{cases} q\xi_i & if\ \xi_i \geq 0 \\ (q-1)\xi_i & if\ \xi_i < 0 \end{cases} \tag{14}$$

$$\xi_i = y_i - h_\theta(x_i) \tag{15}$$

where q is the desired quantile, λ is the regularization factor, L is the number of layers, s_l is the number of units of layer l, and n is the number of samples. Generally, the squared sum error is involved in cost function. However, in case of the quantile regression, the pinball function, (14) is used [5]. The regularization part prevents the neural network from the over-fitting problem.

After the modeling is done, how to validate the obtained MPN model? In this paper, three methods are proposed. Firstly, the overbounding of error is checked by Q-Q plot after normalization using proposed model. Secondly, the accuracy of weighted least square position solution using proposed model is checked. Thirdly, the fault detection and protection level (PL) calculation results using the proposed model are checked based on sloped-based RAIM (receiver autonomous integrity monitoring).

DATA COLLECTION AND ANALYSIS

Figure 7 shows the test car used in the data collection and the equipped sensors in the test car. The test car is equipped with a Novatel GNSS receiver and a tactical grade IMU. The collected data was processed through Novatel Waypoint Inertial Explorer software package to obtain the reference position of user.

Figure 7 Test car for deta collection (left) and equipped sensor (right)

The data was collected in five areas around Toulouse (Figure 8). Total 45 hours of data was collected. All user trajectory is within 6km boundary from the reference station, TLSE. Figure 9 shows the environment of each area. In St. Orens area, there is no high buildings but warehouses with metal plate wall are taking places around the area. In the other four areas, buildings up to 5 floors high are on both sides of route.

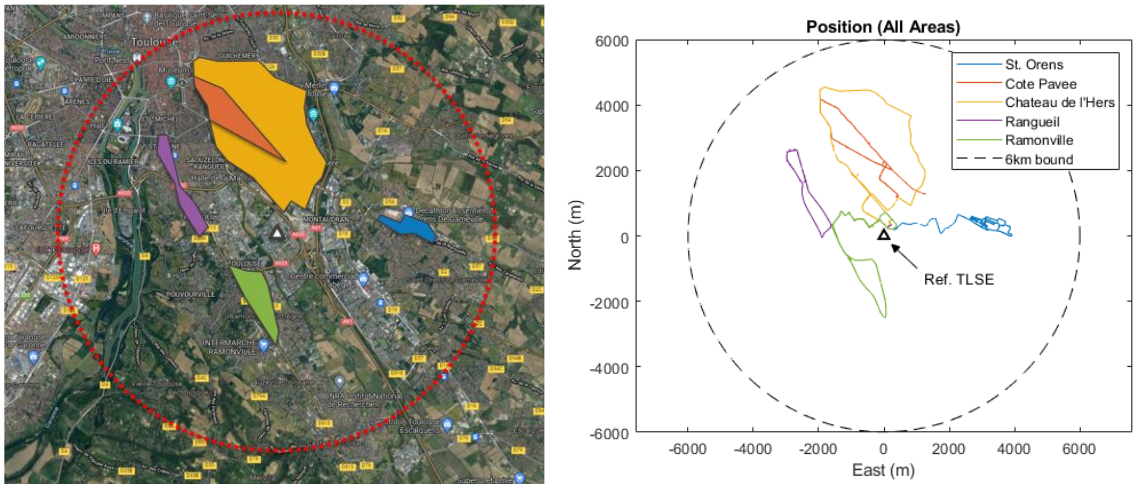


Figure 8 Area of Data Collection (Toulouse, France)

Figure 9 Environment of each area

Figure 10 shows MPN vs. user CN0 relation for GPS and Galileo constellation. It is obvious that two constellations are showing different level of MPN because of the different signal characteristics. Their ranges of CN0 are not matching each other. From these facts, individual modeling for GPS and Galileo is desirable. Figure 11 shows the relation between elevation angle and MPN. It also shows certain amount of correlation. Figure 12 shows the relation between user speed and MPN. Outer boundary for moving state is noticeable pretty clearly. As speed decreases, the boundary gets higher, and when user stops, the MPN becomes substantial. Figure 13 shows the relation between locktime and MPN for static and dynamic states separately. The condition for the static state is the user speed less than 0.3m/s. In the dynamic state, the MPN doesn't last for long time. However, in the static state, many cases of continuous large MPN exist. From these facts, it could be considered to have two separate models for the static and dynamic states.

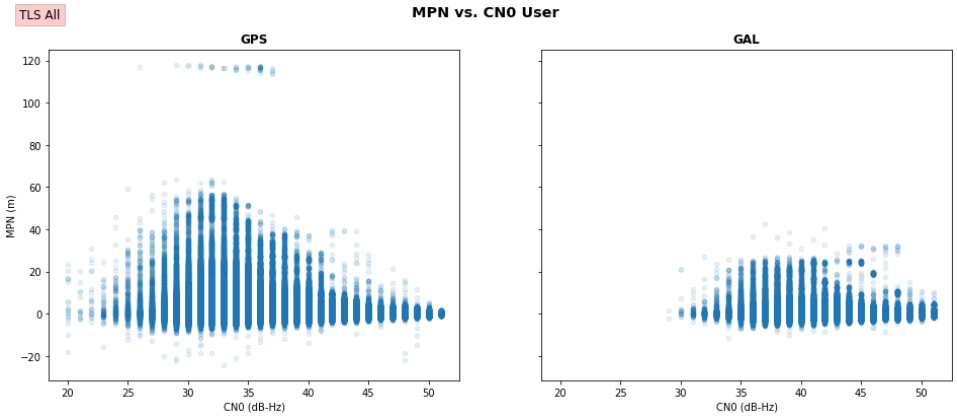


Figure 10 Multipath error vs. user CN0

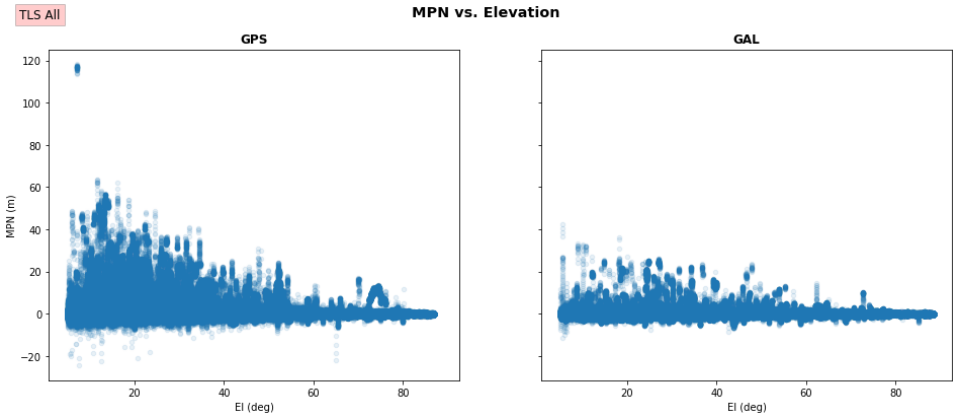


Figure 11 Multipath error vs. elevation angle of satellite

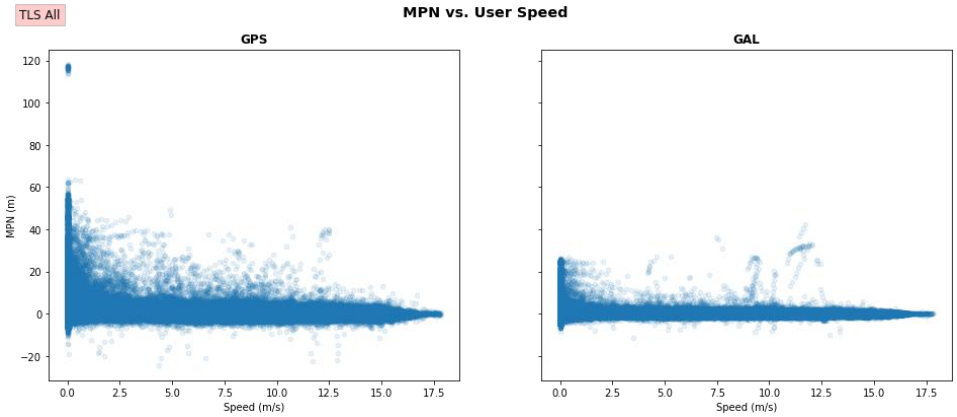


Figure 12 Multipath error vs. user speed

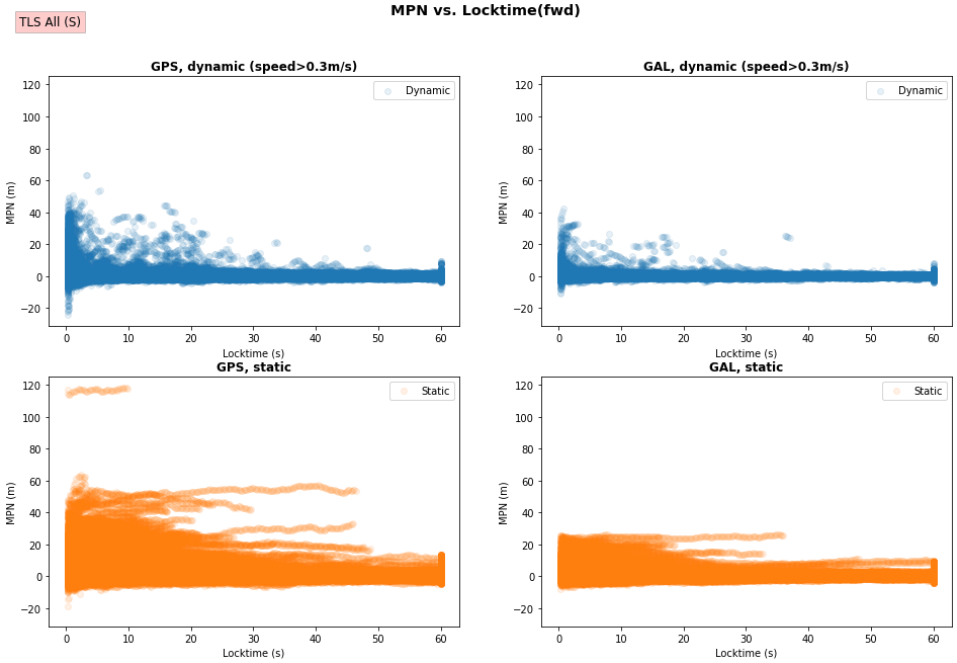
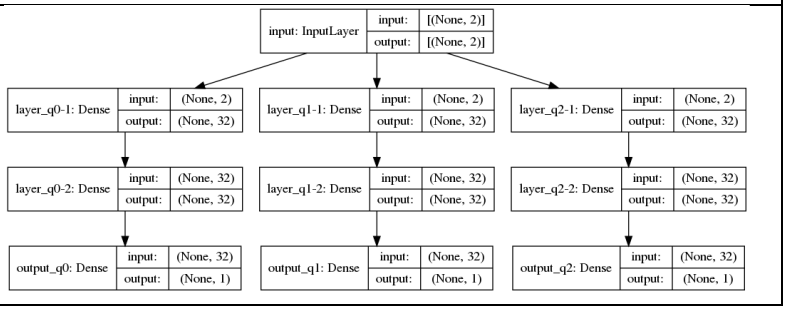


Figure 13 Multipath error vs. locktime for static (up) and dynamic (down) state

MODELING RESULTS

Various structures of neural network model were tested by changing number of units per layer, types of activation function and order of features. In this paper, a well-balanced neural network structure between complexity and computation load for MPN quantile modeling is proposed in Table 1.

Table 1 Proposed neural network model parameters and structure

Model Parameters	Model Structure
Estimated quantiles: 0.95, 0.99, 0.999 Number of hidden layers: 2 layers Number of units per layer: 32 Activation function: ReLU	

In this structure, quantiles of three different desired probabilities are modeled in the model. The number of quantiles or desired probabilities can be changed according to their application. Four different models were considered according to feature types and user state separation. First model – named "SF CN0" is taking CN0 as a single feature. Second model – named "SF ELE" is taking elevation as a single feature. Third model – named "MF ALL" is taking both of CN0 and Elevation angle as multiple features. The last fourth model – named "MF SDS" is same as the third model but it contains two separate models for the static and dynamic states of user.

Figure 14 shows SF CN0 model. The model is drawn up to the point where enough number of samples are collected for desired quantile probability. As expected, Galileo shows lower quantile value than GPS overall. The higher MPN quantile is expected for lower CN0. Figure 15 shows SF ELE model. It is interesting that, for Galileo, the elevation angle of 25 deg has the highest expected 0.999 probability quantile. This came from severe multipath of the region in collected data set and is showing the dependency of model to training data set in case of extreme desired probability. To avoid this, a sufficiently larger data set is required. Figure 16 shows the MF ALL model. Higher MPN quantile is expected for lower elevation angle and CN0. Figure 17 shows MF SDS model, two separate MF models for static and dynamic state of user. Lower MPN quantile is expected in the dynamic user state than in the static state.

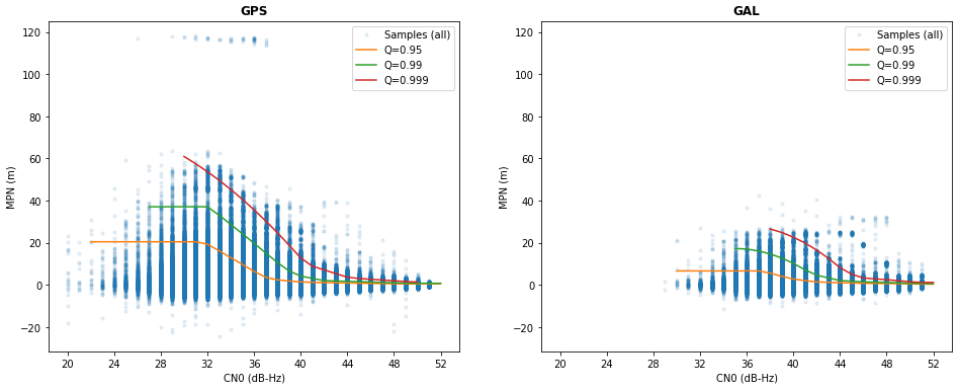


Figure 14 Multipath error modeling results with single feature (CN0) model

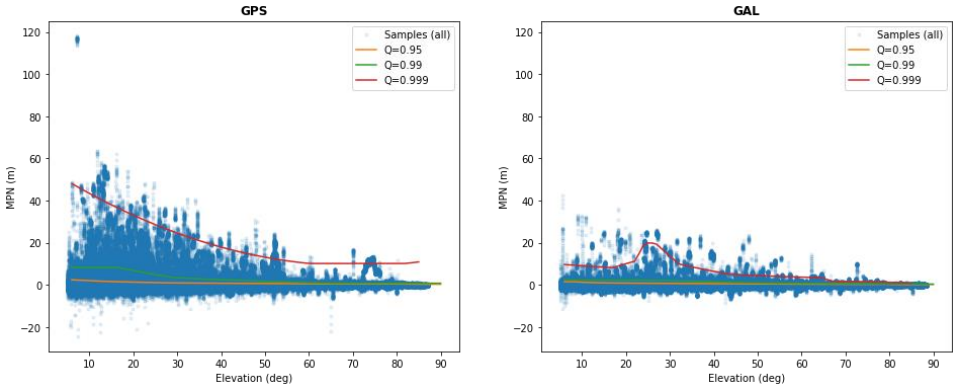


Figure 15 Multipath error modeling results with single feature (elevation) model

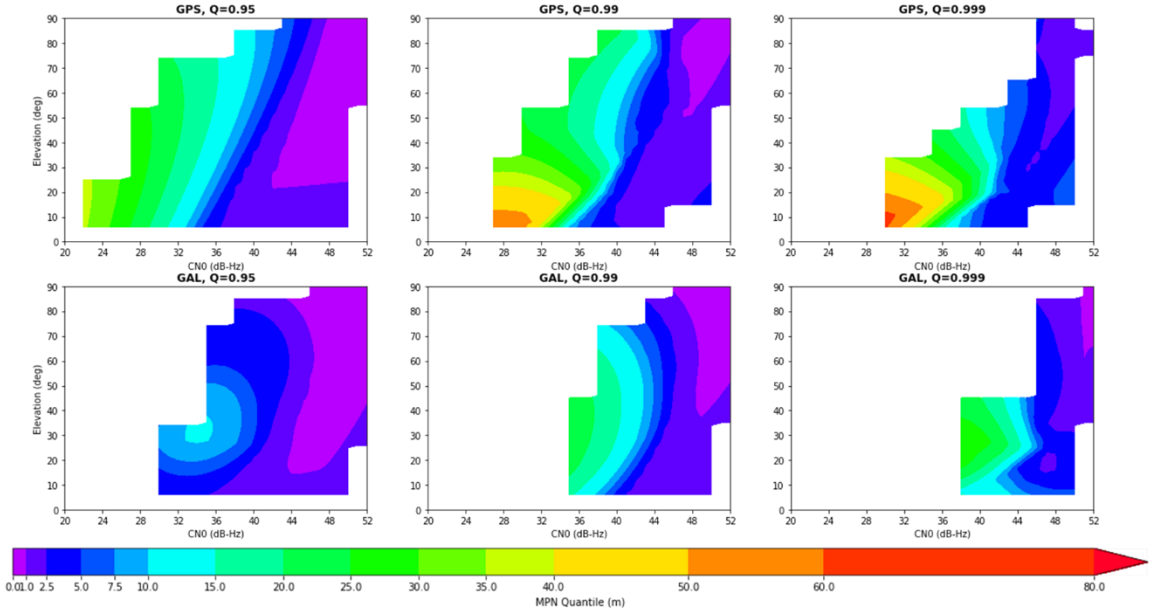


Figure 16 Multipath error modeling results with multiple feature (CN0 and elevation) model

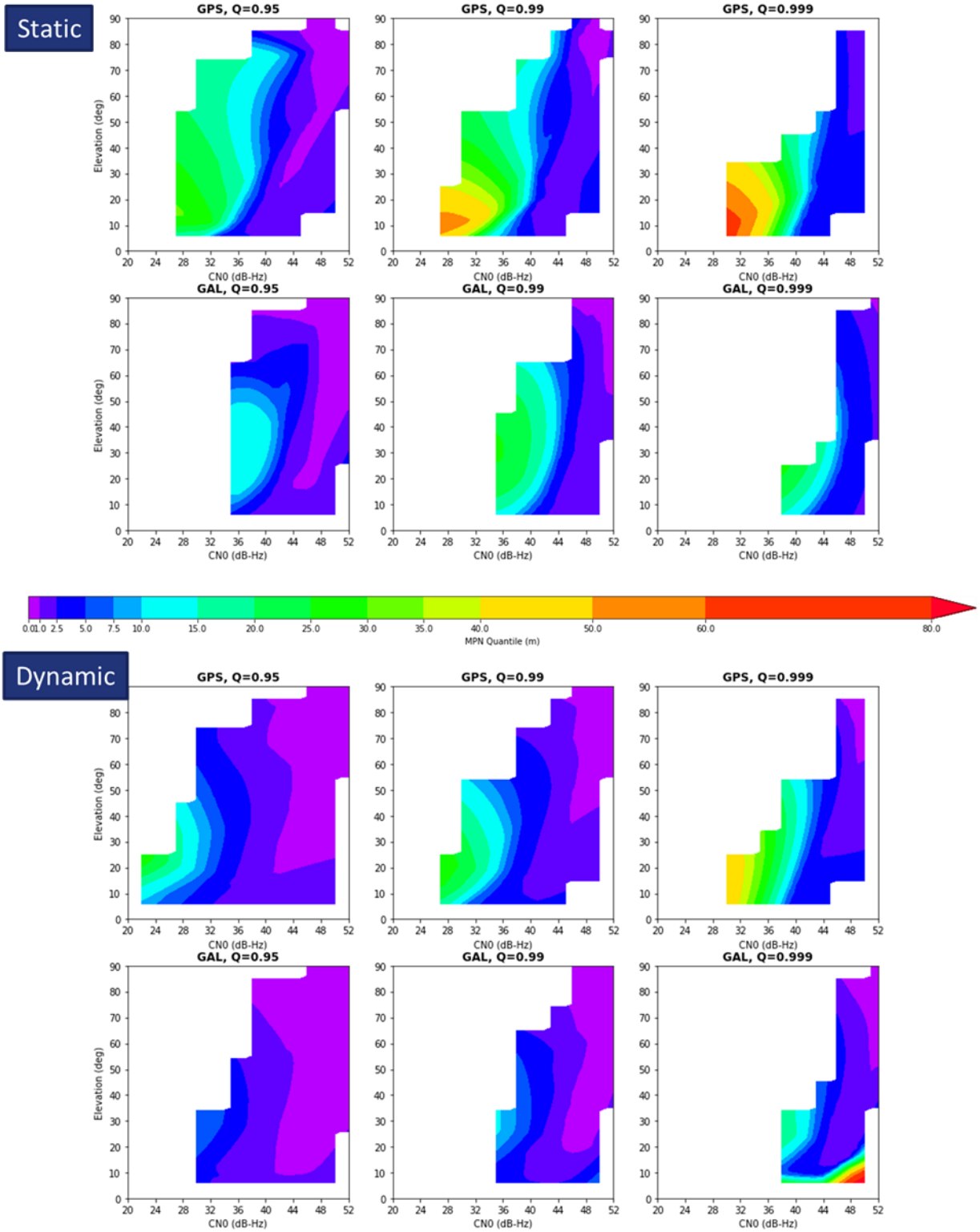


Figure 17 Multipath error modeling results with multiple feature (CN0 and elevation) model for static (above) and dynamic (bottom) state

MODEL VALIDATION

Firstly, the error overbounding results of proposed models are checked by Q-Q plot of normalized MPN in Figure 18. If the curve overpasses the crossing point between the normal curve and desired probability, it means the model is conservative. If the curve underpasses, the model is failed to overbound the error. For GPS, all models are successfully overbounding the MPN. Among the models, the most sophisticated, MF SDS model (the red curve) looks the closest to exact bound. The results for Galileo look similar to those of GPS for 0.95 and 0.99 probabilities. However, for 0.999 probability, the MF SDS is the only model can overbound up to desired probability. The other models are failed to overbound.

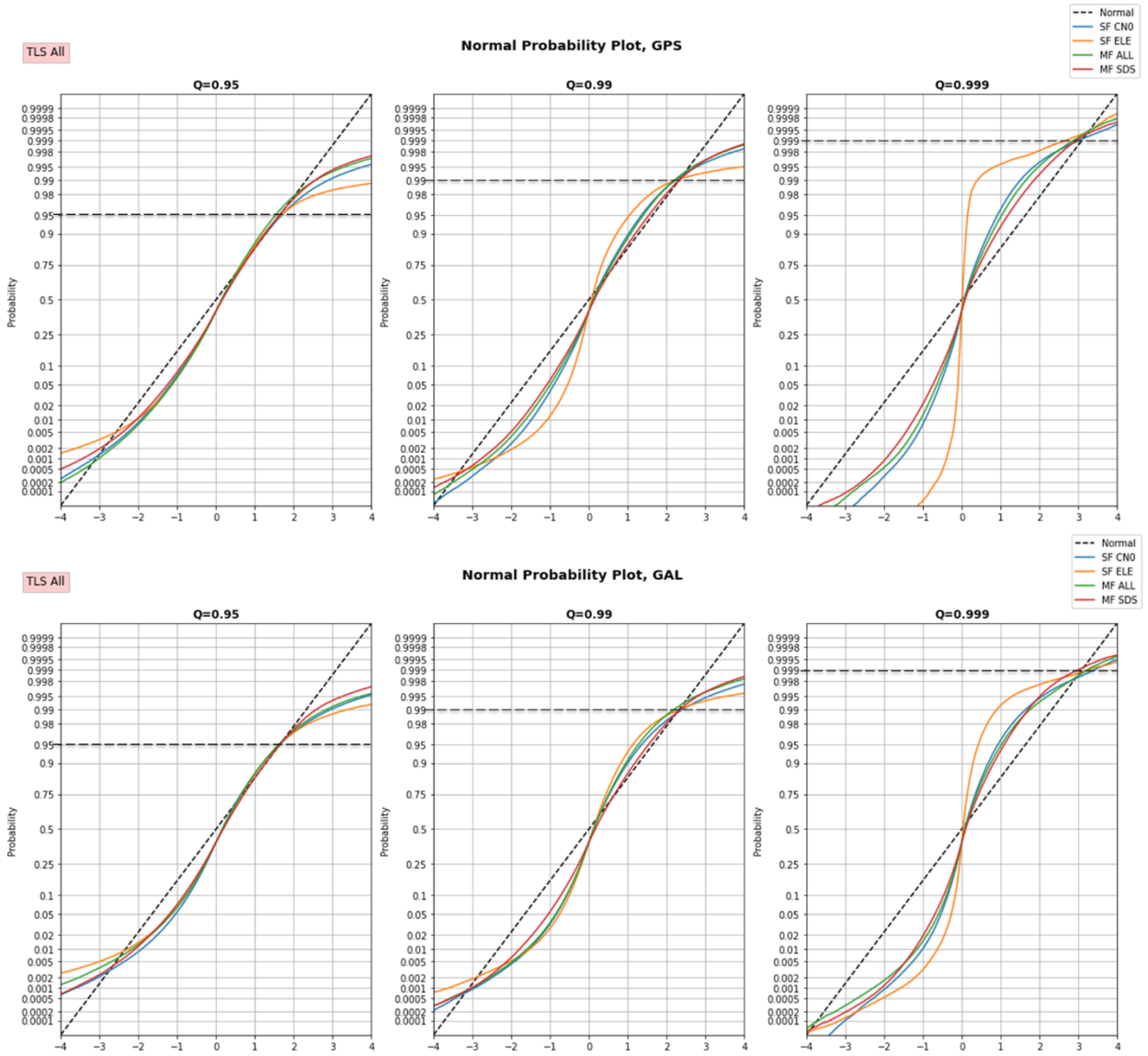


Figure 18 Q-Q plot of normalized multipath error for GPS (above) and Galileo (bottom)

The other validation methods are the comparison of position solution accuracy and slope-based RAIM results. Among the five data collection areas, St. Orens and Ramonville areas will be discussed in this paper. In the validation results of model, the WLS (weighted least square) position solutions of MPN models are compared with the LS (least square) position solution and the WLS of GBAS error model [6]. For the RAIM based validation, empirical standard deviation under open-sky condition (approximately 0.6m) was used for LS, and 0.99 quantile value was used for proposed MPN models. Keep in mind that the LS position solution and GBAS model based WLS position solution are not considering the effect multipath in pseudorange measurement.

The results of St. Orens area will be discussed first. The position solution results are shown in Figure 19. The proposed MPN models are showing better position accuracy than LS and GBAS model based WLS. Among them, the MF SDS model is showing the smallest RMS error. The slope-based RAIM results is shown in Figure 20. The LS and GBAS WLS are showing higher alarm rate than the MPN models because they are not considering multipath error. All results have no epochs of MI (misleading information). The MF SDS model is providing the tightest PL because it is the closest model to the exact overbound. Figure 21 shows the CDF (cumulative distribution function) of PL (protection level) and position error in nominal operation condition. The GBAS WLS looks providing the tightest PL but its availability is limited because of its high alarm rate. Its position error is bigger than the MPN models. The MF SDS model is showing the tightest PL and the smallest position error among the MPN models.
Next, the results of Ramonville area will be shown in Figure 22, Figure 23 and Figure 24. Since the Ramonville area is denser urban area than St. Orens. area, the overall position error level is higher. However, the MPN models are still showing quite good position accuracy. Since Ramonville area has more severe multipath error then St. Orens area, the alarm rate of GBAS WLS is significant. It also has several epochs of MI. Among MPN models, the MF SDS model are still showing small alarm rate and tight PL. The overall level of PL and PE of this area are higher than St. Orens area. The availability and position error of GBAS model is severely degraded. The MF SDS model is still showing the best CDF of PL and position error.

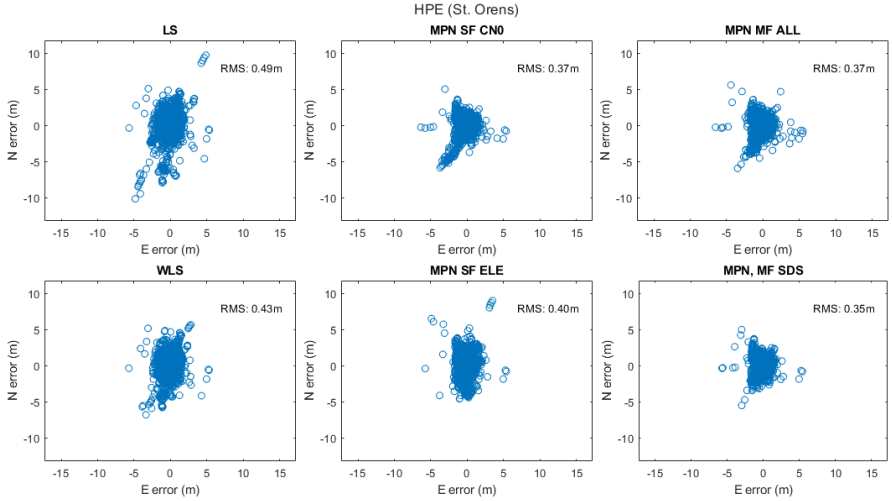


Figure 19 Position solution results for St. Orens area

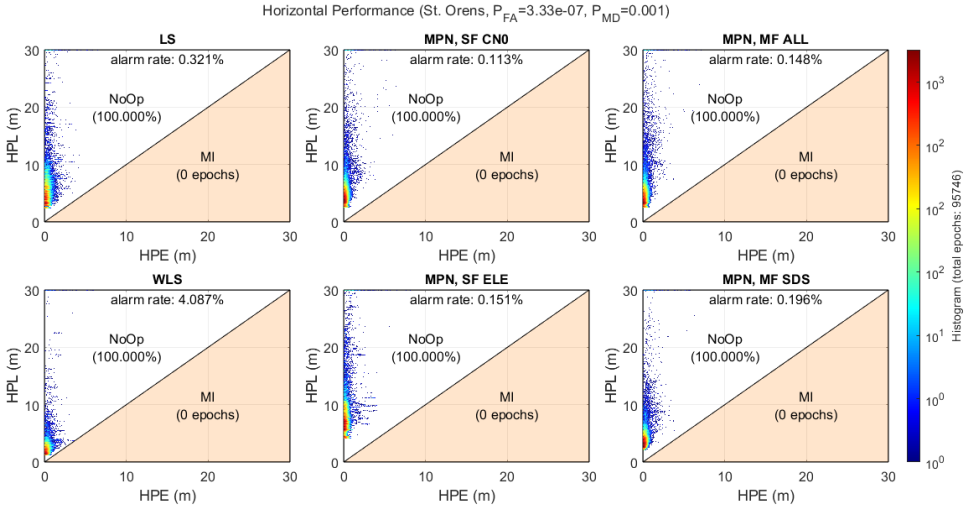


Figure 20 Fault detection and protection level calculation results for St. Orens area

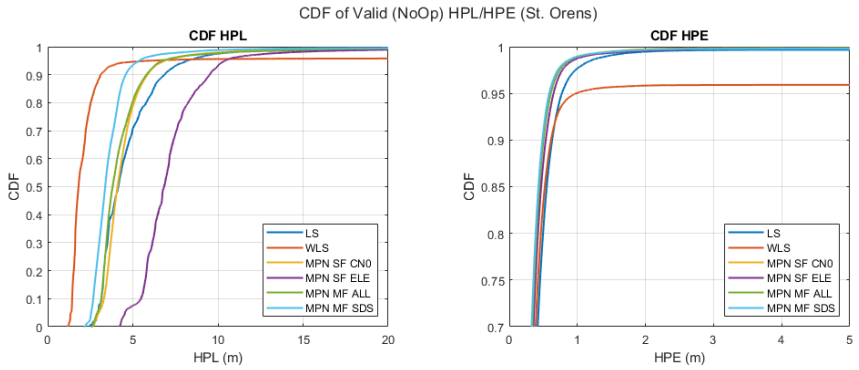


Figure 21 Cumulative density function results of protection level and position error for St. Orens area

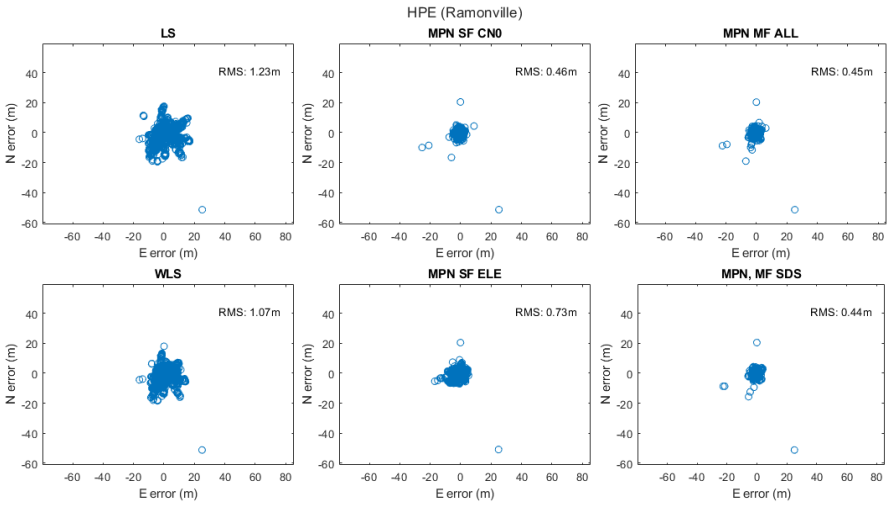


Figure 22 Position solution results for Ramonville area

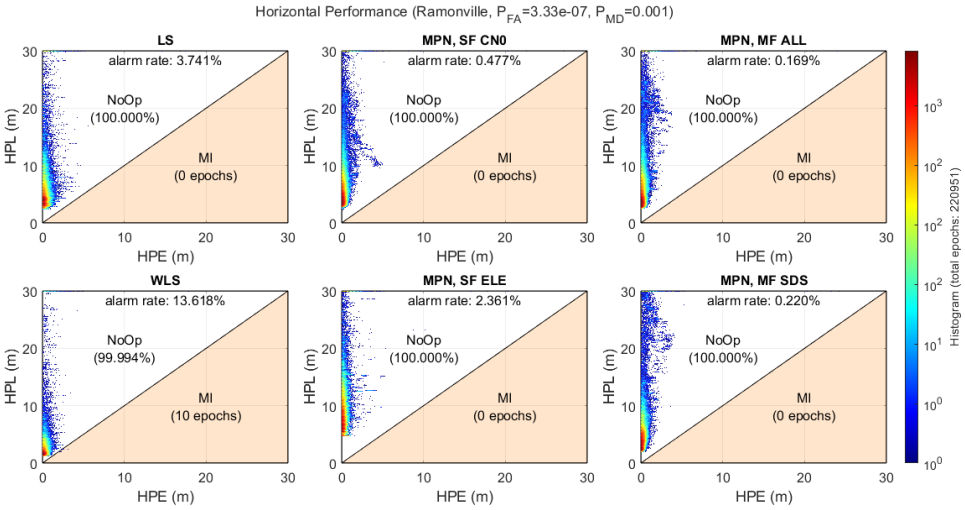


Figure 23 Fault detection and protection level calculation results for Ramonville area

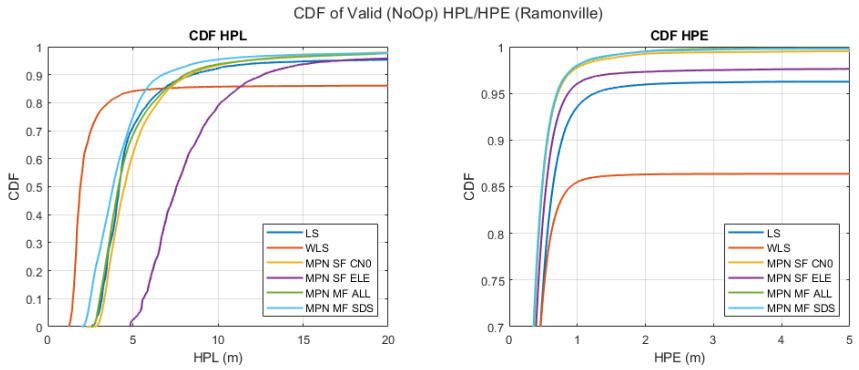


Figure 24 Cumulative density function results of protection level and position error for Ramonville area

CONCLUSIONS

In this paper, the pseudorange multipath error of urban area is collected and analyzed. The correlations between MPN and features (CN0, elevation angle and user speed) are shown. The machine learning based modeling method is proposed for multipath error overbounding. The quantile-based approach overbounds the error up to desired probability by setting quantile probability. The entire modeling scheme and optimized neural network model structure are proposed. The proposed model is well overbounding up to desired probability. It also achieved better position accuracy, lower alarm rate and tighter protection level.

ACKNOWLEDGMENTS

Authors thank the CLUG partners for their support and their work on the CLUG project, and then authors especially thank EUSPA as they are the funder of this project of H2020 program.

REFERENCES

1. UNISIG, “Safety Requirements for the Technical Interoperability of ETCS in Levels 1 & 2 (SUBSET-091),” 2015.
2. Daneshmand, S., Broumandan, A., Sokhandan, N. and Lachapelle, G., “GNSS Multipath Mitigation with a Moving Antenna Array,” *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 49, No. 1, 2013, pp. 693-698.
3. Miura, S., Hsu, L., Chen, F. and Kamijo, S., “GPS Error Correction with Pseudorange Evaluation Using Three-Dimensional Maps,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 6, 2015, pp. 3104-3115.
4. Nobuaki, K., Kobayashi, K., and Furukawa, R. “GNSS multipath detection using continuous time-series C/N0,” *Sensors*, Vol. 20, No. 14, 2020, p. 4059.
5. Takeuchi, I., Le, Q. V., Sears, T. D., Smola, A. J., “Nonparametric Quantile Estimation,” *Journal of Machine Learning Research*, Vol. 7, No. 45, 2006, pp. 1231-1264.
6. RTCA, “Minimum Aviation System Performance Standards for the Local Area Augmentation System (LAAS),” *RTCA document DO-245A*, 2004.