Parallel Deep Quantile Estimation with Gaussian Overbounding for GNSS Multipath Modeling

Florian Roessl and Omar García Crespillo
German Aerospace Center (DLR)

BIOGRAPHY

Florian Rößl received his Master in Geodesy and Geoinformation from the Technical University Munich in 2022. Since then he is working as a scientific employee at the Navigation Department of the German Aerospace Center (DLR). His current field of research includes robust error modeling for safety critical navigation.

Omar García Crespillo is currently Group Leader at the Navigation department of the German Aerospace Center (DLR) where he leads activities in GNSS, inertial sensors, integrated navigation systems and integrity monitoring for safe ground and air transportation systems. He received the Ph.D. degree in Robotics, Control and Intelligent Systems from the Swiss Federal Institute of Technology Lausanne (EFPL), Switzerland, and the M.Sc. (Ing.) in Telecommunication Engineering from the University of Málaga, Spain. He was the recipient of the SESAR Young Scientist Award in 2022 and the Early Career Award of the IEEE Aerospace and Electronics Systems Society in 2023.

ABSTRACT

Global navigation satellite systems (GNSS) is being widely introduced for safety-related applications (beyond aviation). For instance, in the context of train localization, the use of GNSS require the robust modeling of each measurement error component in order to satisfy the safety or integrity requirements. However, due to the challenging railway environment the modeling of the local multipath error remains an active area of research. There is currently no well-accepted multipath error model or approach for railway application that provides a robust and yet not overly conservative error characterization. In order to address this gap, we proposed in previous work to leverage artificial intelligence (AI) for estimating the multipath empirical error distribution using a quantile based description. The quantile-based empirical distribution is then bounded in the cumulative distribution function (CDF) sense in order to provide a parametric and robust model, which can be incorporated into a GNSS integrity monitoring systems. In this work, we explore deeper on one side the impact of different model parameters and design choices of the neural network that estimates the quantile levels on the prediction performance. On the other side, we propose and analyze different architecture possibilities considering the joint estimation of all quantiles together or in a parallel network scheme. The advantages and disadvantages of these design alternatives are analyzed and discussed based on a simplified simulation scenario as well as with real railway data. Finally, the new derived multipath models are applied to a modified H-ARAIM algorithm on the railway scenario.

I. INTRODUCTION

In recent years, significant effort has been put at European level to integrate satellite-based positioning into land-based applications, such as the automotive (European Union Agency for the Space Program, 2024) and railway (European Union Agency for the Space Program, 2023) sectors. The use of GNSS in railway signaling and train control must ensure the stringent safety requirements and provide continuous localization information, while reducing infrastructure costs (European Union Agency for Railways, 2018). An essential aspect for this is safe error characterization of all involved error components. However, different error sources, such as multipath, have a complex nature and its modeling using techniques for aviation multipath do not cover properly the nature of multipath in the railway environment. If these errors are not properly accounted for, they can lead to large positioning errors and, in severe cases, even to hazardous misleading information. This highlights the necessity for an adaptive GNSS multipath error model that can adequately handle the complex error behavior in railway environments. Methods proposed in the literature classify the track areas into use or do not use GNSS using multipath observations (CORDIS, 2022; Gerbeth et al., 2020) or the horizontal position error (HPE) (Kazim et al., 2021), weight observations to reduce their impact (Damy et al., 2016) or model the variance of the railway multipath via satellite geometry and carrier-to-noise ratio (Neri et al., 2021).

More recently, AI has also been proposed to be used in the field of GNSS, in particular to the multipath problem. In (Li et al., 2022), the authors proposed the use of neural network (NN)-based correlation schemes within the GNSS receiver tracking loop for GNSS multipath mitigation. For GNSS non-line-of sight (NLOS) signal classification, the authors in (Kliman & Garcia Crespillo, 2022) employed logistic regression based on GNSS post correlation observations. In (Lee et al., 2023), the authors applied non-linear-regression, using Support-Vector Regression, to create multipath error maps for multipath error

mitigation. In (Zeng et al., 2024), a learning-based spatiotemporal model for environmental characterization and GNSS NLOS signal classification for static locations was presented. In (Zhang et al., 2021), the authors combined a standard NN with long short-term memory (LSTM) for environment characterization, in order to predict satellite visibility and mitigate GNSS pseudorange errors. A GNSS multipath error characterization is estimated using a NN and single quantile regression (SQR) by (No & Milner, 2021). The authors in (Liu & Jiang, 2024), enriched that approach by using multi-sensor information and a LSTM architecture for multipath error characterization via SQR. Summarizing it can be said, that previous work, using AI for error modeling, have either tried to predict the actual error value, exclude affected observations and epochs or estimate the uncertainty based on SQR. However, estimating the uncertainty does not ensure bounding the full probability distribution, which is crucial for integrity applications. For example, heavy-tail distributions can otherwise result in optimistic predictions and lead to potential unbounded situations. Consequently, there is still a gap of robust multipath error models, that can provide a safe error quantification for real-time satellite-based navigation applications in challenging scenarios.

In previous work, we investigated the use of deep multi-quantile regression to estimate jointly different quantiles up to a certain significance level (Roessl & García Crespillo, 2024). The estimated quantiles were then Gaussian overbounded to derive a final parametric and simple model ensuring bounding conditions. In this method, AI is employed to model the highly complex relationship between selected GNSS features and the multipath error. However, the joint estimation of quantiles may result in overly conservative quantiles in some parts of the probability distribution. It was also observed in (Roessl & García Crespillo, 2024), that certain design parameter choices could have a large impact in the behavior of the quantile estimations.

In this work, we propose additionally and alternative architecture to the joint quantile estimation, consisting of a parallel NN deep quantile regression for GNSS multipath modeling. In this approach, each desired quantile is trained with a neural network separately. Furthermore, a sensitivity analysis is performed with respect to model parameters to gain more insights in to the prediction process and explore the effect of certain parameters.

This paper is structured as follows. Section II provides first a short revision of quantile based error overbounding. Potential architectures for quantiles estimation are presented in Section III. Section IV presents a sensitivity analysis of the neural network. A GNSS railway example is given in Section V and discussed in Section VI. Section VII concludes the work.

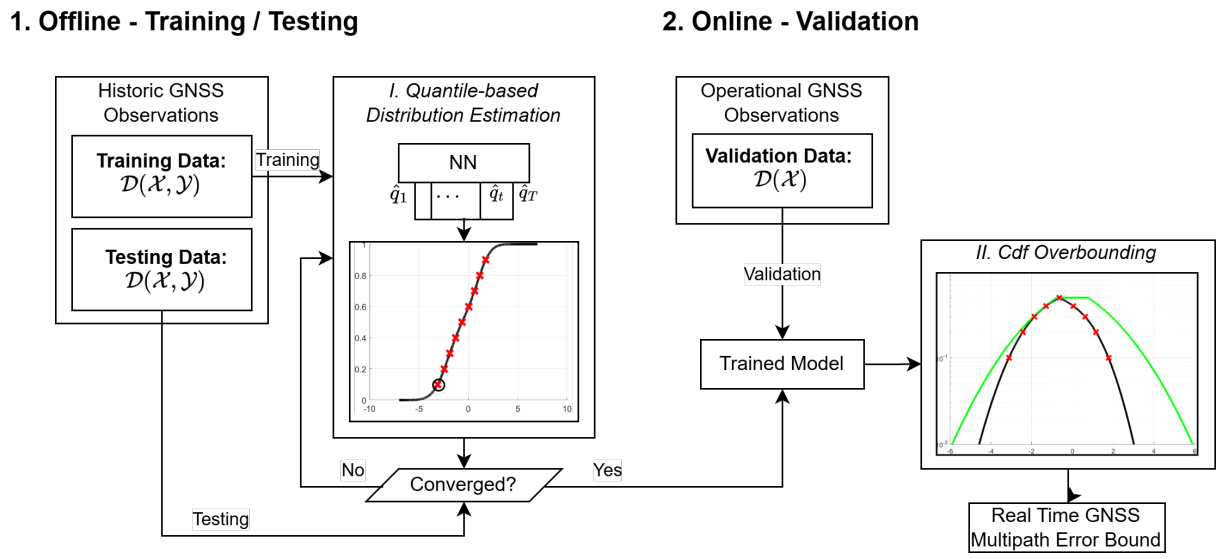


Figure 1: Processing scheme of deep quantile estimation based GNSS multipath error modeling, comprising an offline training and an online validation phase.

II. LEARNING BASED ERROR MODELING

In railway dynamic environment, multipath cannot directly be observed reliably. However, the GNSS receiver provides different measurements in post-correlation that can give some level of information about this error. In order to model multipath based on features whose relationship cannot be easily expressed, we proposed to use a learning based approach. The robust learning based multipath error modeling follows a two step process (Roessl & García Crespillo, 2024). In a first step, an AI based model abstracts the relationship among features $\boldsymbol{X}$ and the target value $\boldsymbol{Y}$ in a supervised-learning approach from training data $\mathcal{D}_{\text{train}}$ to provide a description of the conditional probability distribution function (PDF) $f_{Y|X}(y|x)$. In a second step, that description is transformed into Gaussian bounds to obtain a concise and parametric characterization.

1. Quantile based Distribution Estimation

The proposed learning based modeling approach offers a description of $\boldsymbol{Y} \in \mathcal{Y}$ via its CDF $F_{\boldsymbol{Y}|\boldsymbol{X}}$ as a set of conditional quantiles $\mathbb{Q}_{\mathcal{T}} = \{q_{\tau_1}, \ldots, q_{\tau_t}\}$ of probability levels $\tau \in (0,1)$. During the training process the model abstracts a set of functions, where each of them $g_{\tau,\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{x}) = \hat{q}_\tau$ returns the conditional quantile value $\hat{q}_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$ of target variable $\boldsymbol{Y}$ and specified probability level τ for a given input $\boldsymbol{x}$. A discrete representation of the CDF is therefore created by estimating the distribution via quantiles $\hat{\mathbb{Q}}_{\mathcal{T}}$. The optimization objective function (i.e., learning criterion), we use the pinball or quantile loss function $\mathcal{L}_\tau^{\text{QL}}$ (Koenker, 2005; Koenker & Bassett, 1978):

$$\mathcal{L}_\tau^{\text{QL}}(\zeta) = \begin{cases} \tau \cdot \zeta, & \text{if } \zeta \geq 0 \\ (\tau - 1) \cdot \zeta & \text{otherwise.} \end{cases} \tag{1}$$

The prediction error of the model ζ is calculated as $\zeta = y - g_\tau(\boldsymbol{x})$. In Roessl and García Crespillo (2024) an adapted *conservative* quantile loss $\mathcal{L}_\tau^{\text{CQL}}$ was proposed, that isolates and penalizes non conservative predictions with a hyperparameter α. It can be directly incorporated into the quantile loss as:

$$\mathcal{L}_\tau^{\text{CQL}}(\zeta;\alpha) = \begin{cases} \zeta \cdot (1 + \alpha \cdot c_{\text{ob,l}}(\tau)) \cdot (\tau - 1), & \text{if } \zeta \geq 0 \\ \zeta \cdot (1 + \alpha \cdot c_{\text{ob,r}}(\tau)) \cdot \tau, & \text{otherwise.} \end{cases} \tag{2}$$

The isolation and normalization of non conservative predictions for the left $c_{\text{ob,l}}$ ($\tau \leq 0.5$) and right $c_{\text{ob,r}}$ ($\tau > 0.5$) side of the distribution are:

$$c_{\text{ob,l}}(\tau) = \begin{cases} \dfrac{1}{|\tau - 1|}, & \text{if } \tau < 0.5, \\ \dfrac{1}{2} \cdot \dfrac{1}{|\tau - 1|}, & \text{if } \tau = 0.5, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

and

$$c_{\text{ob,r}}(\tau) = \begin{cases} 0, & \text{if } \tau < 0.5, \\ \dfrac{1}{2} \cdot \dfrac{1}{|\tau|}, & \text{if } \tau = 0.5, \\ \dfrac{1}{|\tau|}, & \text{otherwise.} \end{cases} \tag{4}$$

The learning problem to be optimized during the training phase by the NN can be finally expressed as:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{(\mathbf{x},y)}^{\mathcal{D}_{\text{train}}} \sum_{\tau}^{\mathcal{T}} \mathcal{L}_\tau^{\text{CQL}}(y - g_\tau(\mathbf{x};\boldsymbol{\theta})) + \beta \cdot \mathcal{P}(\mathcal{T}, \mathcal{D}_{\text{train}}) + \gamma \omega^T \omega \tag{5}$$

where $\boldsymbol{\theta}$ are the model parameters and $\mathcal{P}(\mathcal{T}, \mathcal{D})$ is a quantile crossing penalty introduced to prevent quantile intersections with a minimum distance of ν (Fakoor et al., 2023) and a tunable hyperparameter γ for the weight-decay (Murphy, 2022):

$$\mathcal{P}(\mathcal{T}, \mathcal{D}) = \sum_{(x,y)}^{\mathcal{D}_{train}} \sum_{t=1}^{N_{\mathcal{T}}-1} \max\left(g_{\tau_t}(\mathbf{x};\boldsymbol{\theta}) - g_{\tau_{t+1}}(\mathbf{x};\boldsymbol{\theta}) + \nu, 0\right). \tag{6}$$

2. Quantile based Gaussian Bounds

In the second step the quantile distribution is conservatively described by a CDF overbound using a Gaussian distribution $\mathcal{N}(\pm b, \sigma)$. The parameters are determined using the two-step Gaussian overbounding technique as presented in (Blanch et al., 2019), combining symmetric CDF (DeCleene, 2000) and paired overbounding (Rife et al., 2006). First, a left and right side intermediate symmetrical unimodal distribution $\mathbb{F}_{\text{su}}$ are determined based on the quantiles, as presented in (Roessl & García Crespillo, 2024), in order to satisfy the overbounding conditions of zero mean, symmetry and unimodal (DeCleene, 2000). Second, both distributions, $\mathbb{F}_{\text{su, left}}$ and $\mathbb{F}_{\text{su, right}}$ are Gaussian bounded by $\mathbb{F}_{\text{ob,left}}(y;\mu_{\text{left}},\sigma_{\text{left}})$ and $\mathbb{F}_{\text{ob,right}}(y;\mu_{\text{right}},\sigma_{\text{right}})$, and summarized by the maximum operation into one final Gaussian overbound $\mathbb{F}_{\text{ob}}(y;b,\sigma)$. The final overbound can be expressed as:

$$f_{\mathbf{Y}}(y) \prec f_{\text{ob}}(y;b,\sigma), \tag{7}$$

where $\prec$ denotes the overbounding condition such that $\mathbb{F}_{\boldsymbol{Y}} \leq \mathbb{F}_{\text{ob}},\ \forall y \leq 0$ and $\mathbb{F}_{\boldsymbol{Y}} \geq \mathbb{F}_{\text{ob}},\ \forall y > 0$.

Depending on the application and its requirements, only a subset of quantiles might be relevant. That is, for a specific integrity risk, the set of quantiles $\mathcal{Q}_{\mathcal{T}_{\text{IR}}}$ up to $IR/2$ tail probability levels may be necessary. For example, for accuracy purposes, with the centered prediction interval of $95\,\%$, the following subset is defined:

$$\hat{\mathcal{Q}}_{\mathcal{T}_{\text{acc}}} = \left\{ \hat{q}_\tau \in \hat{\mathcal{Q}}_{\mathcal{T}} : \frac{1-0.95}{2} \leq \tau_i \leq \left(1 - \frac{1-0.95}{2}\right) \right\}. \tag{8}$$

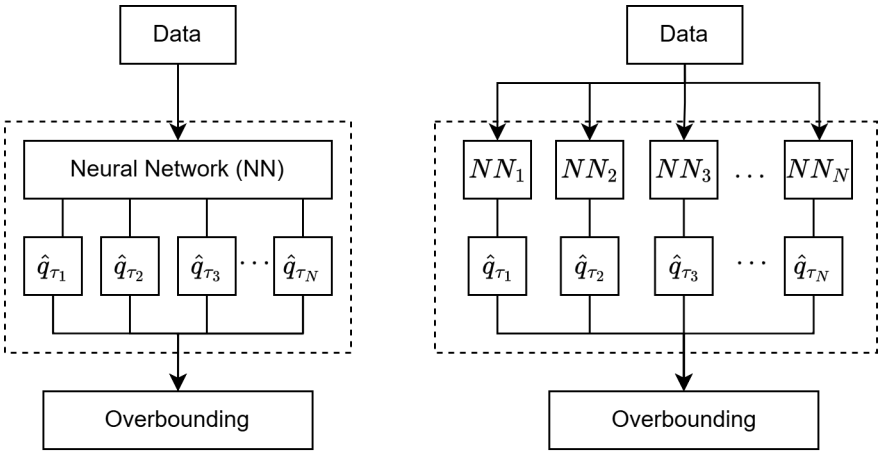
The two step procedure of the proposed methodology can be expressed as functions:

$$g_{\text{AI}}(\mathbf{x}; \mathcal{T}) = \hat{\mathcal{Q}}_{\mathcal{Y}|\mathcal{X}=\mathbf{x}_i}, \tag{9}$$

$$h_{\text{ob}}(\hat{\mathcal{Q}}, \mathcal{T}) \mapsto [b, \sigma]_{\text{AI}}^{T}. \tag{10}$$

III. NEURAL NETWORK ARCHITECTURE ALTERNATIVES

This work investigates the estimation process of the error distribution based on quantiles. Fig. 2 presents two possible architecture layouts for designing the quantile estimation process: a joint and a parallel strategy. In (Roessl & García Crespillo, 2024), we used the joint quantile regression (JQR) strategy for GNSS multipath error modeling, where one NN approximates all selected quantile functions. However, the impact of alternatives strategies was raised as an outlook for future work. Potential advantages of JQR, as shown in (Fakoor et al., 2023), are the robust estimation process as *knowledge* for function modeling is shared among different quantile functions. Instead of estimating all quantiles with a single NN, the parallel quantile regression (PQR) architecture approach estimates only a single quantile function per NN. This allows for further flexibility in the training process and potentially reduced model complexity.



(a) Joint quantile estimation (JQR) **(b)** Parallel quantile estimation (PQR)

Figure 2: Neural network architectures for joint (a) and parallel (b) quantile estimation for distribution modeling.

The optimization objective function $\mathcal{J}_{\text{JQR}}$ used for training the JQR is given in (5) and it is here repeated for convenience:

$$\mathcal{J}_{\text{JQR}}(\boldsymbol{\theta}; \mathcal{D}_{\text{train}}, \mathcal{T}) = \sum_{(\mathbf{x},y)}^{\mathcal{D}_{\text{train}}} \sum_{\tau}^{\mathcal{T}} \mathcal{L}_\tau^{\text{CQL}}(y - g_\tau(\mathbf{x}; \boldsymbol{\theta})) + \beta \cdot \mathcal{P}(\mathcal{T}, \mathcal{D}_{\text{train}}) + \gamma \omega^T \omega. \tag{11}$$

For the PQR architecture, the optimization problem $\mathcal{J}_{\text{PQR}}$ simplifies to minimizing only one quantile probability level τ per neural network. Hence, it is given as:

$$\mathcal{J}_{\text{PQR}}(\boldsymbol{\theta}; \mathcal{D}_{\text{train}}, \tau) = \sum_{(\mathbf{x},y)}^{\mathcal{D}_{\text{train}}} \mathcal{L}_\tau^{\text{CQL}}(y - g_\tau(\mathbf{x}; \boldsymbol{\theta})) + \beta \cdot \mathcal{P}(\mathcal{T}, \mathcal{D}_{\text{train}}) + \gamma \omega^T \omega, \tag{12}$$

where the non-crossing penalty can be calculated during parallel training.

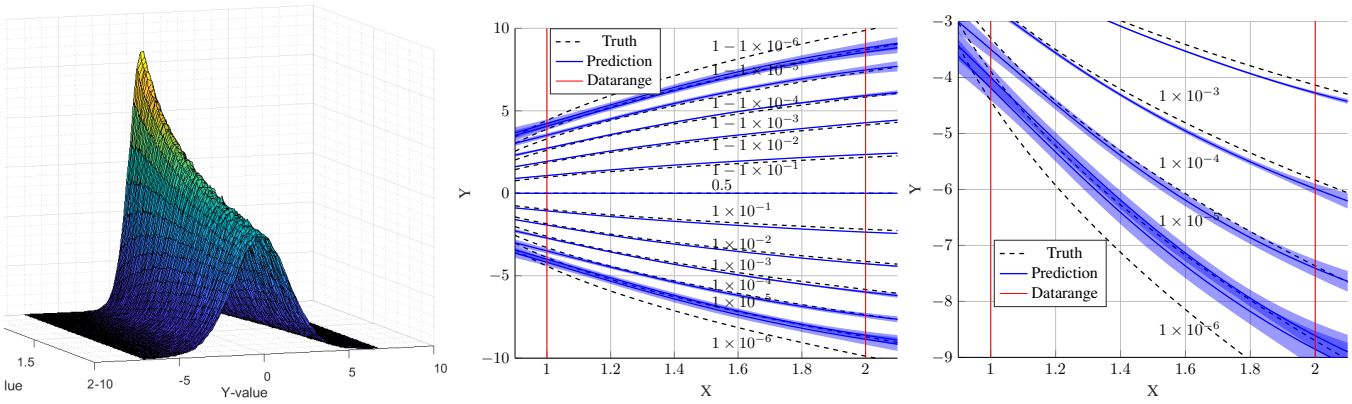
IV. NEURAL NETWORK SENSITIVITY ANALYSIS

For abstracting the relationship between numerical features $\boldsymbol{X}$ and the multipath distribution $\boldsymbol{Y}$ a shallow NN with one input, three hidden and one output layers are used to keep the depth of the model as simple as possible and complex as needed. It was observed in previous work that dataset size and model parameters choices could have a large impact onto the prediction quality. In order to understand this impact, this section provides a sensitivity analysis on different aspects with Monte-Carlo simulations over the quantile learning of a simplified model example.

In this example, the target value $\boldsymbol{y} \in \mathbb{R}^1$ is modeled as a Gaussian mixture model to be dependent on the input feature $\boldsymbol{x} \in \mathbb{R}^1$ via $y(x) \sim \sum_{i=1}^{2} \pi_i \cdot \mathcal{N}(0, \sigma_i(x))$ with weight vector $\boldsymbol{\pi} = [0.8, 0.2]^T$ and standard deviation vector $\boldsymbol{\sigma} = [-2x+5, -4x+10]^T$ where $x \sim \mathcal{U}(1, 2)$. Fig. 3a shows its associated PDF. In total, 30 simulations are performed, each with a dataset size of 1×10^6 for training.

1. Impact of Finite Training Dataset

First, Fig. 3 shows the impact of the dataset size on the prediction quality for different probability levels estimated by the JQR architecture. The predicted quantiles, which describe the core of the distribution, match the truth well, as they are upper bounding and demonstrate a small variance. For small probability levels, e.g. 1×10^{-6}, we can see that the model is no longer able to make upper bounding predictions with respect to the truth. At the same time, the variance of predictions grow. One intuitive explanation is the lack of probability level representativeness due to the limited dataset in the training process. Low probabilities values are not represented enough in the dataset. This effect is present in every data-driven model and must be properly considered in the design process during training, by either selecting only a certain tail probability level or using sufficient data (Goodfellow et al., 2016).



(a) PDF of simulated Gaussian mixture dataset

(b) Predictions

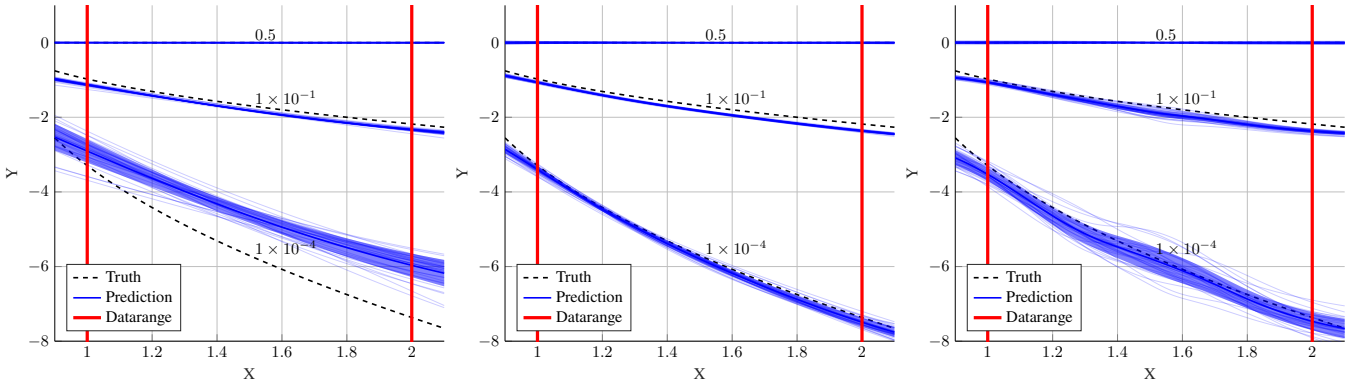
(c) Zoomed predictions $y \in (-9, -3)$

Figure 3: Monte-Carlo simulation based experiment of the variance of predicted quantiles with respect to selected probability levels. The black dashed line represents the true and blue (mean $\pm$ standard deviation) the predicted quantile values of $\boldsymbol{y}$ as a function of $\boldsymbol{x}$ with red denoting the range of the training dataset.

2. Model Complexity

The next experiment investigates the effect of model complexity or capacity on the prediction process. Model complexity usually describes the number of hidden layers and number of neurons used for modeling (Goodfellow et al., 2016), similar to the degree of a polynomial. In particular, we investigate the impact of choosing 4, 128 or 512 neurons in a five layer shallow NN.

Fig. 4 shows the predictions made by the JQR model for different number of neurons. When using a low model complexity, it is possible to obtain upper bounding predictions for probabilities 0.1 and 0.5. However, the NN fails to properly model small probability quantile functions, e.g. 1×10^{-4}, and the model suffers from underfitting (Goodfellow et al., 2016). Increasing complexity improves the prediction of these small probability levels as upper bounding predictions are obtained when using 128 neurons. However, Fig. 4c shows that the variance of the predictions starts to increase when using 512 neurons and the model suffers from overfitting to the training data and is no longer able to generalize equally well to the evaluation data.



(a) JQR$_{(1\text{-}4\text{-}8\text{-}4\text{-}1)}$ **(b)** JQR$_{(1\text{-}128\text{-}256\text{-}128\text{-}1)}$ **(c)** JQR$_{(1\text{-}512\text{-}1024\text{-}512\text{-}1)}$

Figure 4: Monte-Carlo simulation based model generalization performance affected by under/overfitting

3. Number of Quantiles

The impact of the number of selected quantiles $N_{\mathcal{T}}$on the prediction quality is explored using the JQR architecture and PQR as a reference. The model complexity, number of neurons and hidden layers, of the JQR and PQR is not modified for the individual configurations, only the number of output neurons is changed, e.g. (1-128-256-128-5) or (1-128-256-128-201). The centered prediction interval (CPI) of $99.98\,\%$ is modeled using a range of quantiles from 5-201. Fig. 5 shows the true and predicted quantile values as a function of $\boldsymbol{x}$ for probability levels 0.5, 0.025 and 1×10^{-4}. A variation in the prediction performance when using the JQR method is obtained when increasing $N_{\mathcal{T}}$. Especially for the small probability of $0.01\,\%$, the model fails to estimate an upper bounding prediction of the quantile value with respect to the truth. The PQR model does not suffer from that effect as one NN learns only one quantile function, however the variances of the predictions increase.

V. EXPERIMENTAL MULTIPATH MODELING AND H-ARAIM RESULTS

The proposed methodology is evaluated for the GNSS code multipath error modeling of satellite-based train localization. First, the error characterization performance of the model, using both architectures, for unknown data is investigated. Second, a exemplary application of the estimated multipath models along the railway tracks is presented via a modified horizontal advanced receiver autonomous integrity monitoring (ARAIM) (HARAIM) algorithm for integrity monitoring.

1. Railway GNSS Data and Experimental Settings

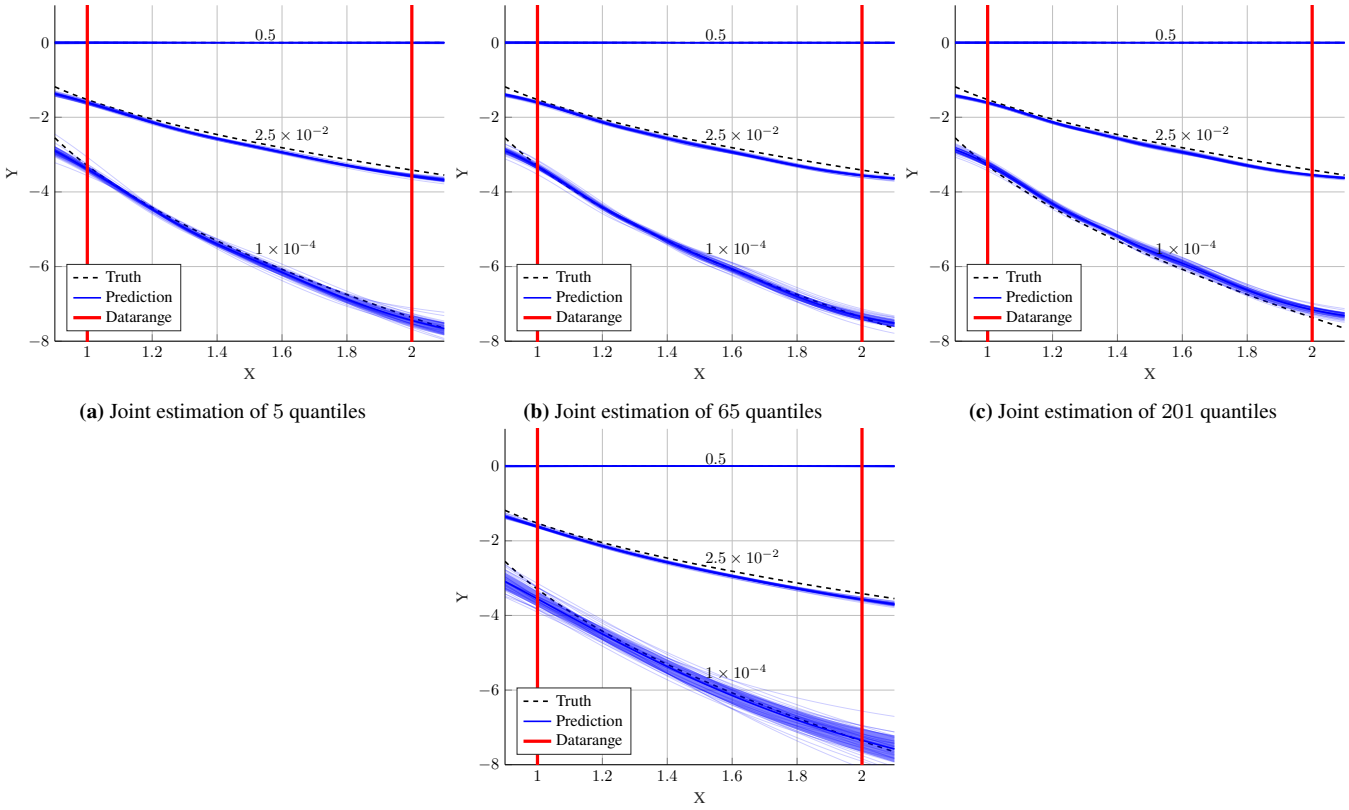
Fig. 6 shows the maintenance train used for GNSS sensor data collection within the European research project *RailGap* in Málaga, Spain (RAILGAP, 2024). The dataset includes circa $27\,\mathrm{h}$ of multi-constellation and multi-frequency GNSS observations in urban and rural environments. The feature values $\boldsymbol{X}$, as described in (Roessl & García Crespillo, 2024), which contain information for multipath modeling, are carrier-to-noise ratio C/N_0, elevation angle θ, local azimuth ψ, pre-normalized $\overline{C/N_0}$, change rate $\Delta C/N_0$, lock time t_{lock}, pseudorange consistency $\overline{\text{PRC}}$, velocity v and permanent multipath error bound σ_{perm} (Kliman et al., 2024). The code-minus-carrier linear combination is applied for estimating the target multipath target value $\boldsymbol{Y}$. It provides an estimate of the GNSS code multipath and receiver noise for a specific satellite observation (Braasch, 1994).

The model is trained for 500 epochs using the Adam optimizer (Kingma & Ba, 2017). In order to obtain more robust results the k-fold cross-validation split strategy, as depicted in Fig. 6, is employed resulting in 5 separate models each trained with a different subset of training data.

Fig. 7 plots the evolution of the total loss (see Eq.(5)), sharpness and interval score for the JQR and PQR architectures during training and testing. Sharpness is a measure of narrowness or the variance of prediction interval width (Gneiting et al., 2007). The authors in (Yan et al., 2023) estimate the expected interval sharpness EIS as:

$$EIS(\hat{\mathbb{Q}}_{\mathcal{T}}) = \frac{1}{t}\sum_{k}^{t}\frac{1}{N_{\mathcal{D}}}\sum_{i}^{N_{\mathcal{D}}}|g_{\tau_{N_{\mathcal{T}}-k}}(\mathbf{x}_i) - g_{\tau_k}(\mathbf{x}_i)|. \tag{13}$$

The interval score is a loss function that can be used for centered interval prediction, but also as an evaluation metric. It is



(a) Joint estimation of 5 quantiles

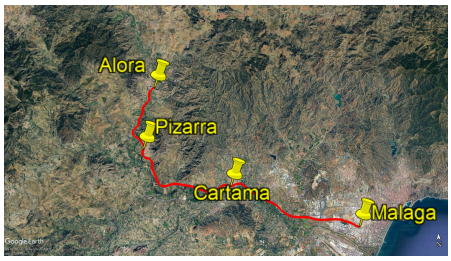
(b) Joint estimation of 65 quantiles

(c) Joint estimation of 201 quantiles

(d) Parallel estimation of 5 quantiles

Figure 5: Monte-Carlo simulation based sensitivity analysis of the dependency between number of estimated quantiles and prediction performance. The black dashed line shows the true quantile values for selected probability levels, blue the mean predictions and the standard deviation of the different models.

(a) ADIF maintenance train used for GNSS field data collection in Málaga, Spain.



(b) RailGap measurement campaign area

Experiment	Data batches					
$k=1$	1	2	3	4	5	6
$k=2$	1	2	3	4	5	6
$k=3$	1	2	3	4	5	6
$k=4$	1	2	3	4	5	6
$k=5$	1	2	3	4	5	6

(c) Cross-validation dataset split into training, testing and evaluation batches

Figure 6: Map of the RailGap measurement campaign train and area and dataset split.

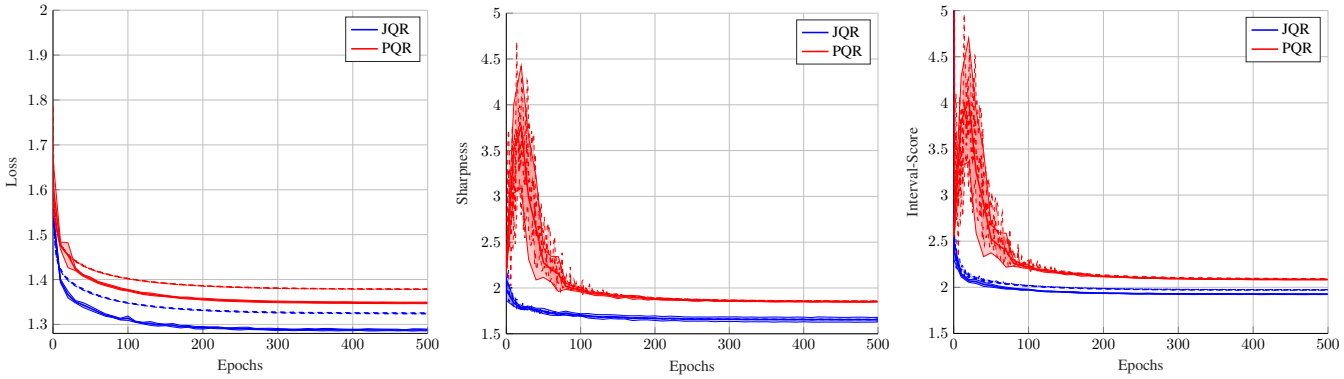
expressed as (Gneiting & Raftery, 2007):

$$S_{\varepsilon}^{int}(l,u;y) = (u-l) + \frac{2}{\varepsilon}(l-y)\cdot \mathbb{1}\{y \le l\} + \frac{2}{\varepsilon}(y-u)\cdot \mathbb{1}\{y > u\}, \tag{14}$$

with the central prediction interval $PI = [l,u]$ and penalty depending on $\varepsilon = 1 - PI$.

2. Multipath Error Modeling

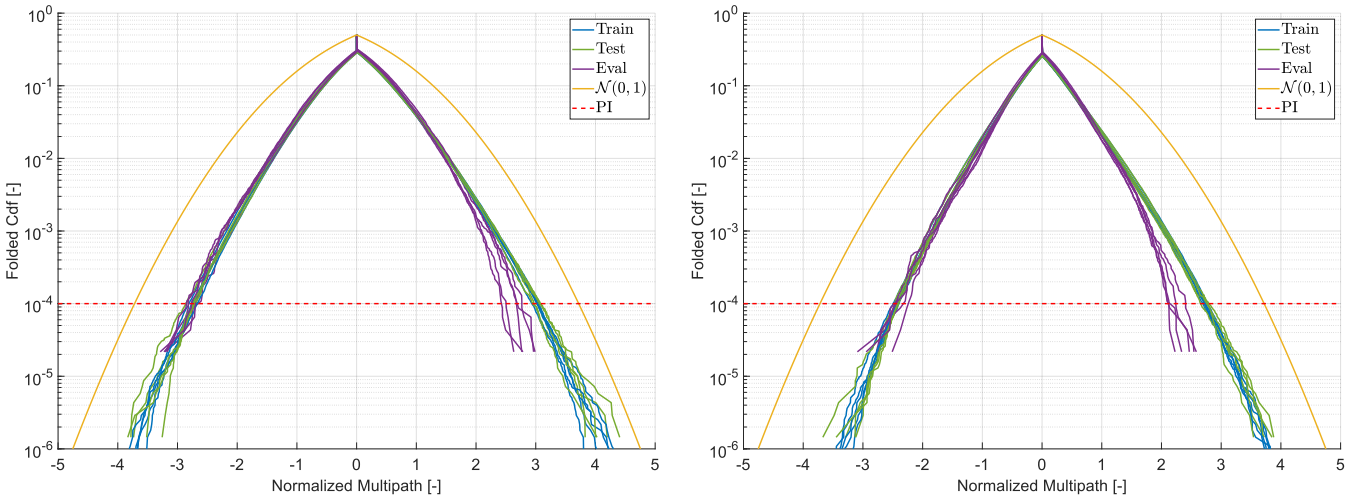
Fig. 8 shows the normalized folded CDF of the training, testing and evaluation dataset using the JQR and PQR architectures. The isolated multipath observations y are normalized with the estimated bounds via $\overline{y} = \text{sgn}(y)\cdot \frac{\max(|y|-b_{\text{ob}},0)}{\sigma_{\text{ob}}}$. In the CPI



(a) Total loss **(b)** Sharpness **(c)** Interval Score

Figure 7: Evaluation metrics along the training (dashed) and testing (full line) procedure for joint (blue) and parallel (red) quantile estimation architectures.

of $99.98\,\%$, as indicated by the red dashed line, all distributions are overbounded by a folded standard normal distribution. Analyzing the results of the different architectures it can be observed that the gap between the standard normal and the normalized distributions is larger for the PQR than for the JQR architecture.



(a) JQR **(b)** PQR

Figure 8: Resulting normalized distributions of the training (blue, 2769327), testing (green, 692332) and evaluation (purple, 46082) datasets using the joint (a), parallel (b) quantile estimation after convergence for the k-fold cross validation experiment.

3. Modified H-ARAIM Results

The proposed model for user multipath characterization is included in a modified HARAIM algorithm, similarly as presented in (Roessl & García Crespillo, 2025; Roessl et al., 2023) by adapting the baseline ARAIM algorithm (Blanch et al., 2015). ARAIM is an integrity monitoring system, which exploits redundant measurements to perform consistency checks and provides a probabilistic bound of the unknown position error (also known as protection level (PL)). Fig. 9 visualizes the flowchart of the ARAIM algorithm using the AI-based multipath error model.

The main adaptations include the replacement of the local user-side error model with the multipath model estimated with AI dynamically over time. In particular, the covariance matrix used for integrity purposes $\mathbf{\Sigma}_{\text{int}}$ is a diagonal matrix with elements σ^2_{int} for each observation i as:

$$\sigma^2_{\text{int}} = \sigma^2_{\text{ura}} + \sigma^2_{\text{tropo}} + \sigma^2_{\text{iono}} + \sigma^2_{\text{rail,int}}, \tag{15}$$

with the variance of the residual satellite orbit and clock σ_{ura}, troposphere σ_{tropo} and ionosphere σ_{iono} error model for the single

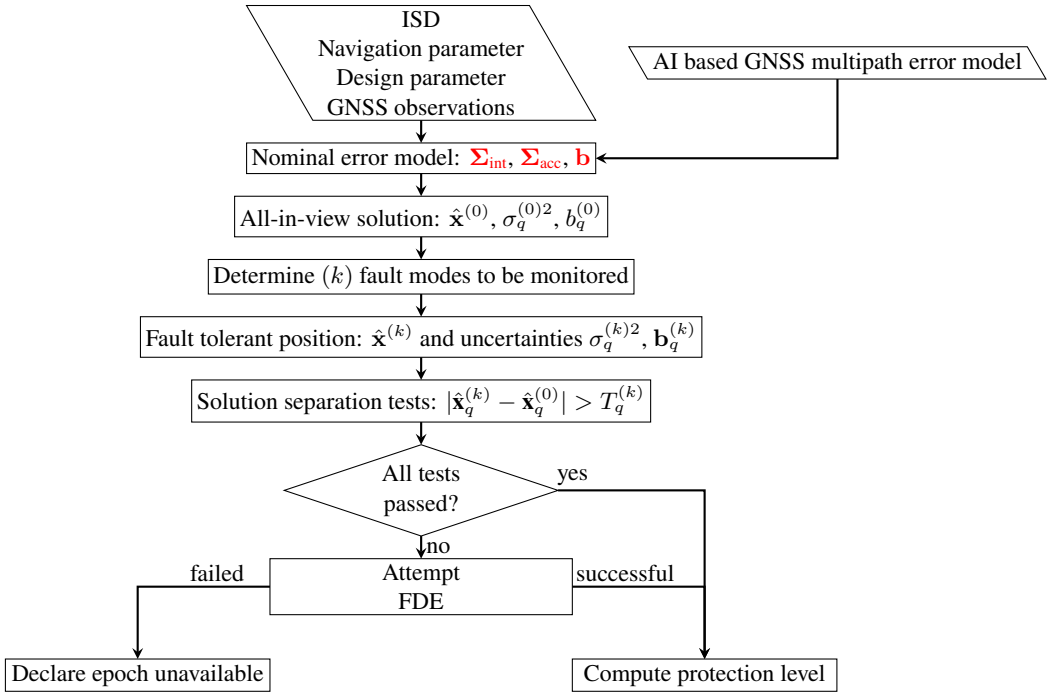


Figure 9: Integration of the learning based GNSS code multipath error model into the ARAIM algorithm with FDE for railway localization

frequency user (Roessl et al., 2023). For railway users, the multipath and noise residual error model $\sigma_{\text{rail,int}}$ and bias $b_{\text{rail,int}}$ are estimated by the proposed method using quantiles $\hat{\mathcal{Q}}_{\mathcal{T}_{\text{int}}}$ and the CDF bounding process. The pseudorange biases $\mathbf{b}$ can be expressed as a sum of the signal-in-space (SiS) nominal bias and the multipath bias by:

$$b = b_{\text{nom}} + b_{\text{rail, int}}, \tag{16}$$

where b_{nom} is the nominal bias, contained e.g., in an integrity support message (ISM) message. The covariance matrix used for accuracy and continuity purposes $\mathbf{\Sigma}_{\text{acc}}$ is formed as a diagonal matrix where each element σ^2_{acc} is:

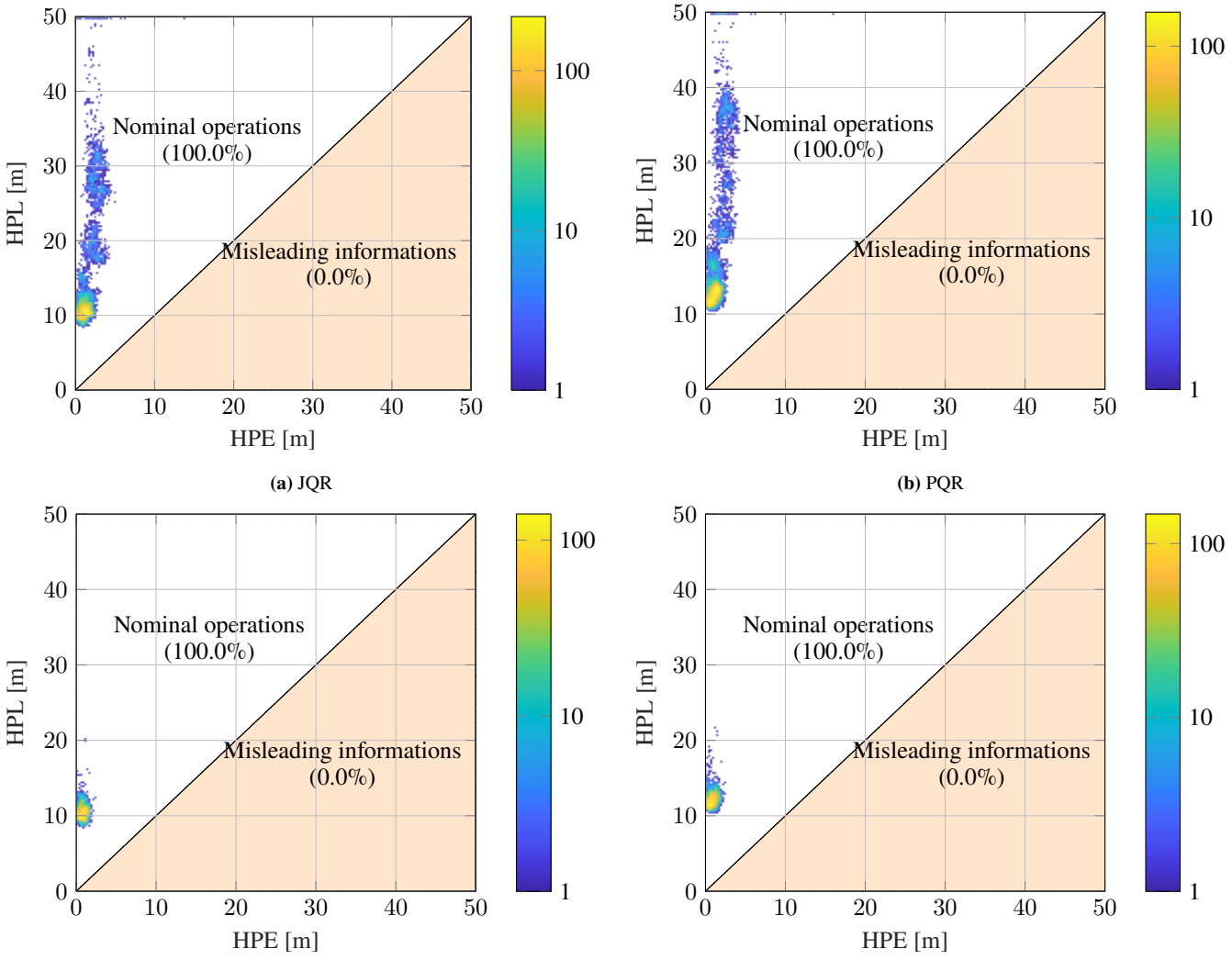
$$\sigma^2_{\text{acc}} = \sigma^2_{\text{ure}} + \sigma^2_{\text{tropo}} + \sigma^2_{\text{iono}} + \sigma^2_{\text{rail, acc}}, \tag{17}$$

where σ_{ura} is the accuracy satellite orbit and clock error and $\sigma^2_{\text{rail, acc}}$ is the variance of the accuracy model for the railway user based on $\hat{\mathcal{Q}}_{\mathcal{T}_{\text{acc}}}$ and estimated by the AI-based model.

Fig. 10 presents the HARAIM results using the learning-based multipath model in the Stanford plot. Estimated horizontal protection level (HPL) are plotted versus the HPE with respect to a multi-sensor post processed ground-truth solution. It is found that HPL $\geq$ HPE holds for all epochs, so that zero misleading information is obtained. Since position errors can be due both to measurement errors and poor geometry, we considered in the second row of Fig. 10 only those results with good geometry (GDOP ≤ 1.5), with a HPL range of 8-20 m.

VI. DISCUSSION

One of the main goals of the Monte-Carlo simulation sensitivity analysis was to explore the impact of model parameters on the prediction performance for quantile based distribution modeling. The results shown in Fig. 3 demonstrate the importance of dataset size for quantile estimation. In particular, for tail probability levels, e.g. 1×10^{-6} and 1×10^{-5}, an increased variance of the estimate was observed, indicating insufficient representativeness due to limited amount of data (1×10^{6}). In Fig. 4 the impact of model complexity was studied. When selecting too few neurons, the model was not able to learn the complex function properly and hence suffered from underfitting. Using too many neurons resulted in overfitting, as the model was not able to generalize well to unknown evaluation data (Goodfellow et al., 2016). However, these effects can be monitored by observing the presented evaluation metrics as well as the variance of the predictions made by the individual models during the training and testing phases. Fig. 5 shows that when using the JQR architecture, the number of estimated quantiles affects the prediction



(a) JQR

(b) PQR

(c) JQR results GDOP $\leq$ 1.5 filtered

(d) PQR results GDOP $\leq$ 1.5 filtered

Figure 10: Modified H-ARAIM results using the proposed learning-based GNSS code multipath error model

performance. If the model complexity is not properly adopted low tail probability quantiles predictions are no longer upper bounding the true distribution. This additional complexity between number of quantiles and model complexity can be avoided by using the PQR architecture, however at the cost of increased variance in the predictions. Comparing the performance of the JQR and PQR architecture, it can be concluded that JQR provides tighter results at the cost of additional complexity in the model design process, because the number of estimated quantiles affects the required number of neurons. The results from the real-data GNSS multipath evaluation demonstrate the capability of the proposed method to abstract the relationship between features and the multipath error from training data. This is shown as the folded normalized error distributions are overbounded by the standard normal distribution. One reason for the conservatism in Fig. 8 may be a result of the normalization process, since in general, each error sample belongs to a different distribution. In the future, a better representation of the distribution results will be investigated.

The modified HARAIM results in the Stanford plot demonstrate the capability of the method to obtain a safe position error quantification via HPL for GNSS based train localization. When considering only good satellite geometry epochs, it can be seen that the large protection levels are mostly due to poor geometry and not due to an extra conservatism introduced by the proposed error model. Future work will investigate the generalization performance of the approach to unknown data of a different datasets and investigate the time correlation of the multipath error.

VII. CONCLUSION

The use of AI for GNSS error modeling for safety purposes is an interesting new and challenging area of research. In this work, the proposed method was capable to abstract from training dataset the relationship between selected post-correlation GNSS features and the observed multipath in order to come up with a robust multipath error description for railway application. Furthermore, thanks to the sensitivity analysis, we provided new insights on the performance of AI predictions, which increases its explainability and important design considerations.

This new method is a potential solution to obtain a safe position error quantification considering local multipath effects along the railway tracks for GNSS based train localization using commercial of the shelf receivers in combination with integrity monitoring systems such as HARAIM or satellite-based augmentation systems (SBAS).

ACKNOWLEDGMENT

This work has been partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No: 101004129 *RAILGAP* project. Furthermore, we would like to thank our collegue Ana Kliman for prodiving the permanent multipath model characterizing the train antenna installation, which is used as an input feature.

REFERENCES

Blanch, J., Walter, T., & Enge, P. (2019). Gaussian Bounds of Sample Distributions for Integrity Analysis. *IEEE Transactions on Aerospace and Electronic Systems*, *55*(4), 1806–1815. https://doi.org/10.1109/TAES.2018.2876583

Blanch, J., Walter, T., Enge, P., Lee, Y., Pervan, B., Rippl, M., Spletter, A., & Kropp, V. (2015). Baseline advanced RAIM user algorithm and possible improvements. *IEEE Transactions on Aerospace and Electronic Systems*, *51*(1), 713–732. https://doi.org/10.1109/TAES.2014.130739

Braasch, M. S. (1994). Isolation of GPS Multipath and Receiver Tracking Errors. *Navigation*, *41*(4), 415–435. https://doi.org/10.1002/j.2161-4296.1994.tb01888.x

CORDIS. (2022). ERTMS on SATELLITE Galileo Game Changer — ERSAT GGC — Project — Results — H2020 — CORDIS — European Commission. Retrieved September 22, 2025, from https://cordis.europa.eu/project/id/776039/results

Damy, S., Majumdar, A., & Ochieng, W. Y. (2016). GNSS-based High Accuracy Positioning for Railway Applications. *Proceedings of the 2016 International Technical Meeting of The Institute of Navigation*, 1003–1014. https://doi.org/10.33012/2016.13479

DeCleene, B. (2000). Defining Pseudorange Integrity - Overbounding. *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, 1916–1924.

European Union Agency for Railways. (2018). ERA's Vision, Mission, Values and Tasks. Retrieved November 17, 2024, from https://www.era.europa.eu/agency-you/agency/vision-mission-values-tasks

European Union Agency for the Space Program. (2023). Report on Rail: User Needs and Requirements. Retrieved April 2, 2025, from http://www.euspa.europa.eu/

European Union Agency for the Space Program. (2024). Report on Road and Automotive: User Needs and Requirements. Retrieved February 19, 2025, from https://www.euspa.europa.eu/sites/default/files/documents/Report%20on%20Road%20and%20Automotive%20User%20Needs%20and%20Requirements.pdf

Fakoor, R., Kim, T., Mueller, J., Smola, A. J., & Tibshirani, R. J. (2023). Flexible Model Aggregation for Quantile Regression. https://doi.org/10.48550/arXiv.2103.00083

Gerbeth, D., Crespillo, O. G., Pognante, F., Vennarini, A., & Coluccia, A. (2020). Framework to Classify Railway Track Areas According to Local GNSS Threats. *2020 European Navigation Conference (ENC)*, 1–11. https://doi.org/10.23919/ENC48637.2020.9317368

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(2), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Kazim, S. A., Tmazirte, N. A., & Marais, J. (2021). On The Impact Of Temporal Variation On GNSS Position Error Models. https://doi.org/10.33012/2021.17863

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. https://doi.org/paper

Kliman, A., & Garcia Crespillo, O. (2022). Characterization of GNSS Multipath in Nominal Open-Sky Scenario for Safe Railway Localization. In *DGON POSNAV 2022*.

Kliman, A., Grosch, A., & García Crespillo, O. (2024). Permanent train-side GNSS multipath characterization considering time-correlation for safe railway localization. *2024 European Navigation Conference (ENC)*.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge university press.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33. https://doi.org/10.2307/1913643

Lee, Y., Wang, P., & Park, B. (2023). Nonlinear Regression-Based GNSS Multipath Dynamic Map Construction and Its Application in Deep Urban Areas. *IEEE Transactions on Intelligent Transportation Systems*, *24*(5), 5082–5093. https://doi.org/10.1109/TITS.2023.3246493

Li, H., Borhani-Darian, P., Wu, P., & Closas, P. (2022). Deep Neural Network Correlators for GNSS Multipath Mitigation. *IEEE Transactions on Aerospace and Electronic Systems*, 1–23. https://doi.org/10.1109/TAES.2022.3197098

Liu, R., & Jiang, Y. (2024). Overbounding Multipath Error in Urban Canyon With LSTM Using Multi-Sensor Features. *IEEE Transactions on Intelligent Transportation Systems*, 1–15. https://doi.org/10.1109/TITS.2024.3376812

Murphy, K. P. (2022). *Probabilistic Machine Learning: An introduction*. MIT Press. probml.ai

Neri, A., Capua, R., Filip, A., Ruggeri, A., & Baldoni, S. (2021). Integrity Bounds for Rail and Road Applications Based on Local Hazard Maps. *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, 4157–4169. https://doi.org/10.33012/2021.18079

No, H., & Milner, C. (2021). Machine Learning Based Overbound Modeling of Multipath Error for Safety Critical Urban Environment, 180–194. https://doi.org/10.33012/2021.17874

RAILGAP. (2024). Railway ground truth and digital map. Retrieved September 29, 2024, from https://cordis.europa.eu/project/id/101004129/results

Rife, J., Pullen, S., Enge, P., & Pervan, B. (2006). Paired overbounding for nonideal LAAS and WAAS error distributions. *IEEE Transactions on Aerospace and Electronic Systems*, *42*(4), 1386–1395. https://doi.org/10.1109/TAES.2006.314579

Roessl, F., Crespillo García, O., Heirich, O., & Kliman, A. (2023). A Map Based Multipath Error Model for Safety Critical Navigation in Railway Environments. *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 446–457. https://doi.org/10.1109/PLANS53410.2023.10140130

Roessl, F., & García Crespillo, O. (2024). Robust GNSS Multipath Error Modeling Based on Deep Quantile Regression with Gaussian Overbounding. *Proceedings of the 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024)*, 1402–1415. https://doi.org/10.33012/2024.19762

Roessl, F., & García Crespillo, O. (2025). Integrity Monitoring of GNSS Railway Localization with a Map-based Multipath Error Model. *Unpublished*.

Yan, X., Su, Y., & Ma, W. (2023). Ensemble Multi-Quantiles: Adaptively Flexible Distribution Prediction for Uncertainty Quantification. *IEEE transactions on pattern analysis and machine intelligence*, *45*(11), 13068–13082. https://doi.org/10.1109/TPAMI.2023.3288028

Zeng, K., Li, Z., Zhao, H., Xie, K., Xie, S., Niyato, D., Chen, W., & Zheng, Z. (2024). A Spatiotemporal Information-Driven Cross-Attention Model With Sparse Representation for GNSS NLOS Signal Classification. *IEEE Internet of Things Journal*, *11*(19), 31892–31908. https://doi.org/10.1109/JIOT.2024.3423016

Zhang, G., Xu, P., Xu, H., & Hsu, L.-T. (2021). Prediction on the Urban GNSS Measurement Uncertainty Based on Deep Learning Networks With Long Short-Term Memory. *IEEE Sensors Journal*, *21*(18), 20563–20577. https://doi.org/10.1109/JSEN.2021.3098006