Railway GNSS Multipath Error Modelling Approach with both Train-side and Operational Environment Characterization

Ana Kliman, Florian Roessl, Anja Grosch, Omar García Crespillo
German Aerospace Center (DLR)

BIOGRAPHY

Ana Kliman obtained a bachelor and Master's degree in Geodesy and Geoinformatics from University of Zagreb, Croatia. After which she joined DLR's institute of Navigation and Communication, Integrated Navigation Integrity group. Her topics of interest include high accuracy, high integrity GNSS with an emphasis on railway applications.

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

Anja Grosch received the German diploma in Computer Engineering from the Ilmenau University of Technology, Germany in 2007. In March 2008, she has joined DLR's navigation department. Since then she has been working on multi-sensor systems designed for safety-of-life applications for civil aviation, railway and automotive. Her main focus is the fusion of GNSS and INS and the development of suitable integrity concepts. This includes the sensor error modeling, fault detection and exclusion strategies and performance bounding.

Omar Garcia 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

Railway transportation systems have high accuracy and high integrity demands for safe localization. In the future, railway signaling is expected to rely on onboard sensors like Global Navigation Satellite Systems (GNSS) in order to reduce installation and maintenance costs. GNSS position determination can be however highly degraded because of the presence of multipath caused by the train roof and the challenging railway environment. In this work, we propose to model multipath error in the pseudorange as a combination of the multipath introduced, on one hand, by the vehicle structure and antenna installation and, on the other hand, by the additional reflections of surrounding buildings and objects. First, we derive a safe multipath error model for the antenna installation on the train roof. Second, we use the obtained error model as a reference model to normalize and characterize multipath in along the railway tracks. The role of different Fault Detection and Exclusion (FDE) algorithms in ensuring model compliance is also evaluated. Finally, the methodology is evaluated and tested with real railway dataset. The characterization of both the permanent train multipath and the operational environment multipath contributes to a safe and robust complete GNSS error description for safety critical navigation in the railway environment.

I. INTRODUCTION

Recent developments in railway localization systems have targeted replacing track-side based signaling systems, such as balises, with an on-board based localization solution. Most of the on-board localization system proposals have leaned towards implementing a fusion solution of different sensors, with Global Navigation Satellite Systems (GNSS) playing a key role. On-board based localization systems can potentially reduce installation and maintenance costs of signaling systems compared to infrastructure ones as well as ensure cross-country interoperability, while at the same time maintaining the accuracy and robustness needed for safe operations. However, for safety related operations, apart from the position information, it is required to have the information about its integrity. In some applications, such as civil aviation, the information about the trustworthiness or integrity of the localization solution is ensured by using augmentation systems. However, GNSS positioning can be more challenging in railway environments than in civil aviation applications. For that reason, using augmentation systems that correct or check the integrity of system and Signal in Space (SiS) errors is not sufficient. The environment in which railways operate regularly range from dense urban areas to suburban and rural areas and GNSS-based localization might be highly affected by various local GNSS threats such as multiple reflected signals (i.e., multipath), Non-line-of-sight (NLOS) signals or interferences caused by fast changing and dynamic environment. This poses a big challenge to determine safe and robust models for local GNSS errors, so that the existing algorithms can be used to correctly quantify the positioning error and ultimately its integrity.

In order to safely quantify the positioning error, all the different GNSS error sources have to be carefully modelled for the railway environment and suitable detection and exclusion methods have to be designed.

The most commonly used techniques for multipath and NLOS fault detection and exclusion are: Code Minus Carrier (CMC) monitor, Code Minus Carrier time Difference (CMCD) Monitor (Caamano et al., 2020), Signal to Noise Ratio Monitor (Marais et al., 2014) and Cycle Slip Detector (Sanz Subirna & Juan Zornoza, 2013). Additionally, a railway map can provide a topological position and consequently improve localization by placing the posit within the track areas (García Crespillo et al., 2014). The main advantage of dealing with multipath error at the post-correlation stage is that algorithms can be implemented with commercial of-the-shelf antennas and receivers, and they do not rely on complex techniques, this makes them cost efficient solutions. However, the detection and estimation of multipath errors at the post-correlation level usually require a certain level of a-priori statistical knowledge or assumptions about the nominal error and fault models. For land applications, it is in general difficult to obtain reliable models due to the complex and varying environments that are typically found. That is the case since, according to (Groves, 2013), signal reflections in land applications are primarily attributed to the surrounding environment, including factors such as ground surfaces, nearby buildings, structures, vehicles, and trees. However, multipath errors caused by the vehicle structure and antenna installation is also significant. This is further explored in (Kliman & García Crespillo, 2022), where it was shown that multipath highly depends on the specific antenna installation and the imminent surroundings of the antenna (e.g. train rooftop). Therefore, considering the antenna installation multipath could allow for the design of more robust error model and fault detection algorithms.

In this work, we propose to model multipath error at the pseudorange level as a combination of multipath caused, on one hand, by the vehicle structure and antenna installation and, on the other hand, by the additional reflections of surrounding buildings and objects. The multipath caused by the antenna installation and the vehicle structure (e.g. train rooftop) in this work is referred to as (permanent) train-side multipath, while the multipath caused by the surrounding objects is referred to as the operational environment multipath, as shown by Figure 1.

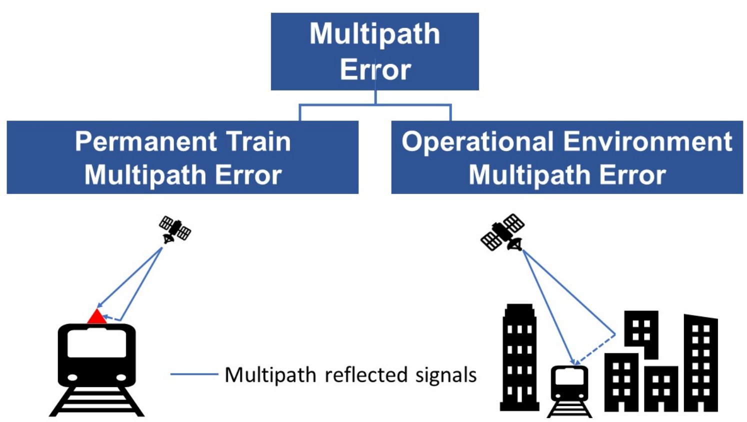


Figure 1: Proposed mulipath error separation.

Multipath error caused by the surrounding operational environment is expected to be changing during the train's operation, while multipath caused by the vehicle structure and antenna installation is expected to have similar stochastic properties for a given satellite elevation and azimuth with respect to the user. Although it is arguable whether physical multipath effect can be decomposed in this way, this allows us to develop a practical methodology to model the total multipath error in post-correlation. In other words, because of the varying nature of multipath decomposing the total multipath error contribution into two parts allows us to model separate parts of the multipath error which is otherwise still challenging to model. In addition to modeling the separate, distinct multipath errors in order to obtain a multipath error model which is not overly conservative, which would negatively affect the availability of measurements, in other words, in order to model the nominal multipath error, we are applying different Fault Detection and Exclusion (FDE) algorithms on the dynamic set of data. Those mainly being Code Minus Carrier Delta (CMCD) (Caamano et al., 2020) with a Cycle Slip detection and Signal to Noise Ratio threshold (Strode & Groves, 2016). The impact the FDE algorithms and their sensitivity have on the multipath outliers detection and exclusion is also introduced.

In this paper, we compare the impact of different fault detection and exclusion techniques and propose a methodology to derive a conservative error model of the permanent train-side multipath error considering time correlation. Finally, the methodology is evaluated and tested with real railway datasets obtained within the European Union project Railgap.

II. METHODOLOGY

The graphic representation of the steps is given in Figure 2.

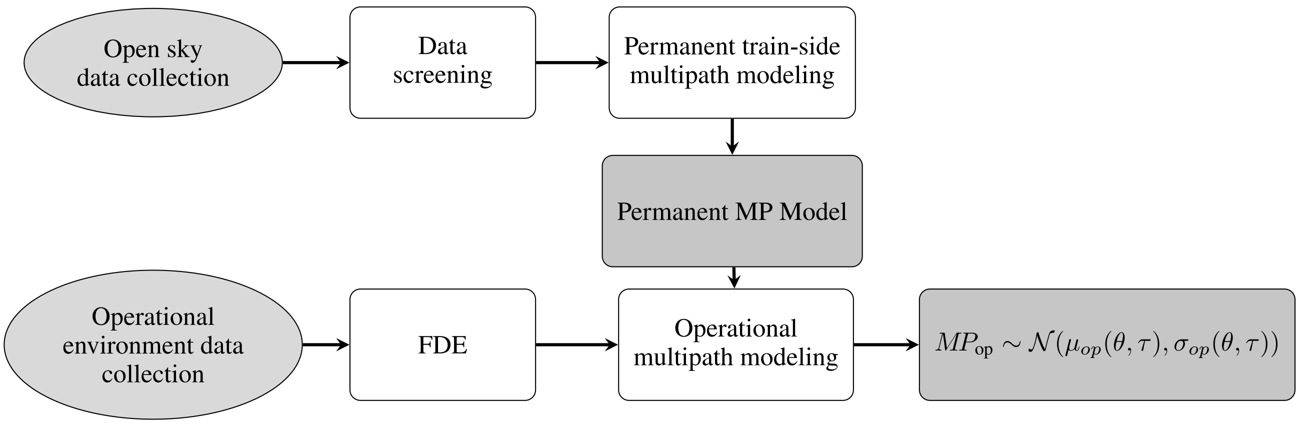


Figure 2: Overview of processing steps.

The methodology can be considered to be divided into two separate units; the permanent train-side multipath error modelling and the operational environment multipath modelling. Both units have the same basic procedure flow: first, an appropriate data collection is introduced, followed by a data screening or Fault detection and exclusion (FDE). The output of the permanent train-side multipath modelling is an elevation dependent multipath error model of the permanent train-side multipath. This model is used as a reference model and as an input to the operational multipath modelling. The result of the operational environment multipath model is both elevation and location dependent. The two steps are described in more detail in the following subsections.

1. Permanent Train-side Multipath Error

Figure 3 gives a graphic overview of different methodology steps that are considered within the *Permanent train-side multipath modeling* block of the Figure 2.

In order to model the permanent train-side multipath behaviour, it is first necessary to obtain GNSS measurements that are not affected by the varying multipath observed during railway operations. This can be done by collecting data in a static, open-sky scenario as suggested in (Kliman & García Crespillo, 2022).

By collecting data in an open sky, static environment, we try to minimize the contribution of the additional multipath introduced by the surroundings of the train.

This is further ensured by an additional data screening process. This step entails the inspection of the measurement that are affected by the environment around the train since placing a train in a perfect open sky environment is very challenging. The health flag of the satellites is checked as well as the availability of dual frequency as an additional healthy criteria. In addition, we also check the availability of both code and phase measurements.

After the *Data screening*, isolation of the permanent train-side code measurement multipath error has to be performed. To that end Code-Minus-Carrier (CMC) is used. CMC is the difference between pseudorange and carrier-phase measurements and can be used as an observable for the multipath error, thermal noise error and ionospheric error Braasch, 1994. This is due to the fact that the range, satellite and receiver clock biases and tropospheric delay are common to both measurements (Marais et al., 2000).

It is expressed as a difference of pseudorange measurements ρ_k^s and corresponding carrier phase measurements ϕ_k^s for satellite s at time k. As shown in (Caamano et al., 2020), for a single receiver, satellite s and frequency j can be expressed as follows:

$$CMC_k^s = \rho_k^s - \phi_k^s = MP_{\rho,k}^s + (2I_k^s - \lambda N_k^s) + (\varepsilon_k^s - mp_{\phi,k}^s - \xi_k^s), \tag{1}$$

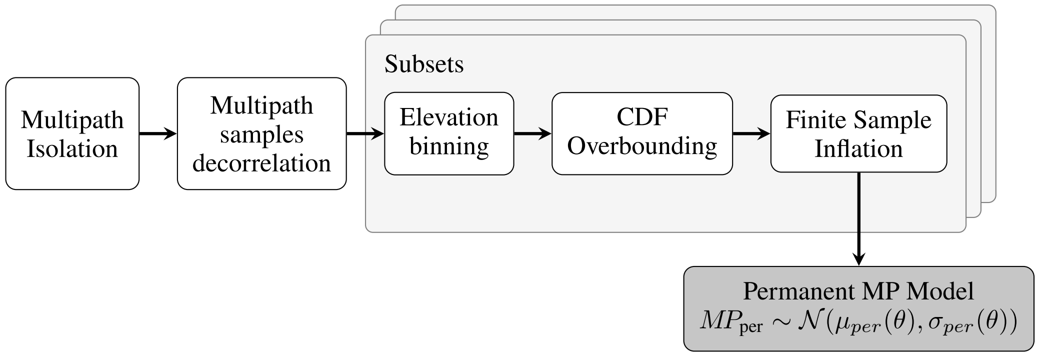


Figure 3: Overview of the *Permanent train-side multipath modeling* steps.

The CMC method removes all the code and carrier phase measurements common error terms. It is commonly assumed that carrier-phase multipath and noise are significantly smaller in magnitude compared to code multipath and noise and they can be neglected. The ionospheric error can be estimated and removed as a linear combination of two carrier-phase measurements received for the same satellite as shown in for example (Circiu, 2020).

The resulting linear combination of carrier-phase ambiguities can be eliminated by leveraging the fact that integer ambiguities remain constant during a continuously tracked cycle-slip free period (Caamano et al., 2020), allowing for bias estimation and removal. From that comes the following expression for satellite s, and time period K:

$$\widehat{MP}^{s}_{\text{per},k} = CMC^{s}_{\text{Dfree},k} - \frac{1}{K}\sum_{p=k-K}^{k} CMC^{s}_{\text{Dfree},p}. \tag{2}$$

The isolated multipath is obtained from code and phase measurement time series. It is therefore expected that multipath samples are time correlated. The derivation of sample distributions (or the computation of a variance) based on correlated samples would lead to a bias estimation. Samples must be therefore selected to guarantee they are decorrelated (Kliman et al., 2024). That is done by selecting samples separated by at least their decorrelation time. The decorrelation time is determined based on the normalized autocorrelation of all the time series. The necessary time lag is estimated by finding the time lag for which their normalized median autocorrelation is within a threshold of ± 0.2. The 0.2 threshold is typically in statistics considered as the limit for weak correlation (Circiu, 2020).

Next steps are repeated separately for each of the newly created sets of independent samples as it can be seen in (Kliman et al., 2024). Firstly, elevation dependent binning is performed, since multipath errors are highly elevation dependent. Secondly, a cdf overbound for every elevation bin is calculated following the methodology in (Blanch et al., 2019) in order to obtain a final parametric model.

The last step performed for every subset is an inflation factor based on the number of available finite samples.. The final model variance is obtained as the median variance from the different subsets (Circiu, 2020).

$$MP_{\text{per}} \sim \mathcal{N}(\mu_{per}(\theta), \sigma_{per}(\theta)). \tag{3}$$

By doing so we obtained an elevation dependent model, considering sample correlation, elevation dependent and constellation and frequency specific of the permanent train-side multipath error.

2. Fault Detection and Exclusion Algorithms

A set of measurements collected across different runs are used to evaluate the trains operational environment. Which are in the Figure 2 labeled as *Operational environment data collection*. Followed the operational environment data collection is the Fault Detection and Exclusion (FDE). The FDE is an important step since by not applying any FDE the outliers are also entering the modeling which would lead to a conservative model. Implementing FDEs allows us to tighten the model, avoiding issues with availability and it allows us to model the nominal error distribution. An overview of different FDE that can be applied for railway applications can be found in (García Crespillo et al., 2018). When applied per channel, multiple FDE can be applied in a cascaded fashion (Figure 4) to improve the overall detectability. However, even though we have used the SNR threshold and the CMCD with Cycle slip as FDEs for evaluations in Section III, other FDS could also be implemented in the general proposed methodology, and that is represented by dots in Figure 4.

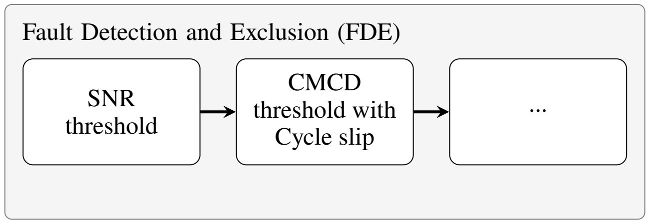


Figure 4: Fault Detection and Exclusion processing.

SNR threshold. The Signal-To-Noise (SNR) ratio is an observable available in GNSS receivers (including the commercial ones) and is often utilized to detect the presence of interference (Bhuiyan et al., 2014). It is a metric that quantifies the strength of a satellite signal relative to the background noise at the receiver.

The satellite elevation angle has a great impact, since lower elevation signals have been travelling longer through the atmosphere, and hence, encounter a stronger signal attenuation. On the other side, the antenna gain and immediate installation setup play a crucial role in the signal attenuation and multipath (Kliman & García Crespillo, 2022), which affects the SNR. Finally, the specific receiver tracking loops configuration may be also be a factor. In other words, measurements with lower SNR are more likely to be considered outliers (Marais et al., 2006), which are more frequent at lower elevations. In this paper, the SNR is employed as a threshold, which does not require nominal calibration.

Code-Minus-Carrier Difference Detector. Multipath or NLOS reception most likely have a different effect on code and phase measurements. Hence, one common approach is the investigate the difference of both in order to isolate and detect multipath and NLOS events, the so called Code-Minus-Carrier (CMC) (Caamano et al., 2020). One main problem associated with CMC as shown in (Blanco-Delgado & de Haag, 2011), is the need to remove the contribution of the carrier phase integer ambiguity. This is normally achieved by removing the average of CMC estimates over a long satellite observation period, however it might be difficult to achieve in urban scenarios with constant loss of tracking. Therefore, using the time-difference of the CMC estimates might be beneficial, the so called CMC Difference (CMCD). By differentiation the CMC over time, many error quantities are removed, such as the integer ambiguities and others reduced to their time-variation quantities.

$$\Delta CMC_k = \frac{1}{\Delta t}(CMC_k - CMC_{k-1}) = 2\dot{I}_k^s + \dot{MP}_k^s - \dot{mp}_{\phi,k}^s, + \dot{\epsilon}_k - \dot{\xi}_k^s, \tag{4}$$

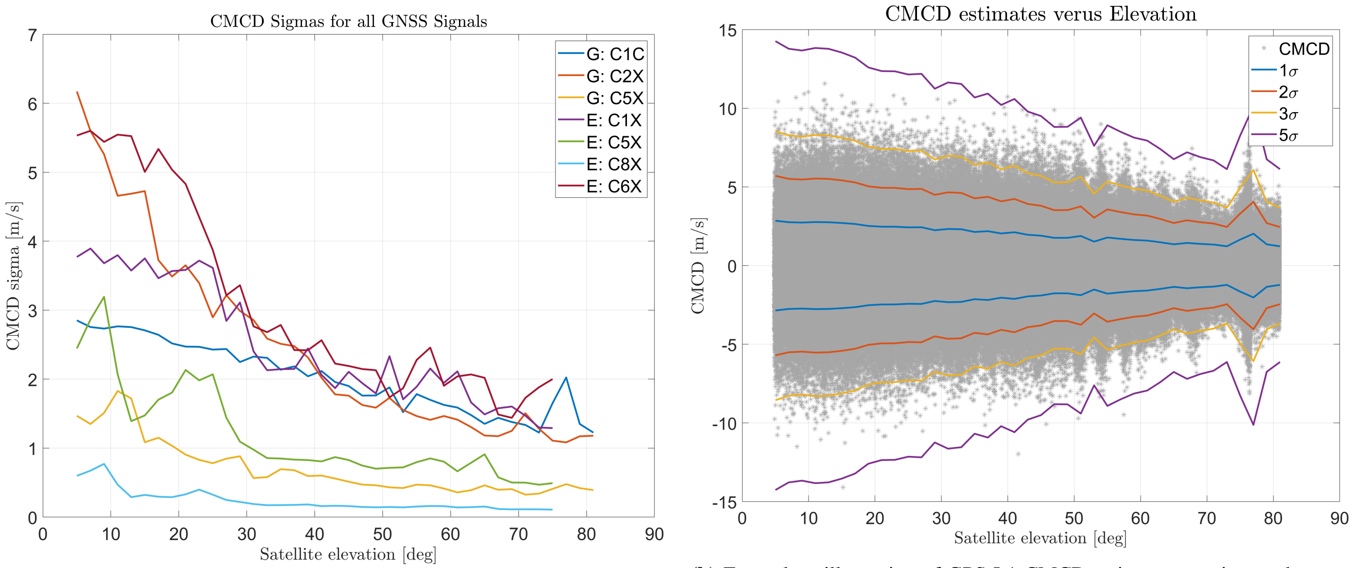
where Δt is the time difference between two consecutive measurement epochs, $2\dot{I}_k$ represents twice the ionospheric delay rate, $\dot{MP}_k^s$, $\dot{mp}_\phi^s$, $\dot{\epsilon}_k$ and $\dot{\xi}_k^s$ are the rates of multipath and noise of code and phase, respectively. As seen in (Jiang et al., 2017), the ionospheric rate component is negligible in nominal conditions in comparison to the expected uncertainty of the rate of multipath and noise. Additionally, the contribution of the phase multipath and noise can be assumed to be negligible compared to the code. Therefore, the CMCD can be directly represented as noisy rate of multipath change over time:

$$\Delta CMC_k \approx \dot{MP}_k^s + \dot{\epsilon}_k \tag{5}$$

Similar to the multipath isolation described previously, in the observed CMCD, there is also a quantity caused by the installation itself, hence, a calibration for each installation is needed. This calibration is done for a long-term static, open-sky scenario. The results for different GNSS signals are depicted in Figure 5a. The threshold of the CMCD is selected based on an inflation of these calibration results and is satellite elevation depend. This can be seen exemplarily for GPS L1 signals observed in Figure 5b. When the observed CMCD estimate exceeds this threshold a fault is triggered and both code and phase measurements are discarded from further processing. Evidently, the higher the threshold is set the less measurements are discarded. To indicate the drop of measurement availability for different thresholds, Figure 6b shows the number of measurement exclusions due to different CMCD inflation factors. That is by selecting the threshold to be twice the calibration results could lead to an exclusion of 25% of all Galileo E5a measurements received.

Please note that these results is only valid if no phase cycle slip has occurred during two consecutive epochs. Hence, the usage of a cycle slip detector is mandatory before the CMCD detector can be applied.

Cycle Slip Detector. Cycle slip is perceived as jumps of integer number of wavelengths. Different methods can be used for cycle slip detection, operating over undifferenced, single-differenced or double-differenced measurements between pairs of satellites (Sanz Subirna & Juan Zornoza, 2013). The cycle slip detector of choice, Melbourne-Wübbena (MW) combination, is based on code and carrier phase measurements. The MW combination provides a noisy estimate of the wide lane ambiguity (Sanz



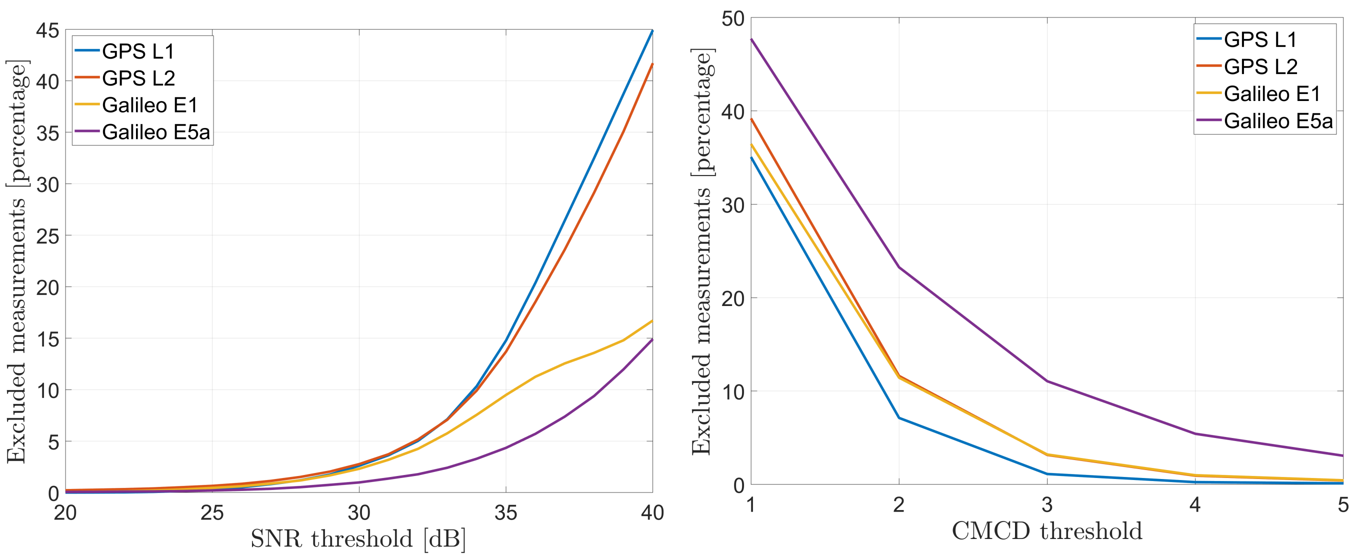
(a) CMCD Calibration result for different GNSS signals.

(b) Exemplary illustration of GPS L1 CMCD estimate over time and corresponding CMCD thresholds based on different inflation factors.

Figure 5: CMCD Detector calibration and usage.

Subirna & Juan Zornoza, 2013). Even though in this case we have only analysed the code measurements, we have used the cycle slip detector and removed the instances (epochs) of code measurements from further processing in the epochs when the carrier phase cycle slip was detected.

However, it is important to keep in mind that using SNR threshold test and joined the CMCD and Cycle slip detector can lead to a significant decrease of available measurements. Figure 6a and Figure 6b show the percentage of measurements that are being excluded after applying different SNR and CMCD thresholds, respectively. This emphasises the importance of considering the total number of measurements observed when selecting the SNR threshold and the CMCD inflation factor, since high thresholds exclude large number of measurements.



(a) Percentage of measurements exclusions for different SNR thresholds.

(b) Percentage of measurements exclusions for different CMCD inflation factor thresholds.

Figure 6: Measurements exclusions with different FDE.

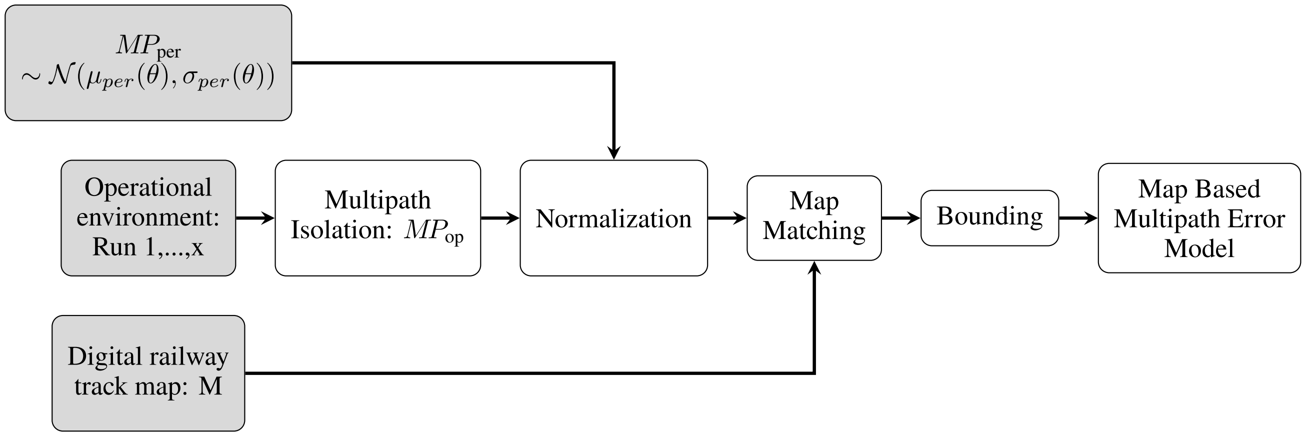


Figure 7: Overview of the operational multipath MP_{op} modeling steps.

3. Operational Environment Multipath Error

The second component of our error model is the operational environment multipath MP_{op} observed during operation within the tracks. In this work, we consider the derivation of a *map of multipath* along the tracks following the approach in (Roessl et al., 2023). A description of the location specific error can be obtained by collecting GNSS data at different times and hence different realizations, since the error shows a certain degree of repeatability for the same location. Figure 7 illustrates the involved steps of the modeling process.

Normalisation. We model the error observed along the tracks to be dependent on both the permanent and the operational components. Hence, for isolating the operational error we propose to eliminate the contribution of the permanent train-side error via a normalization step:

$$MP_{\text{op}} = \frac{\widehat{MP}}{\overline{\sigma}_{\text{per}}(\theta)}, \tag{6}$$

with $\overline{\sigma}_{\text{per}}(\theta) = \sigma_{\text{per}}(\theta) + b_{\text{per}}(\theta)$.

We presume that by repeating the process for several train runs, one can observe different satellite geometries and thus different elevation and azimuths and location specific characteristics causing multipath, despite slight changes in the environment, like foliage of the trees or passing trains.

Map-Matching. In a subsequent step, the observations MP_{op} are projected into a discrete digital track map exploiting the fact that a train is restricted to move only on the tracks, similar to the work of (Gerbeth et al., 2020). A train's position is generally defined by a topological railway location $\boldsymbol{\tau} = \{id, s\}$ on a track map, identified by a track identifier id and an along-track position s (Allan, 2000). Digital railway maps are increasingly used to support digital operations like map matching and train localization (Heirich, 2020). These maps typically include three layers: topological, geometrical, and sensor-specific information. Considering those three layers every track point $\mathbf{m}$ in a discretized digital map M can be characterized by:

$$\mathbf{m} = \{\boldsymbol{\tau}, \mathbf{x}, \boldsymbol{\iota}\}$$

The entire map M is then a set of track points: $\text{M} = \{\mathbf{m}_1, \ldots, \mathbf{m}_n\}$. The goal is to assign every observed point to one of the points in the discretized digital map. However, since a train's GNSS position may not align perfectly with the track map due to different errors, like track map and measurement errors, the nearest track points are identified through map-matching based on a horizontal radius to account for the uncertainties (Roessl et al., 2023).

Bounding. An empirical error distribution can be obtained as a track segment is passed within several runs. For each track location, an empirical distribution of observations is used to quantify multipath errors, but this method is computationally heavy and complex. To simplify, the two-step Gaussian bounding methodology is employed just like with the overbounding of the permanent train-side multipath error in order to obtain a parametric error description. This Gaussian overbounding method provides a conservative error description for each discrete track point, facilitating efficient error quantification and storage. The resulting digital map presents the geo-referenced and topological coordinates along with the Gaussian bounds for each track point and is defined as:

$$MP_{\text{op}}(\boldsymbol{\tau}) \sim \mathcal{N}(\pm b_{\text{op}}(\boldsymbol{\tau}), \sigma_{op}(\boldsymbol{\tau})). \tag{7}$$

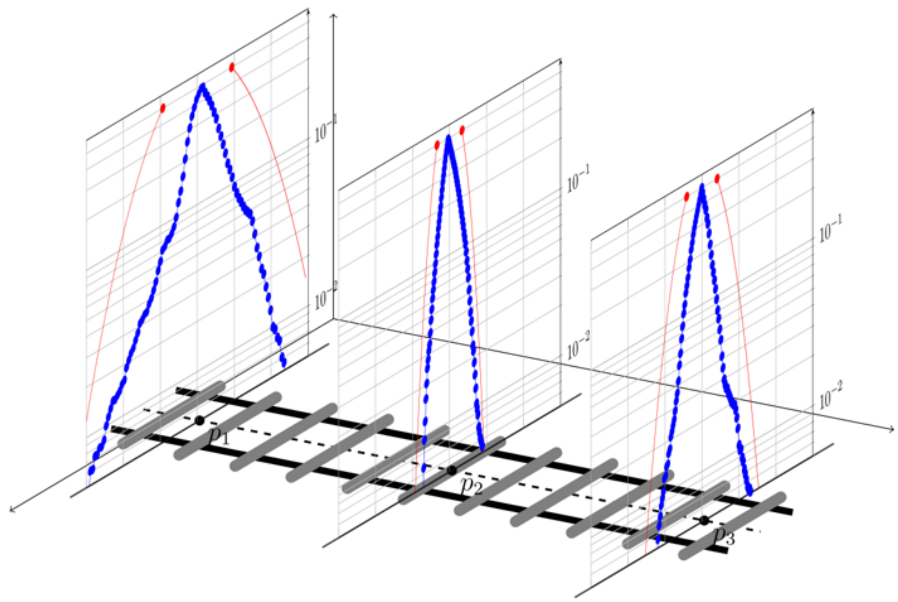


Figure 8: Digital track map based multipath error model providing an error description for each track point.

Figure 8 illustrates aggregated data distribution and resulting Gaussian overbound for the track points.

Map Based Multipath Error Model. The final model can hence be expressed as:

$$MP(\theta,\boldsymbol{\tau}) \sim \mathcal{N}(\pm\hat{b}(\theta,\boldsymbol{\tau}), \hat{\sigma}(\theta,\boldsymbol{\tau})), \tag{8}$$

with the bias $\hat{b}(\theta,\boldsymbol{\tau}) = \overline{\sigma}_{\text{per}}(\theta) \cdot b_{\text{op}}(\boldsymbol{\tau})$ and standard deviation $\hat{\sigma}(\theta,\boldsymbol{\tau}) = \overline{\sigma}_{\text{per}}(\theta) \cdot \sigma_{\text{op}}(\boldsymbol{\tau})$

III. EXPERIMENTAL SETUP

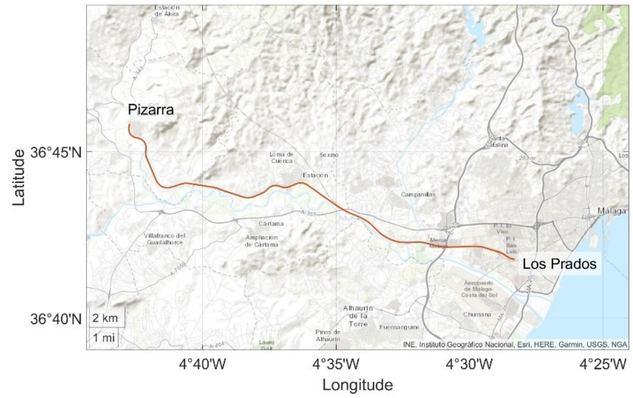
Results presented in this paper are obtained based on a data collected within the European Union project RAILGAP (RAILGAP, 2024). Data set was collected during a measurement campaign in Málaga, Spain. The train setup consisted of a maintenance diesel engine train Plasser Theurer draisine. The GNSS antenna installed on the train was an Antcom G8 antenna and it was connected with a Javad Delta-3N receiver. The receiver was collecting raw measurements with a sampling rate of 10 Hz. GPS L1, L2 and L5 as well as Galileo E1 and E5a measurements were collected.

The data for the permanent train-side multipath modeling was collected in an open-sky and static conditions during two consecutive data collections equaling to a total of around 23 hours. The location of the train during the data recording is shown in Figure 9a. During the collection of the data the train was placed in an as close as possible to an open-sky scenario, given the operational constraints of railway lines. That allowed for the isolated multipath error to be considered as the multipath caused exclusively by the antenna installation. Furthermore, the continuous recording enables the assumption that the integer ambiguities of the carrier-phase measurement can be successfully estimated and removed. Additionally, it allows us to obtain measurements from a variety of different elevation angles and azimuths.

For the dynamic measurements we selected a run from February, 2nd, 2024. The train left Pizarro train station at 12:17pm and arrived to Los Prados maintenance depot 30 minutes later. The selected run is around 25 km long and is depicted in Figure 9b. During the ride there were no tunnels and only short bridges, which allowed for a generally good satellite visibility.

In addition to the measurements collected during the RAILGAP project the methodology requires the map information and base station providing GNSS code and phase measurements for the same time span. For the latter a RTK (Real-Time Kinematic) station from HxGN SmartNet network was used. The station named MLGA reference station is located in Málaga, Spain. Please note the processing has been performed in post-processing, hence, real-time capabilities are not required. The base station measurements are used for the implementation of the Forward-Backward RTK/INS Smoothing solution, which is used for the ground truth estimation needed for the residual extraction.

(a) Train location during static measurements.



(b) Trajectory of the dynamic measurements.

Figure 9: Experimental data collection.

IV. RESULTS

1. Permanent Train-side Multipath Error Models

Following the methodology steps introduced in Figure 3, we first isolate the code multipath of the measured static, open-sky GNSS signals, exemplary for GPS L1 it is shown in Figure 10a and for Galileo E1 Figure 10c. Secondly, we derive the time lag of the correlated multipath estimates. The result of time lag estimation based on the normalized autocorrelation for exemplary GPS L1 data set is given in Figure 10b and for Galileo E1 in Figure 10d.

The time lag used for the decorrelation of the samples is the median time lag based on the normalized autocorrelation of all available time series's and equals in this case to 80 seconds for GPS L1 and 113 seconds for Galileo E1 measurements.

Following, we generated different subsets each containing only independent samples. The multipath estimates of each subset are separated into different elevation bins such that in each elevation bin we have the minimum number of 300 samples. That resulted in 19 elevation bins for GPS L1 and 15 for Galileo E1.

Next, we calculate the cdf overbound, resulting in a left and right hand bias and sigma estimate, we consider only the maximum of both estimates. Finally, we inflated this sigma estimate to account for the limited number of samples. We repeated the cdf overbound and the finite sample inflation for all generated data subsets of independent samples. Each sigma estimate was inflated with the exact inflation factor based on the number of samples in the corresponding elevation bin.

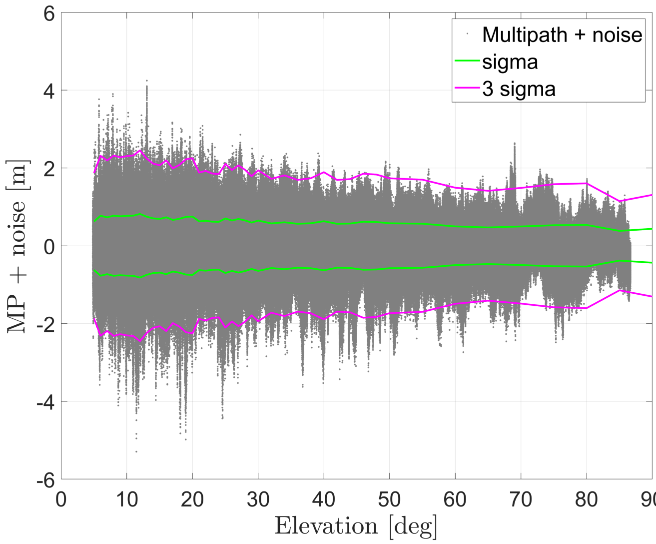
By computing the overbounding sigmas of all the subsets and inflating them to account for the limited number of samples, and finally providing the model based on only one of the independent subsets we insure that the one independent subset is selected after an insight to the full data set and that the limited number of samples are tackled by adopting the inflation factor as seen in (Kliman et al., 2024).

The final step of the methodology is selecting the sigma estimate for the final model selection, which was in this case the median sigma inflated estimate. The final models for all observed constellations and frequencies is given in Figure 11.

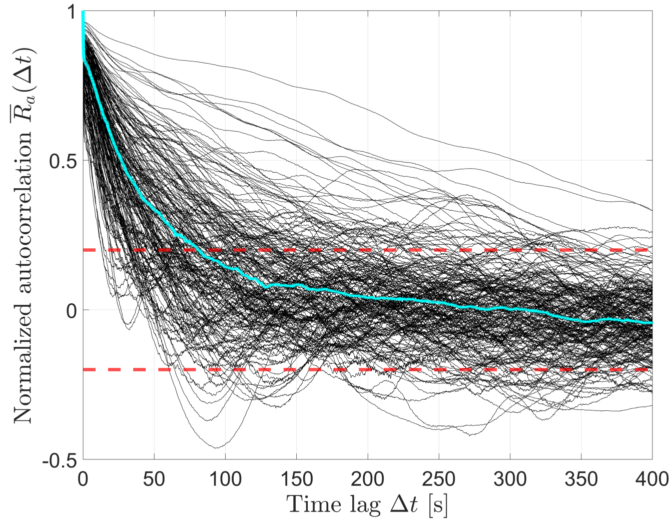
2. Impact of the Fault Detection and Exclusion Algorithms

In order to evaluate the impact of the different FDE detectors, we investigate the code measurement residuals obtained in comparison with the reliable trajectory, which has been obtained by a forward-backward RTK/INS smoothing process. This trajectory determination implies: complete FDE chain including Signal-to-Noise-Ratio threshold test and CMCD detector with Cycle slip detection. The forward and backward chain included a separate EKF FDE mechanisms consisted of a EKF innovation screening and normalized innovation square test. Based on the obtained trajectory, the Single Difference (SD) pseudorange residuals of the user and the base station are computed by taking into account precise satellite orbits, precise user clock and base station clock bias estimation, ionospheric correction and relativistic path range correction. By computing the SD pseudorange, the satellite clock bias is removed, the ionospheric and the tropospheric delays can be assumed to be removed presuming the same impact at the user and base station receiver due their proximity (short baselines). Additionally, the multipath error of the base station is considered to be negligible since the base station is a continuously operating reference station and as such it is installed in a way to avoid multipath errors.

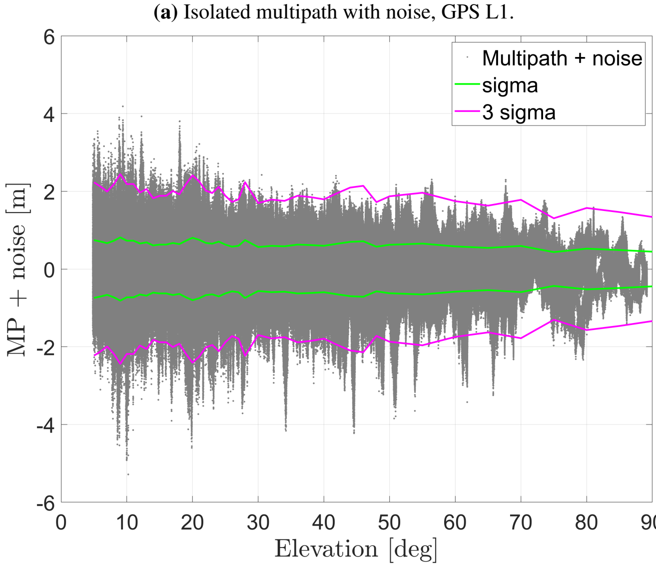
The impact of the GNSS FDEs is tested by enabling different detectors setup. Additionally, the obtained SD pseudoranges are normalized with the previously obtained permanent train-side multipath model. The investigated combinations of the FDEs



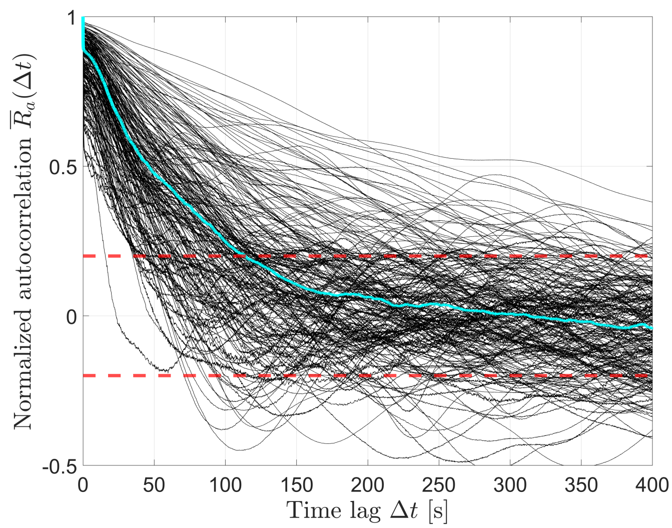
(a) Isolated multipath with noise, GPS L1.



(b) Normalized autocorrelation, GPS L1.



(c) Isolated multipath with noise, Galileo E1.



(d) Normalized autocorrelation, Galileo E1.

Figure 10: Isolated multipath with noise and Normalized autocorrelation for exemplary GNSS signals.

detectors in this paper are the following:

1. No GNSS FDE,
2. SNR threshold test,
3. CMCD detector (including Cycle Slip detector),
4. Full FDE chain (SNR threshold test, Cycle slip and CMCD detector).

To illustrate their impact on GPS L1 and Galileo E1, Figure 12 is shown the empirical folded cdf of the observed SD code residuals. The corresponding folded cdf of a normal distribution, with zero mean and variance one, is shown in the same figure by a dashed grey line. It is included to indicate whether the residuals of the used FDE chain are overbounded by the standard Normal distribution. It can be clearly seen that by applying different FDE the heavy tails are reduced, and the distribution is closer to a Normal distribution then when no FDEs are used. In addition, the cdf values lesser than 10^{-4} are not being investigated due to the limited number of samples.

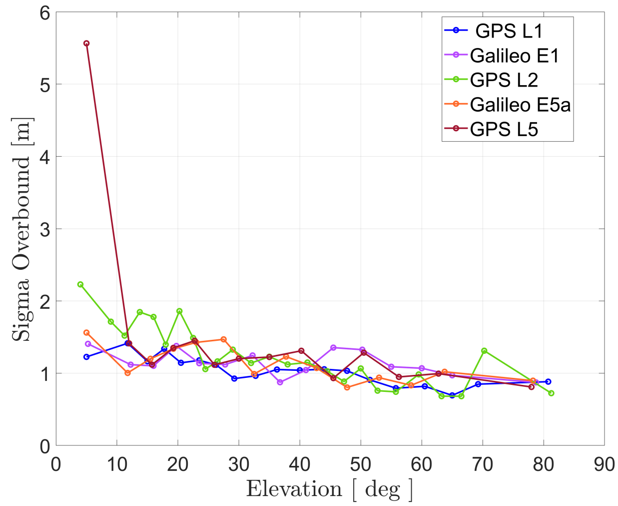
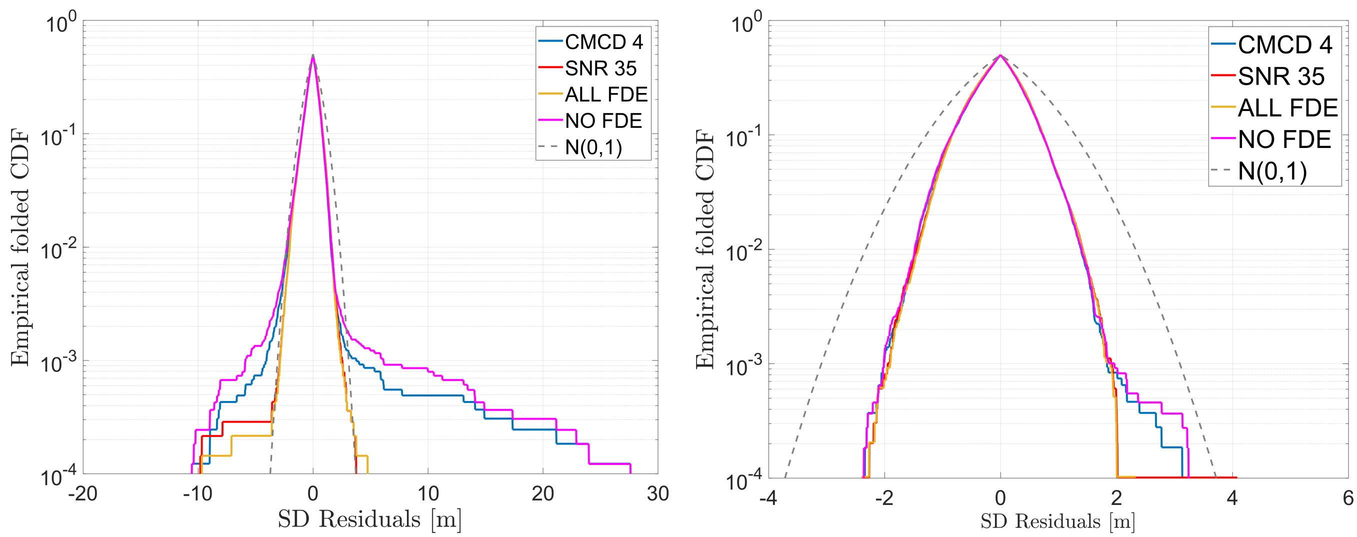


Figure 11: Final models of the permanent train-side multipath.



(a) Comparison of FDE impacts, GPS L1.

(b) Comparison of FDE impacts, Galileo E1.

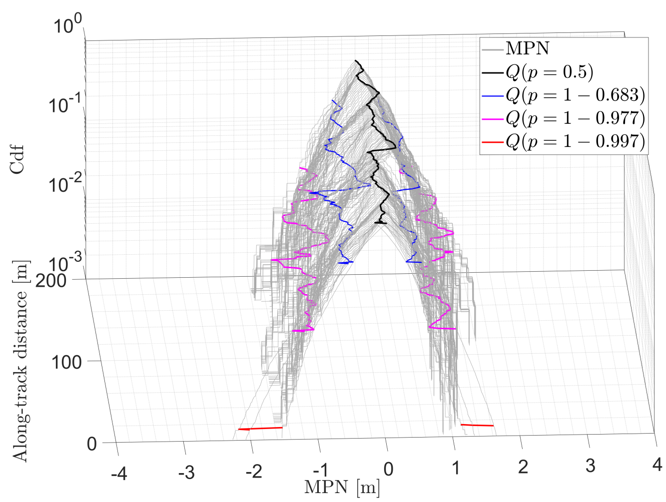
Figure 12: Comparison of FDE impacts for ememplar GNSS signals.

3. Operational Environment Multipath Error Models

Figure 13 depicts the resulting operational multipath distributions and associated error bounds for GPS L1, as obtained through the application of all FDE algorithms. Figure 13a illustrates the folded cdf of the normalized multipath and noise error, with the y-axis representing the along-track dimension. The variation in the distributions is indicated by the use of different empirical quantiles, which are coloured differently. For example, the median quantile (black) illustrates significant deviations, confirming the hypothesis for employing an overbounding characterization. Figure 13b shows an exemplary track map of the dataset colored with the resulting error bound of the operational multipath for GPS L1. The track points are colored with $\hat{b}_{\text{op}} + 1.96 \cdot \hat{\sigma}_{\text{op}}$. As expected, a location-dependent variation is observed, with larger error bounds occurring in built-up areas, such as the vicinity of bridges.

V. CONCLUSION

In conclusion, this work presents a methodology to derive a robust GNSS multipath error model, particularly tailored for railway environments. By separately characterizing the permanent train-side multipath effects and the operational environment additional multipath, the study provides a comprehensive full approach for assessing and modeling multipath errors. The



(a) Empirical distribution of projected and normalized multipath

(b) GPS L1 error bound

Figure 13: Projected operational multipath and resulting error bound for GPS L1 after applying all FDE.

developed multipath error model supports operational multipath assessment and can be used to improve robustness of PVT algorithms and ensure its integrity especially when combined with SBAS or ARAIM. Furthermore, the analysis of the impact of Fault Detection and Exclusion (FDE) algorithms on multipath error distribution highlights their significant role in improving the accuracy of GNSS positioning for railways, while at the same time not being too conservative and jeopardising availability. The methodology offers specific modeling parameters suited for the railway context, contributing to the safe and robust application of GNSS-based localization. Looking forward, the extensive dataset from the RailGap project presents an opportunity to further refine the results. Future work will focus on incorporating elevation dependent and frequency dependent thresholds and time-correlated models, which will enhance the reliability and applicability of the proposed multipath model for dynamic and complex railway environments for different users.

ACKNOWLEDGEMENTS

The data analysed in this work was collected during the European GSA H2020 project RailGap (Grant Agreement Number: 101004129) and the work was partially funded by the same project. Special thank you to the partners responsible of the data collection ADIF, CEDEX and INECO.

REFERENCES

Allan, J. J. (2000). *Computers in railways VII*. WIT Press.

Bhuiyan, M. Z. H., Kuusniemi, H., S., S., & Airos, E. (2014). The impact of interference on gnss receiver observables – a running digital sum based simple jammer detector. *Radioengineering*, *3*, 898–906.

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

Blanco-Delgado, N., & de Haag, M. U. (2011). Multipath analysis using code-minus-carrier for dynamic testing of gnss receivers. *2011 International Conference on Localization and GNSS (ICL-GNSS)*, 25–30. https://doi.org/10.1109/ICL-GNSS.2011.5955254

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

Caamano, M., García Crespillo, O., Gerbeth, D., & Grosch, A. (2020). Detection of GNSS Multipath with Time-Differenced Code-Minus-Carrier for Land-Based Applications. *2020 European Navigation Conference (ENC)*. https://doi.org/10.23919/enc48637.2020.9317340

Circiu, M.-S. (2020). *Integrity aspects for dual-frequency dual-constellation ground based augmentation system (gbas)* [Doctoral dissertation, RWTH Aachen].

García Crespillo, O., Heirich, O., & Lehner, A. (2014). Bayesian gnss/imu tight integration for precise railway navigation on track map. *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, 999–1007. https://doi.org/10.1109/PLANS.2014.6851465

García Crespillo, O., Konovaltsev, A., Marais, J., Sabina, S., Vennarini, A., Coluccia, A., Neri, A., Aguila, A., Razzano, E., & Ranauro, F. (2018). Local gnss threat detection methods for virtual balise placement in railway applications. *2018 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST)*, 1–7.

Gerbeth, D., García Crespillo, O., 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

Groves, P. (2013). *Principles of gnss, inertial, and multisensor integrated navigation systems*. Artech House.

Heirich, O. (2020). *Localization of Trains and Mapping of Railway Tracks*. Verlag Dr. Hut.

Jiang, Y., Milner, C., & Macabiau, C. (2017). Code carrier divergence monitoring for dual-frequency gbas. *GPS Solutions*, *21*, 769–781. https://api.semanticscholar.org/CorpusID:125909110

Kliman, A., & García 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 [To appear]. *2024 European Navigation Conference (ENC)*.

Marais, J., Flancquart, A., & Lefebvre, S. (2006). Satellite availability in a railway mountainous environment: Can we use satellite positioning for safety applications? *World Congress on Railway Research (WCRR), Montréal*.

Marais, J., Meunier, B., & Berbineau, M. (2000). Evaluation of gps availability for train positioning along a railway line. *Vehicular Technology Conference Fall 2000. IEEE VTS Fall VTC2000. 52nd Vehicular Technology Conference (Cat. No. 00CH37152)*, *5*, 2060–2067.

Marais, J., Meurie, C., Attia, D., Ruichek, Y., & Flancquart, A. (2014). Toward accurate localization in guided transport: Combining gnss data and imaging information. *Transportation Research Part C: Emerging Technologies*, *43*, 188–197.

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

Roessl, F., García Crespillo, 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.

Sanz Subirna, J., & Juan Zornoza, M., J. M.and Hernandez-Pajares. (2013). *Gnss data processing, volume 1: Fundamentals and algorithms*. ESA Communications.

Strode, P. R., & Groves, P. D. (2016). Gnss multipath detection using three-frequency signal-to-noise measurements. *GPS solutions*, *20*, 399–412.