# Robust Statistics vs. Machine Learning vs. Bayesian Inference: Insights into Handling Faulty GNSS Measurements for Robust Vehicle Localization

Haoming Zhang, *Member, IEEE*

***Abstract*—This paper presents research findings on handling faulty measurements (i.e., outliers) of global navigation satellite systems (GNSS) for vehicle localization under adverse signal conditions in field applications, where raw GNSS data are frequently corrupted due to environmental interference such as multipath, signal blockage, or non-line-of-sight conditions. In this context, we investigate three strategies applied specifically to GNSS pseudorange observations: robust statistics for error mitigation, machine learning for faulty measurement prediction, and Bayesian inference for noise distribution approximation.**

**Since previous studies have provided limited insight into the theoretical foundations and practical evaluations of these three methodologies within a unified problem statement (i.e., state estimation using ranging sensors), we conduct extensive experiments using real-world sensor data collected in diverse urban environments.**

**Our goal is to examine both established techniques and newly proposed methods, thereby advancing the understanding of how to handle faulty range measurements, such as GNSS, for robust, long-term vehicle localization. In addition to presenting successful results, this work highlights critical observations and open questions to motivate future research in robust state estimation.**

## I. Motivation

Vehicle navigation in ***large-scale field environments*** commonly depends on Global Navigation Satellite Systems (GNSS) to maintain globally consistent state estimates. However, in complex environments such as urban canyons and forest, GNSS signals are frequently degraded by multipath propagation and non-line-of-sight (NLOS) conditions. These effects lead to measurement outliers with time-varying noise characteristics that are often ***heavy-tailed***, ***skewed***, and ***multi-modal***, making them difficult to model a priori [1].

Consequently, state estimators based on least-squares techniques may degrade or even diverge when relying on inconsistent measurement models with overconfident, hand-crafted noise parameters, as they oftentimes underestimate the complexity and magnitude of uncertainty. As a result, developing a vehicle localization approach that ensures robust ***long-term operation*** across diverse environments remains a significant challenge.

### A. Related Work

Given that faulty sensor observations pose a critical challenge to trustworthy autonomy in field applications, state-of-the-art methods typically address this issue by leveraging five key concepts:

Haoming Zhang was with the Institute of Automatic Control, RWTH Aachen University, Aachen, Germany, and now with the Learning Systems and Robotics Lab, Technical University of Munich, Munich, Germany, and with the Munich Institute of Robotics and Machine Intelligence (MIRMI).

Corresponding: haoming.zhang@tum.de

*1) **Multisensor Fusion**:* Many approaches emphasize fusing additional sensor modalities alongside GNSS observations to enhance state observability in the presence of signal loss or to mitigate the effects of faulty measurements. Previous studies have extensively explored the integration of various local sensors into graph-based optimization frameworks, such as near-field range signals [2], lidar [3], cameras [4, 5], or multiple sensors [6], alongside GNSS for state estimation, particularly in SLAM applications.

In our previous work [7, 8], we proposed a generalized GNSS/multisensor state estimator based on continuous-time factor graph optimization by incorporating Gaussian process motion priors, originally introduced in [9]. The flexibility comes from the fact that we choose estimation timestamps independent of any particular sensor frequency. This feature, as well as the smoothing effect of a motion prior, provides robustness in the presence of any particular sensor dropout.

**Limitations:** Even when redundant state constraints are available through multiple sensor inputs, the state estimators reviewed above still assume that measurement noise follows a Gaussian distribution. This assumption tends to underestimate the complexity and magnitude of real-world uncertainty. As a result, they often struggle to effectively downweight faulty sensor observations. Without advanced engineering and fine-tuned parameters, the performance of multisensor fusion systems is compromised, limiting their scalability and reliability for long-term operations in large-scale environments.

Therefore, robust state estimation for handling faulty measurements (outliers) remains an active area of research.

*2) **Outlier Exclusion**:* Outlier exclusion serves as a discriminative strategy for handling measurement outliers in state estimation problems. Rather than relying on statistical modeling of the outlier distribution, this approach uses heuristic methods to detect and exclude faulty GNSS measurements.

Methods in this category typically employ consensus checking (e.g., RANSAC) [10] and integrity monitoring techniques [11], where two further principles are commonly employed: statistical testing [12], which evaluates measurement consistency under redundancy, and interval analysis [13, 14], which bounds uncertainties through conservative set representations to detect anomalies.

**Limitations:** RANSAC-like algorithms are known to become inefficient and unstable when dealing with high outlier rates ($> 50\,\%$) [15], and they struggle to scale effectively in high-dimensional problems [16]. Similarly, integrity monitoring assumes sufficient healthy measurements to maintain consistency. When outliers dominate or too few measurements are available, the estimated protection level becomes unreliable.

*3) **Robust Error Modeling**:* Robust error modeling presents another effective approach to outlier rejection, commonly achieved by shaping the cost function to enhance resilience against outliers.

In this category, M-estimation (maximum likelihood type estimates) has received considerable attention for rejecting GNSS measurement outliers [17, 18]. Their utility has also been validated in harsh environments with significant multipath and NLOS effects [19, 20].

While M-estimation is well-established, its effectiveness relies on selecting a well-parameterized kernel function. Recent work has focused on adaptive M-estimation, where parameters are tuned based on current error statistics. [21] proposed a self-learning Huber loss using GNSS residuals, showing slight performance gains. [22] introduced a two-stage method to tune the generalized kernel from [23], leveraging past measurements to adapt to outlier behavior.

**Limitations:** Most studies using non-adaptive M-estimators focus on hard-redescending kernels like Huber and Tukey, overlooking the potential of soft-redescending functions for robust GNSS-based vehicle localization. They also lack analysis of parameter tuning and the trade-off between robustness and efficiency. Although some works [24]–[26] evaluate various kernels and settings, they are not tailored to range-based localization.

Adaptive M-estimation methods are typically limited to simple, symmetric noise models and rely on past residuals, which may fail when noise patterns shift abruptly. In low-visibility GNSS scenarios, estimator efficiency becomes as crucial as robustness.

These gaps highlight the need for a comprehensive study of M-estimators across categories, focusing on their mathematical properties and suitability for robust GNSS-based vehicle localization.

*4) **Bayesian Inference for Distribution Approximation**:* Another recent work has tackled measurement outliers by integrating robust error modeling into the Variational Bayes framework. In simpler cases, measurement noise is modeled with a conjugate prior (e.g., Gaussian) from the exponential family, enabling efficient inference [27, 28].

When noise exhibits complex characteristics, such as heavy tails, multimodality, or asymmetry, beyond the capacity of exponential-family priors, mixture models serve as universal approximators. Gaussian mixture models, in particular, have been widely applied to fit noise distributions from measurement residuals, effectively capturing diverse noise patterns [29].

**Limitations:** However, approximated noise distributions are often complex, nonlinear, and non-convex, making it difficult to extract reliable covariance estimates for least-squares solvers, potentially leading to suboptimal results.

Moreover, sampling measurement residuals poses challenges due to the lack of ground-truth errors. While prior work [30] addresses state uncertainty in residuals, distribution shifts can still occur, especially in dynamic, online settings like vehicle motion. Additional challenges, such as sampling efficiency and its context awareness, remain open research questions.

*5) **Learning-based Methods**:* As environment-dependent outliers may be challenging to identify and can exhibit complex noise patterns that are difficult to model with traditional statistical methods, learning-based methods[1] have also gained attention for handling faulty GNSS observations.

Pre-trained models enable online GNSS fault detection. While classic machine-learning methods can also achieve high accuracy [31, 32], deep learning (DL) better handles time-correlated, nonlinear GNSS data by extracting semantic features and improving generalization [33, 34]. In our previous work [35], we proposed a novel neural network architecture to predict measurement outliers following the noise characteristics of the GNSS observations. In addition, we provided an in-depth analysis comparing classical models with deep networks, as well as evaluating performance on in-distribution and out-of-distribution data.

**Limitations:** However, while previous studies highlight the potential of learning-based methods for handling GNSS outliers, they have not thoroughly evaluated their real-world feasibility, limitations, or generalization across diverse datasets and environments.

Moreover, as there is currently no evidence that learning-based methods can generalize across varying scenarios in vehicle localization, further investigation is needed into how learning concepts can be effectively integrated to enhance the robustness of state estimation applications.

### *B. Objective*

Recognizing the limitations of the aforementioned studies, this paper summarizes the author's key research findings and lessons learned on handling faulty GNSS measurements and aims to evaluate the feasibility and limitations of three prominent method categories using data from real-world measurement campaigns.

In particular, we outline the following objectives:

- **Analyze various M-estimators** for GNSS-based localization, aiming to develop a systematic understanding of how different kernel functions balance robustness and efficiency.
- **Evaluate offline learning-based methods**, including both classical machine learning models and deep neural networks, to predict faulty GNSS measurements. We provide an in-depth comparison between these model types and assess their performance on both in-distribution and out-of-distribution datasets.
- **Validate online variational Bayesian inference** for approximating GNSS pseudorange noise distributions, with a particular focus on distribution shifts in residual samples and identifying remaining open challenges.

## II. Preliminaries

### *A. GNSS Pseudorange Model for Vehicle Localization*

In localization approaches that tightly fuse the GNSS observations, pseudorange is commonly used [36]. The pseudorange

[1]Although many learning-based methods are rooted in statistical inference and thus share similar theoretical foundations, we intentionally categorize inference-based and learning-based methods as distinct approaches.

$\rho$ represents a geometric distance between the phase center of the GNSS antenna and the associated satellite, which contains several range delays due to satellite orbit bias and atmospheric delays. It can be modeled with respect to the antenna position as

$$\rho_k = \underbrace{\left\|\boldsymbol{p}^e_{\text{ant}} - \boldsymbol{p}^e_{\text{sat},k}\right\|}_{\text{1-D geometric range}} + \underbrace{c_b - c_{b,\text{sat}} + T + I}_{\text{assumed to be known}} + \underbrace{M + w_{\rho,k}}_{\text{unknown (Fig.1a)}} , \tag{1}$$

where the vectors $\boldsymbol{p}^e_{\text{ant}}$ and $\boldsymbol{p}^e_{\text{sat},k}$ represent the positions of the GNSS antenna and $k$-th satellite in the Earth-centered, Earth-fixed (ECEF) frame, respectively. The variables $c_b$ and $c_{b,\text{sat}}$ represent the bias due to receiver clock delay and satellite clock delay. The tropospheric, ionospheric and multipath delays are denoted as $T(t)$, $I(t)$, and $M$. The pseudorange noise is $w_\rho$.

Throughout this research, we assume that all clock errors and atmospheric delays can be estimated or eliminated using additional correction data.

### B. GNSS Measurement Corruption

Fig. 1 presents an overview of GNSS measurement corruption, illustrating various sources of error in pseudorange observations. These include delays induced by satellite and receiver clocks, signal propagation through the ionosphere and troposphere, as well as instrumental and infrastructure-related effects (see Fig. 1 a). While some of these errors, such as clock offsets or atmospheric delays, can be accurately eliminated, others, like multipath and NLOS receptions, introduce non-deterministic disturbances that are difficult to eliminate. Fig. 1b further distinguishes between line-of-sight (LOS) and NLOS scenarios, highlighting how environmental obstructions and reflections can significantly corrupt the received signal.

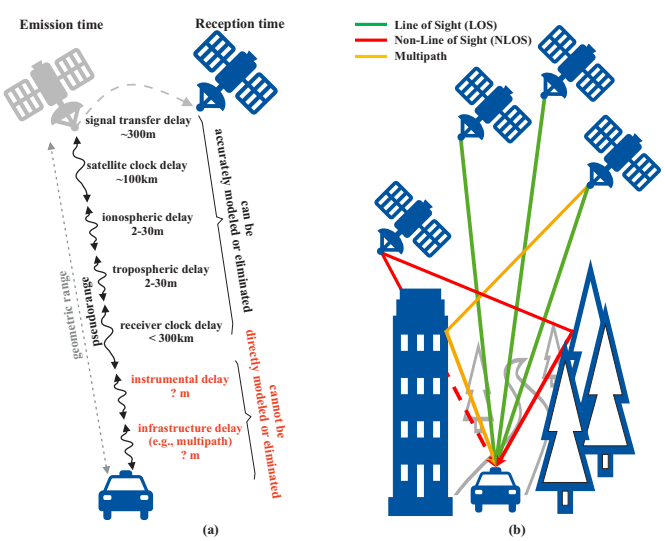



**Figure 1:** GNSS delays and interference.

## III. Methodology

### A. GNSS/Multisensor Fusion

To develop and evaluate these methods, we utilize a novel multisensor state estimator based on continuous-time factor graph optimization (FGO) [8], which serves as a flexible testbed for, e.g., deriving GNSS measurement residuals used in training and assessing the proposed uncertainty models.

### B. Error Mitigation using M-Estimators

In robust statistics, M-estimators offer a fixed-form, easily implemented approach that integrates prior information more simply than L- and R-estimates, while enabling simultaneous outlier detection during estimation [25].

Recalling the nonlinear least-square optimization formulated within a *maximum-a-posteriori* (MAP) problem, where the cost function takes the form

$$J(\boldsymbol{e}) = \|\boldsymbol{e}\|^2 \triangleq \|\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})\|^2_{\boldsymbol{\Sigma}} . \tag{2}$$

Under the Gaussian noise, the cost function in (2) becomes quadratic, commonly referred to as the $L2$ (non-robust) loss.

In this context, M-estimators aim to reshape the quadratic loss by introducing a symmetric, positive-definite kernel function $\rho(\cdot)$ with a unique minimum at zero. This modifies the original MAP problem as follows

$$\hat{\boldsymbol{X}}^{\text{MAP}} = \underset{\boldsymbol{X}}{\text{argmin}} \sum_{k=0}^{K} \rho(\boldsymbol{e}_k). \tag{3}$$

The kernel function $\rho(\cdot)$ in (3) is typically designed to reduce the influence of a large residual $\boldsymbol{e}$, which is often caused by outliers. Numerous kernel functions[2] are available in M-estimation, typically classified by their influence functions into three categories: *robust-monotonic*, *robust-soft-redescending*, and *robust-hard-redescending*.

Two key concepts, *robustness* and *relative efficiency*, are used to evaluate M-estimators, with a trade-off noted by [37]: "The more robust an estimator is, the less efficient it is". Consequently, relative efficiency serves as a fundamental reference for tuning M-estimators.

For an in-depth study of handling GNSS outliers using M-estimators, we select six representative M-estimators, each parameterized at four efficiency levels, as summarized in Table I.

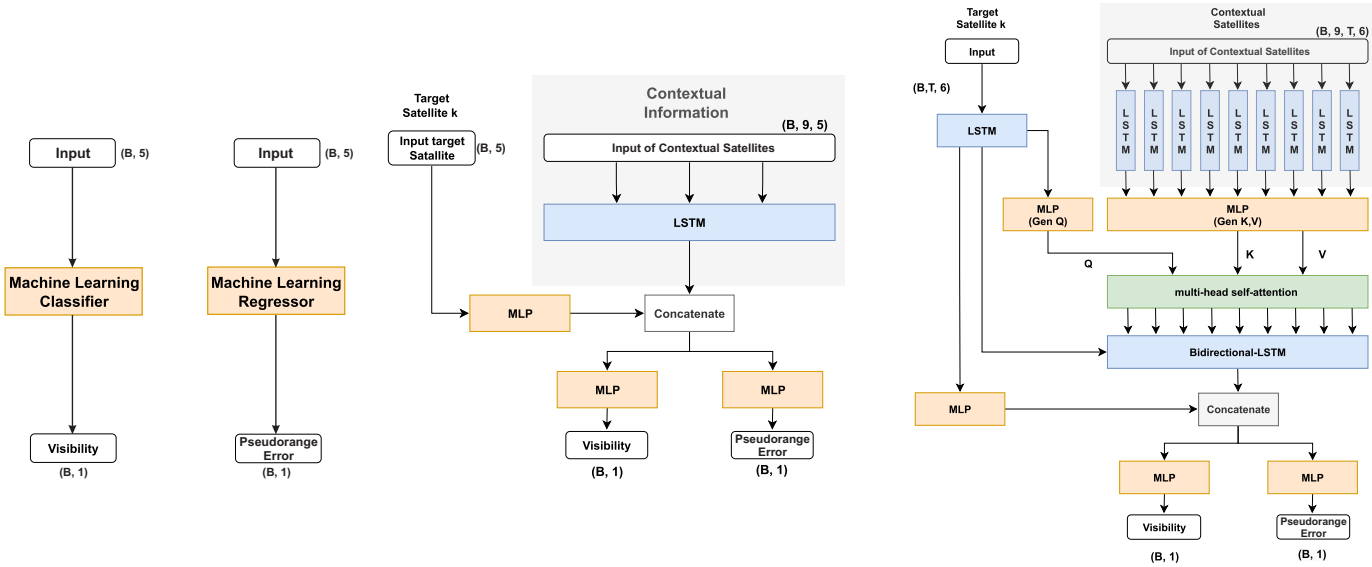
**Table I:** Parameter $c$ of six M-estimators at four efficiencies.

| Kernel \ Efficiency | 95% | 90% | 85% | 80% | Kernel Category |
|---|---|---|---|---|---|
| Fair | 1.3998 | 0.6351 | 0.3333 | 0.1760 | monotonic |
| Cauchy | 2.3849 | 1.7249 | 1.3737 | 1.1385 | soft-redescending |
| Geman-McClure (GM) | 3.8557 | 2.8937 | 2.5731 | 2.2926 | soft-redescending |
| Welsch | 2.9846 | 2.3831 | 2.0595 | 1.8383 | soft-redescending |
| Huber | 1.3450 | 0.9818 | 0.7317 | 0.5294 | hard-redescending |
| Tukey | 4.6851 | 3.8827 | 3.4437 | 3.1369 | hard-redescending |

### C. Machine Learning for Outlier Prediction

To address the limitations of current learning-based approaches for faulty GNSS prediction (see Section I-A5), we

[2]Due to space limitations, we do not provide detailed technical explanations and instead refer readers to [24, 25].




(a) Architecture of classic machine learning models.

(b) Architecture of the reference deep-learning model [34].

(c) Architecture of our previously proposed model [35].

**Figure 2:** Learning models studied in this work.

take into account the noise characteristics of GNSS observations in urban environments and propose a transformer-enhanced Long Short-Term Memory (TE-LSTM) network in our previous work [35].

We study the proposed network alongside a baseline LSTM network [34] and two classical machine learning models: support vector machine (SVM) and extreme gradient boosting (XGBoost), aiming to provide a comprehensive analysis comparing classical models with deep networks, as well as evaluating performance on in-distribution and out-of-distribution data.

We investigate two learning tasks: NLOS classification and pseudorange error prediction as a regression task, using the input feature set

$$\boldsymbol{\mathcal{F}} = \left\{ \mathrm{El}_t^k,\ \mathrm{Az}_t^k,\ \mathrm{C/N0}_t^k,\ \sigma_{\mathrm{LS},t}^k,\ \mathrm{RSS}^k \right\}, \tag{4}$$

where the feature El, Az, and $\mathrm{C/N_0}$ represent the elevation and azimuth angle, and carrier-to-noise ratio of the $k$-th satellite at timestamp $t$, respectively. The variable $\sigma_{\mathrm{LS}}$ denotes the pseudorange residual computed via an iterative least-squares method during pre-processing, which is further used to derive the root-sum-square (RSS) of pseudorange residuals within a specific time window.

Fig. 2 illustrates the architectures of all learning-based models investigated in this work, where the variables $B$ and $T$ denote the batch size and time window length, respectively. The classical models are trained separately for the classification and regression tasks, whereas the deep learning models are trained jointly with corresponding loss terms for both tasks.

For a detailed explanation of all models, we refer the reader to [35].

### *D. Bayesian Inference for Noise Distribution Approximation*

Following the previous work [29], we further explore the online learning paradigm by employing variational Bayesian inference to train a Gaussian Mixture Model (GMM) that approximates the pseudorange noise distribution[3].

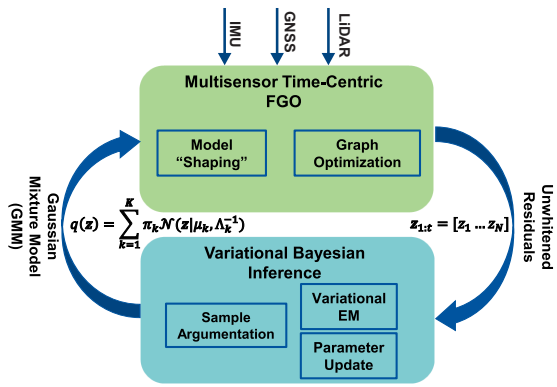



**Figure 3:** Concept of nested estimation with noise model update using variational Bayesian inference following [29].

Fig. 6 illustrates the concept of the proposed algorithm within a nested estimation framework. We employ multi-sensor factor graph optimization (FGO) to obtain "near-optimal" state estimates, which are then used to compute the pseudorange residuals. In a parallel stage, these residuals are sampled over a fixed time window, and a variational expectation-maximization (EM) algorithm is applied to approximate a GMM for each satellite [38, Chapter 10].

Due to limited space, we omit the detailed mathematical foundations of the inference algorithm and refer the reader to [38] for a detailed description. Instead, we highlight two key challenges that extend beyond previous work [29], along with our proposed solutions.

The pseudorange residual $\varepsilon_{\rho,t}^k$ from satellite $k$ depends on the current state $\boldsymbol{x}_t$, while the GMM $\Theta_{t-1}^k$, estimated over $[t-1-\Delta t,\ t-1]$, reflects a slightly outdated model. This temporal mismatch (distribution shift) is its main limitation.

[3]This work is currently in preparation for publication.

In this context, we observed that two key issues arise from this. The first issue arises whenever the vehicle is in motion. As it moves relative to a mostly static environment, its position with respect to reflective surfaces changes, leading to varying residuals caused by the same structures. Although the Gaussian mixture model includes a component with a reasonable mean, it does not account for this relative positional change. Second, if the environment changes abruptly, the outdated GMM trained with past residuals may fail to capture new outliers, leading to poor state estimates. This creates a feedback loop: degraded estimates lead to inaccurate residuals, which further distort the GMM and worsen the next estimate.

To mitigate motion-induced distribution shifts, we propose Maximum Posterior Mean Augmentation (MPMA). The MPMA models a posterior over the mean-shift $\Delta\mu$ given a residual $\varepsilon$, using the dominant GMM component $\mathcal{N}(\mu_k, \Lambda_k^{-1})$:

$$p(\Delta\mu|\varepsilon) \propto \mathcal{N}(\mu_k + \Delta\mu, \Lambda_k^{-1}) \cdot \mathcal{N}(0, \tilde{\Sigma}) \tag{5}$$

with $\tilde{\Sigma} = \frac{1}{9}\Delta r_{j,t}^2$, ensuring 99% of $\Delta\mu$ lies within probabilistic bounds.

To find the most probable mean shift $(\Delta\mu)^{\mathrm{MAP}}$, we take the negative logarithm of (5), drop constant terms and obtain

$$(\Delta\mu)^{\mathrm{MAP}} = \arg\min_{\Delta\mu} \mathcal{L}(\Delta\mu). \tag{6}$$

By minimizing the convex loss

$$\mathcal{L}(\Delta\mu) = (\varepsilon - \mu_k - \Delta\mu)^T \Lambda_k (\varepsilon - \mu_k - \Delta\mu) + \Delta\mu^T \tilde{\Sigma}^{-1} \Delta\mu, \tag{7}$$

yielding the closed-form solution

$$(\Delta\mu)^{\mathrm{MAP}} = \frac{\Lambda_k \tilde{\Sigma}(\varepsilon - \mu_k)}{1 + \Lambda_k \tilde{\Sigma}}. \tag{8}$$

To address the second issue, we extend the state estimator into a hybrid optimization framework inspired by [39], where each pseudorange factor is modeled using a multi-hypothesis (MH) noise model:

$$\eta(\varepsilon) = \begin{cases} \sqrt{w(\Sigma^{-1/2}\varepsilon)}\, \Sigma^{-1/2}\varepsilon, & d = 0 \\ \Lambda_k^{1/2}(\varepsilon - \mu_k), & d = 1 \end{cases}, \tag{9}$$

with $\Sigma$ denoting the covariance of the Cauchy hypothesis, and $\mu_k$, $\Lambda_k \in \Theta$ representing the mean and precision of the dominant GMM mode. The discrete selector $d \in \{0, 1\}$ determines the active noise model and is jointly optimized with continuous variables based on the following pseudo-probability

$$p(\varepsilon|d = 0) = \frac{1}{(2\pi)^{D/2}} |\Lambda_k|^{1/2} \left(1 - \exp\left\{-\frac{1}{2}(\varepsilon - \mu_k^*)^T \Lambda_k (\varepsilon - \mu_k^*)\right\}\right), \tag{10}$$

where we reuse $\mu_k^* = \mu_k + \Delta\mu$ and $\Lambda_k$ from (8) and denote the residual's dimension by $D$, i.e. $\varepsilon \in \mathbb{R}^D$.

## IV. Data and Essential Experimental Results

### A. Data Specification

We conduct data from real-world measurement campaigns to train and test the investigated methods introduced above. Table II summarizes the essential data characteristics.

The data from Aachen (AC) captures a 17 km route through the city of Aachen, containing different driving scenarios: open-sky, semi-/dense-urban. The Hong Kong data (HK) comprises measurements from multiple satellite constellations, with up to 74 unique satellites observed across different timestamps. Data from Düsseldorf (DUS) with raw GPS and Galileo (GAL) measurements does not have labels and thus, are only used for testing.

**Table II:** Data specification. The average, maximal, and minimum number of the received satellite are denoted with $n_{\mathrm{avg}}^{\mathrm{sat}}$, $n_{\mathrm{max}}^{\mathrm{sat}}$, and $n_{\mathrm{min}}^{\mathrm{sat}}$. $\sigma_{\mathrm{max}}^{\rho}$ denotes the maximal pseudorange error. $R^{\mathrm{LOS}}$ and $R^{\mathrm{NLOS}}$ represent the ratio of LOS and NLOS labels.

| Data | Duration (s) | $n_{\mathrm{avg}}^{\mathrm{sat}}$ | $n_{\mathrm{max}}^{\mathrm{sat}}$ | $n_{\mathrm{min}}^{\mathrm{sat}}$ | $\sigma_{\mathrm{max}}^{\rho}$ (m) | $R^{\mathrm{LOS}}$ (%) | $R^{\mathrm{NLOS}}$ ( %) |
|---|---|---|---|---|---|---|---|
| HK | 10573 | 16 | 25 | 4 | 532.5 | 60.62 | 39.38 |
| HK (GPS only) | 10573 | 5 | 9 | 1 | 276.08 | 60.47 | 39.53 |
| AC (GPS only) | 2366 | 6 | 8 | 0 | 640.32 | 77.62 | 22.38 |
| DUS (GPS+GAL) | 810 | 7 | 11 | 2 | 355.74 | - | - |

### B. Handling Outliers using M-Estimators

To study the M-estimators, we divide the data from Aachen City into four sequences, as shown in Table III.

**Table III:** Sequence definition containing four scenarios in the dataset collected in Aachen.

| Seq. | Type | Location | Start | End | Specification |
|---|---|---|---|---|---|
| 1 | Open-Sky | Campus Melaten | 0 s | 150 s | no GNSS blockage |
| 2 | Urban | Kaiser Platz | 750 s | 980 s | heavy GNSS blockage |
| 3 | Sub-Urban | Europaplatz | 1200 s | 1400 s | slight GNSS blockage |
| 4 | Sub-Urban | Eurogress | 1430 s | 1580 s | slight GNSS blockage |
| 5 | Total route | - | 0 s | 2240 s | - |

The error metrics of the mean 2-D localization errors in our experiments are presented in Table IV. Recalling the definitions of robustness and efficiency, the error metrics reveal that the robustness of all M-estimators improves as efficiency decreases. Consequently, the positioning error decreases, provided that sufficient observations remain available to prevent the state estimator from diverging.

*1) Monotonous M-Estimator:* Monotonous M-estimators like Fair assign gradually decreasing influence to residuals. Fair rapidly increases influence for small errors and flattens beyond its tuning parameter, leading to over-penalization of mildly corrupted data. As a result, it shows consistently higher positioning errors. Additionally, its influence never reaches zero, reducing robustness compared to hard-redescending estimators.

*2) Hard-Redescending M-Estimator:* Hard-redescending estimators such as Tukey and Huber suppress outliers more aggressively. Tukey fully ignores large residuals, but its effectiveness heavily depends on the tuning parameter. Mis-tuning leads to valid data being discarded, limiting performance. Huber, though widely used, doesn't reject outliers outright and

**Table IV:** Mean 2-D error of all sequences is presented. Values marked in red indicate the worst accuracy, while those in green highlight the best results.

| Eff. | Ker. \ Seq. | #1 (Open Sky) | #2 (Urban) | #3 (Light-Urban) | #4 (Light-Urban) | #5 (Total Route) |
|---|---|---|---|---|---|---|
| 95% | Fair | 0.35 | 4.09 | 1.74 | 3.08 | 2.51 |
| | Cauchy | 0.35 | 2.12 | 1.05 | 1.60 | 1.38 |
| | GM | 0.50 | 0.67 | 0.38 | 0.78 | 0.99 |
| | Welsch | 0.47 | nan | nan | nan | 3.75 |
| | Huber | 0.38 | 3.26 | 1.62 | 2.77 | 2.09 |
| | Tukey | 0.38 | 0.82 | 0.95 | 0.69 | 1.11 |
| 90% | Fair | 0.35 | 2.97 | 1.42 | 2.78 | 2.02 |
| | Cauchy | 0.22 | 0.76 | 0.44 | 2.26 | 1.10 |
| | GM | 0.30 | 1.80 | 0.41 | 1.80 | 1.20 |
| | Welsch | 0.50 | 0.50 | nan | nan | 0.99 |
| | Huber | 0.35 | 2.78 | 1.33 | 2.54 | 1.91 |
| | Tukey | 0.38 | 0.82 | 0.95 | 0.69 | 1.11 |
| 85% | Fair | 0.30 | 1.90 | 1.19 | 2.18 | 1.59 |
| | Cauchy | 0.45 | 1.13 | 0.51 | 0.98 | 1.07 |
| | GM | 0.31 | 1.05 | 1.25 | 0.64 | 1.15 |
| | Welsch | 0.38 | 0.82 | 0.95 | 0.69 | 1.11 |
| | Huber | 0.34 | 2.30 | 1.37 | 2.45 | 1.75 |
| | Tukey | 0.38 | 0.82 | 0.94 | 0.69 | 1.11 |
| 80% | Fair | 0.34 | 1.72 | 1.05 | 1.84 | 1.41 |
| | Cauchy | 0.44 | 0.73 | 0.53 | 0.92 | 1.04 |
| | GM | 0.36 | 0.47 | 0.69 | 0.46 | 1.09 |
| | Welsch | 70.38 | nan | nan | nan | 70.38 |
| | Huber | 0.31 | 1.58 | 1.25 | 2.08 | 1.49 |
| | Tukey | 0.38 | 0.82 | 0.95 | 0.69 | 1.11 |

behaves more like Fair beyond the threshold. Neither shows consistent improvement over soft-redescending estimators in our tests.

*3) Soft-Redescending M-Estimator:* Soft-redescending estimators (Cauchy, GM, Welsch) offer a gradual reduction in influence for large residuals, balancing robustness and accuracy. Cauchy and GM perform best across all efficiencies (Table IV), assigning reasonable weights even to smaller residuals potentially affected by noise. Welsch, however, showed instability, likely due to its exponential weighting, especially in urban scenarios. Overall, Cauchy and GM estimators provide the best trade-off between robustness and performance for range-based localization.

*4) **Discussions**:* Our findings highlight how to effectively apply M-estimators for handling measurement outliers, showing that aggressive down-weighting with low-efficiency estimators does not necessarily improve localization performance.

However, we do not propose a universal strategy for tuning hyperparameters, as generalizing M-estimation for long-term, large-scale localization remains challenging. Outlier characteristics and environmental variability create a difficult trade-off between robustness and efficiency in GNSS-based localization, where measurement corruption often coincides with signal loss. Even well-tuned estimators may fail to ensure reliable performance across diverse scenarios.

While prior work on generalized kernel functions and self-tuning estimators has improved generalization, consistent performance under dynamic GNSS conditions remains an open problem, especially when the noise characteristics of incoming measurements change abruptly—such as during multipath reception. These sudden shifts can disrupt the tuning process and degrade performance.

This motivates a new open question: "**Can we predict measurement outliers using learning-based methods before integrating them into the estimator?**" Or more generally: "**Can we leverage learning to recover the information needed for tuning M-estimators?**"

**Table V:** Performance Metrics using AC Dataset (reproduced from [35]).

| Model | | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| SVM | LOS | 0.80 | 1.00 | 0.89 | 0.80 |
| | NLOS | 0.39 | 0.01 | 0.01 | |
| SVM* | LOS | 0.91 | 0.72 | 0.80 | 0.72 |
| | NLOS | 0.38 | 0.70 | 0.49 | |
| XGboost | LOS | 0.86 | 0.88 | 0.87 | 0.79 |
| | NLOS | 0.47 | 0.42 | 0.45 | |
| DL [34] | LOS | 0.86 | 0.83 | 0.84 | 0.76 |
| | NLOS | 0.40 | 0.47 | 0.43 | |
| DL (ours, [35]) | LOS | 0.86 | 0.90 | 0.88 | 0.81 |
| | NLOS | 0.51 | 0.42 | 0.46 | |

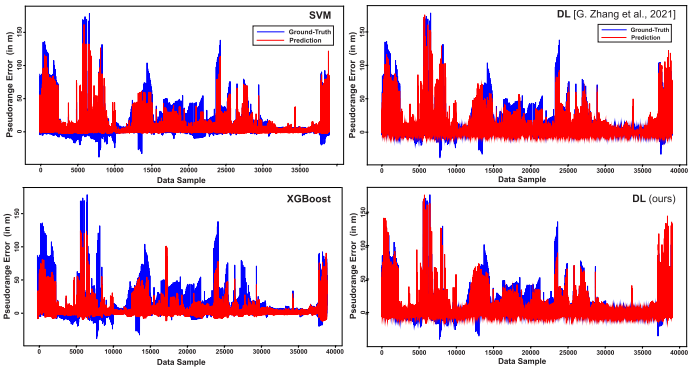



**Figure 4:** Results of pseudorange error prediction of all models using the HK data.

## C. Handling Outliers using Learning-based Methods

To evaluate the effectiveness and generalization of the learning-based methods in Section III-C, we compare them with classical ML models (SVM, XGBoost) and deep learning baselines from [34, 35], using the same dataset. Performance metrics are summarized in Table V[4].

*1) NLOS Detection:* The HK dataset, with a balanced LOS/NLOS ratio (Table II), served as an ideal benchmark. Classical models (SVM, XGBoost) performed well, slightly outperforming the deep learning baseline [34] in precision and recall. Our reimplementation of the baseline yielded similar results, with slightly improved precision.

In contrast, the AC dataset posed a challenge due to its LOS-heavy imbalance. While all models maintained high LOS accuracy, NLOS classification dropped. SVM struggled most, with a recall of just 0.01. Retraining on a balanced subset (SVM*) significantly improved performance.

*2) Pseudorange Error Prediction:* In addition to NLOS classification, we evaluated all models on pseudorange error prediction, a more challenging task critical for robust GNSS positioning. We compared predicted and ground truth errors in Fig. 4, offering qualitative insights across error magnitudes. While all models captured general error trends, none reliably predicted large negative errors.

This limitation is concerning, as large negative errors can significantly degrade localization. Their rarity in the training data likely biases models toward small or positive errors, reducing generalization to critical, fault-prone cases.

[4]Due to limited space, we only present partial results and refer readers to [35] for further details.

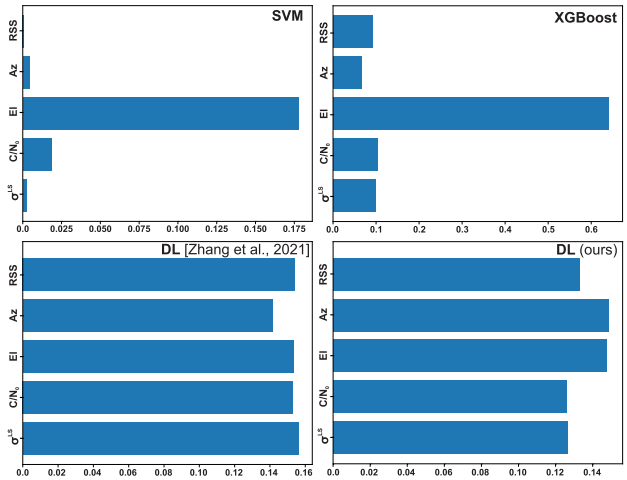



**Figure 5:** Feature Permutation Importance for NLOS classification (reproduced from [35]).

*3) Feature Importance:* To better understand model behavior, we analyzed permutation feature importance for all learning models (Fig. 5). This highlights which features most influence predictions and how feature reliance affects generalization.

Classical models like SVM and XGBoost showed strong dependence on a few features, particularly elevation angle and $\mathrm{C/N_0}$. While informative, these features may overlook contextual cues crucial in urban environments. In contrast, deep learning models exhibited a more balanced reliance across features, indicating their ability to capture complex interactions and contextual patterns.

*4) **Discussion**:* These results support the conclusion: **deep models generalize better**, though classical models excel on clean, balanced data. In addition, we think that learning-based methods are **not yet suited for back-end estimation but are effective for pre-processing**, aiding outlier detection and correction.

In addition, discriminative modeling with pre-trained models relies on a global model assumed to generalize across applications. This raises a key question: "**Does a truly global model for outlier characteristics exist?**"

Put differently: "**Do we need a global model at all?**" Or can robustness be enhanced through a "local" model that is updated in time and online? This leads us to the next method.

### *D. Handling Outliers using Variational Bayesian Inference*

We implemented the proposed variational Bayesian inference to update the noise models in the pseudorange factors and evaluated its performance alongside Gaussian noise and the Cauchy estimator. Table VI summarizes the 2-D error and FGO runtime statistics across the full dataset. In addition, Fig. 6 depicts the fitted mixture distribution alongside reference loss functions.

The Gaussian model lacks robust residual whitening, making it sensitive to outliers, which results in poor state estimates. Similarly, the standalone GMM model fails under severe outliers, preventing its use on the full dataset. Due to their limited robustness, both models are excluded from further analysis. In contrast, our MH-GMM with MPMA, which incorporates distribution-shift elimination, improves accuracy by $40\,\%$ compared to the M-estimator and by $11\,\%$ the naive MH-GMM.

**Table VI:** Noise model statistics from a measurement campaign conducted along an $11\,\mathrm{km}$ route in Aachen.

| Noise Model | 2D Error (m) | | Computation Time (ms) | |
|---|---|---|---|---|
| | mean | std | mean | std |
| Gaussian | 22.31 | 28.49 | 12.33 | 5.16 |
| M-estimator (Cauchy) | 0.90 | 0.72 | **11.93** | 4.88 |
| GMM [29] | Failed | | | |
| MH-GMM w. MPMA | **0.54** | **0.33** | 18.27 | 6.17 |
| MH-GMM w/o. MPMA | 0.61 | 0.36 | 17.87 | **4.63** |

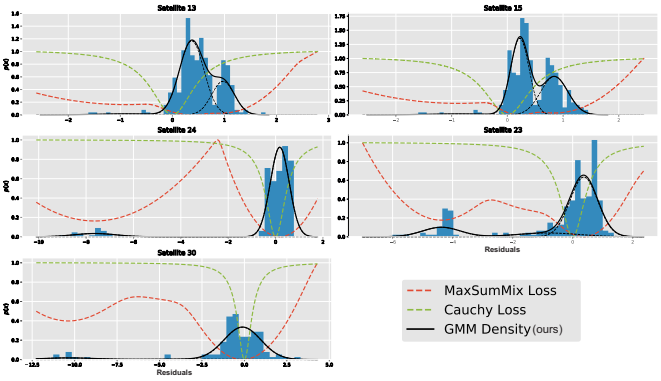



**Figure 6:** Demonstration of the online-approximated GMM distributions for different satellites while driving through a tree-rich area in Aachen.

***Discussions:*** However, although the proposed inference method, updating only the noise models, can significantly improve the overall performance of vehicle localization, issues such as the smoothness of the estimated trajectory remain unresolved. Fig. 7 exemplifies this issue, showing that the estimator using MH-GMM exhibits high variations in 2-D errors due to unsmooth estimates. This may be caused by an inaccurate estimation of the switching variable $d$, which results in frequent transitions between different noise models.

Furthermore, we did not explicitly investigate **sample augmentations** in the current work, although this is crucial in the online learning setting. We also did not examine the **effect of the non-convexity of the approximated GMM** in solving the nonlinear least-squares problem, where alternative techniques such as Hessian approximation [40] could be explored.

## V. Conclusion

Building on the results of this research, our objective is to explore various techniques for handling measurement outliers in range observations to enable robust vehicle localization in the field. Our methods, experimental evaluations, limitations, and open questions are summarized in this paper, providing a foundation for future work in enabling resilient vehicle localization in challenging field scenarios.

Taken together, we emphasize the importance of outlier handling for achieving safe and reliable autonomy, highlighting the potential of learning-based methods and Bayesian inference as promising directions for future research.

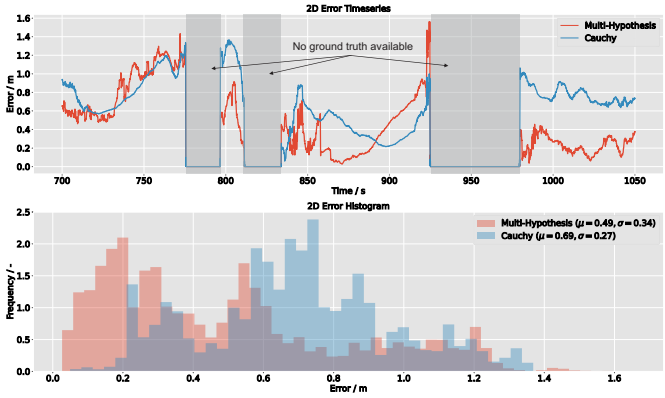



**Figure 7:** Plot of 2-D error and residual histogram in a highly urbanized area in Aachen (Kaiser Platz).

## References


[1] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, vol. 22, 11 2017.

[2] C. Villien and B. Denis, "UWB-aided GNSS/INS fusion for resilient positioning in GNSS challenged environments," in *2023 IEEE/ION Pos., Locat. and Nav. Sympo. (PLANS)*, 2023, pp. 167–178.

[3] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled LiDAR inertial odometry via smoothing and mapping," *2020 IEEE/RSJ Int. Conf. on Intell. Robots and Syst. (IROS)*, pp. 5135–5142, 2020.

[4] S. Cao, X. Lu, and S. Shen, "GVINS: tightly coupled GNSS–Visual–Inertial fusion for smooth and consistent state estimation," *IEEE Trans. on Robot.*, vol. 38, no. 4, pp. 2004–2021, 2022.

[5] W. Wen, X. Bai, and L.-T. Hsu, "3D vision aided GNSS real-time kinematic positioning for autonomous systems in urban canyons," *NAVIGATION*, vol. 70, no. 3, 2023.

[6] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super Odometry: IMU-centric LiDAR-Visual-Inertial estimator for challenging environments," in *2021 IEEE/RSJ Int. Conf. on Intell. Robots and Sys.*, 2021, pp. 8729–8736.

[7] H. Zhang, X. Xia, M. Nitsch, and D. Abel, "Continuous-Time factor graph optimization for trajectory smoothness of GNSS/INS navigation in temporarily GNSS-denied environments," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9115–9122, 2022.

[8] H. Zhang, C.-C. Chen, H. Vallery, and T. D. Barfoot, "GNSS/Multisensor fusion using continuous-time factor graph optimization for robust localization," *IEEE Trans. on Robotics*, vol. 40, pp. 4003–4023, 2024.

[9] T. Y. Tang, D. J. Yoon, and T. D. Barfoot, "A White-Noise-on-Jerk motion prior for continuous-time trajectory estimation on SE(3)," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 594–601, 2019.

[10] Q. Zhang, H. Lin, L. Ding, Q. Chen, T. Zhang, and X. Niu, "RANSAC-based fault detection and exclusion algorithm for single-difference tightly coupled GNSS/INS integration," *IEEE Trans. on Intelli. Vehi.*, vol. 9, no. 2, pp. 3986–3997, 2024.

[11] H. Jing, Y. Gao, S. Shahbeigi, and M. Dianati, "Integrity monitoring of GNSS/INS based positioning systems for autonomous vehicles: State-of-the-art and open challenges," *IEEE Trans. on Intell. Transport. Sys.*, vol. 23, no. 9, pp. 14 166–14 187, 2022.

[12] J. E. Angus, "Raim with multiple faults," *NAVIGATION*, vol. 53, no. 4, pp. 249–257, 2006.

[13] V. Drevelle and P. Bonnifait, "High integrity GNSS location zone characterization using interval analysis," in *Proc. of ION GNSS+ 2009*, 2009, pp. 2178 – 2187.

[14] S. Bhamidipati, S. Kousik, and G. Gao, "Set-valued shadow matching using zonotopes," in *Proceedings of ION GNSS+ 2021*, 2021, pp. 2373 – 2390.

[15] P. Speciale, D. P. Paudel, M. R. Oswald, T. Kroeger, L. V. Gool, and M. Pollefeys, "Consensus maximization with linear matrix inequality constraints," in *2017 IEEE CVPR*, 2017, pp. 5048–5056.

[16] V. Tzoumas, P. Antonante, and L. Carlone, "Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees," in *2019 IEEE/RSJ Int. Conf. on Intell. Rob. and Sys.*, 2019, pp. 5383–5390.

[17] N. Knight and J. Wang, "A comparison of outlier detection procedures and robust estimation methods in GPS positioning," *J. of Navigation*, vol. 62, pp. 699 – 709, 10 2009.

[18] O. G. Crespillo, D. Medina, J. Skaloud, and M. Meurer, "Tightly coupled GNSS/INS integration based on robust m-estimators," in *2018 IEEE/ION PLANS*, 2018, pp. 1554–1561.

[19] D. Medina, H. Li, J. Vilà-Valls, and P. Closas, "On robust statistics for GNSS single point positioning," in *2019 IEEE Intelli. Transport. Sys. Conf. (ITSC)*, 2019, pp. 3281–3287.

[20] P. Chauchat, A. Bellés, D. Medina, and J. Vilà-Valls, "Insights on adaptive robust filtering for navigation under harsh time-varying environments," in *2024 32nd Europ. Sig. Process. Conf. (EUSIPCO)*, 2024, pp. 1227–1231.

[21] Y. Ding, P. Chauchat, G. Pages, and P. Asseman, "Learning-enhanced adaptive robust gnss navigation in challenging environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9905–9912, 2022.

[22] S. Das and J. N. Gross, "Analysis of scale-variant robust kernel optimization for nonlinear least-squares problems," *IEEE Trans. on Aero. and Elec. Sys.*, vol. 59, no. 6, pp. 7529–7538, 2023.

[23] J. T. Barron, "A general and adaptive robust loss function," in *2019 IEEE/CVF CVPR*, 2019, pp. 4326–4334.

[24] Z. Zhang, "Parameter estimation techniques: a tutorial with application to conic fitting," *Image and Vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.

[25] D. De Menezes, D. M. Prata, A. R. Secchi, and J. C. Pinto, "A review on robust m-estimators for regression analysis," *Computers & Chemical Engineering*, vol. 147, p. 107254, 2021.

[26] K. MacTavish and T. D. Barfoot, "At all costs: A comparison of robust cost functions for camera correspondence outliers," in *2015 12th Conf. on Computer and Robot Vision*, 2015, pp. 62–69.

[27] J. Courts, A. G. Wills, and T. B. Schön, "Gaussian variational state estimation for nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5979–5993, 2021.

[28] A. Goudar, W. Zhao, T. D. Barfoot, and A. P. Schoellig, "Gaussian variational inference with covariance constraints applied to range-only localization," in *2022 IEEE/RSJ Int. Conf. on Intell. Robots and Sys.*, 2022, pp. 2872–2879.

[29] T. Pfeifer, "Adaptive estimation using Gaussian mixtures," PhD thesis, Technischen Universität Chemnitz, Chemnitz, Germany, October 2023.

[30] W. Zhao, A. Goudar, M. Tang, X. Qiao, and A. P. Schoellig, "Uncertainty-aware Gaussian mixture model for UWB time difference of arrival localization in cluttered environments," in *2023 IEEE/RSJ Int. Conf. on Intelli. Rob. and Sys. (IROS)*, 2023, pp. 5266–5273.

[31] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," in *2017 IEEE 20th Int. Conf. on Intelli. Transport. Sys. (ITSC)*, 2017, pp. 1–6.

[32] T. Suzuki and Y. Amano, "NLOS multipath classification of GNSS signal correlation output using machine learning," *Sensors*, vol. 21, no. 7, 2021.

[33] A. Siemuri, K. Selvan, H. Kuusniemi, P. Valisuo, and M. S. Elmusrati, "A systematic review of machine learning techniques for GNSS use cases," *IEEE Trans. on Aero. and Elec. Sys.*, vol. 58, no. 6, pp. 5043–5077, 2022.

[34] G. Zhang, P. Xu, H. Xu, and L.-T. Hsu, "Prediction on the urban GNSS measurement uncertainty based on deep learning networks with long short-term memory," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20 563–20 577, 2021.

[35] H. Zhang, Z. Wang, and H. Vallery, "Learning-based NLOS detection and uncertainty prediction of GNSS observations with transformer-enhanced LSTM network," in *2023 IEEE 26th Int. Conf. on Intelli. Transpor. Sys.*, 2023, pp. 910–917.

[36] P. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*. Artech House, 2013.

[37] J. S. Albuquerque and L. T. Biegler, "Data reconciliation and gross-error detection for dynamic systems," *AIChE Journal*, vol. 42, no. 10, pp. 2841–2856, 1996.

[38] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[39] K. J. Doherty, Z. Lu, K. Singh, and J. J. Leonard, "Discrete-continuous smoothing and mapping," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 395–12 402, 2022.

[40] V. Korotkine, M. Cohen, and J. R. Forbes, "A Hessian for Gaussian mixture likelihoods in nonlinear least squares," *IEEE Rob. and Auto. Lett.*, vol. 9, no. 9, pp. 7891–7898, 2024.