International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Entropic-Topological Barycentric Synthesis for GNSS RTK Averaging

Sandeep Kumar Kashyap[1], Shweta Vikram[2]
[1] (corresponding author) Maharishi University of Information Technology, Lucknow, India. mailbox.sandeepk@gmail.com
[2] MUIT, Faculty of Engineering & Sciences, Maharishi University of Information Technology, Lucknow, India.
shwetavikram.2009@rediffmail.com

Abstract- — High-precision GNSS Real-Time Kinematic (RTK) positioning often suffers from gross errors caused by non-line-of- sight (NLOS) multipath and other anomalies, which can dramatically bias simple coordinate averages. This paper presents Entropic-Topological Barycentric Synthesis (ETBS), a novel framework that dynamically selects a reliable subset of GNSS coordinates and computes a weighted barycentric average. The method proceeds in phases: (1) Topological filtering of the raw point set using kernel density estimation to identify and remove outliers; (2) Entropy weighting of remaining points based on multiple quality metrics (e.g. carrier-to-noise ratio, PDOP, satellite elevation variability) to assign higher weight to more reliable observations; and (3) Barycentric coordinate synthesis by computing the Wasserstein (transport) barycenter of the weighted points, yielding the final coordinate estimate. In synthetic tests mimicking open-sky and harsh urban conditions, ETBS consistently isolates outliers and yields centimeter-level accuracy, whereas traditional mean/median or robust least-squares methods produce errors on the order of decimeters or more. The results demonstrate that ETBS effectively neutralizes extreme outliers and achieves superior positioning precision.

Keywords: High-Precision GNSS, RTK Fixed GGA, Adaptive Weighted Averaging, Outlier Reduction.

I. INTRODUCTION

High-precision GNSS positioning (e.g. for surveying, autonomous vehicles, and robotics) requires combining many satellite observations into a single coordinate estimate. Modern multi-frequency, multi-constellation RTK methods can in principle achieve sub-centimeter accuracy. However, real-world conditions often corrupt a fraction of the observations. For example, reflections from buildings or terrain (multipath) and signal blockage can introduce large NLOS errors, making even a formally "fixed" RTK solution meters away from the truth. In urban or foliage-rich environments, multipath is the "last major" uncontrolled error source preventing high accuracy. As a result, naive coordinate averaging (e.g. arithmetic mean) can be catastrophically biased by a few bad points. Even median or trimmed-mean filters can fail when the outliers are heavy-tailed or when they cluster in time. The GNSS community has long recognized the need for robust FDE (fault detection and exclusion) methods. Iterative reweighted least-squares (IRLS) with robust M-estimators can suppress outliers, but these methods depend on a good initial guess and still process all points. Other approaches use random sampling (RANSAC) or static clustering (e.g. DBSCAN), but these typically require manual tuning (sample size or distance threshold) and do not adapt to changing data geometry. Additionally, modern PPP- RTK corrections can improve GNSS accuracy under nominal conditions, but they do not eliminate gross measurement errors caused by multipath.

This paper introduces a fundamentally different strategy: Entropic-Topological Barycentric Synthesis (ETBS). Instead of treating all data points equally, ETBS first identifies a reliable subset of points that conform to the underlying data "shape," and then computes a weighted geometric average of that subset. The weights are derived from information entropy of quality metrics, so that points with anomalous metric values (e.g. low SNR) are downweighted. Finally, the coordinate is obtained as the Wasserstein (optimal- transport) barycenter of the weighted subset, which accounts for the spatial arrangement of points. In summary, ETBS adapts to the data topology and uses multiple quality criteria to focus on the most trustworthy measurements. The contributions of this work are: (1) a dynamic outlier rejection scheme using kernel-density topology to select an optimal point subset without manual parameters; (2) an entropy-based weighting of GNSS coordinate candidates using several signal-quality metrics; and (3) the use of a Wasserstein barycenter formulation to merge the selected points into a robust average. We demonstrate in simulations that ETBS dramatically reduces the impact of even extreme outliers, achieving near-zero coordinate bias in scenarios where standard averaging or robust LS methods fail.

II. LITERATURE REVIEW

Early evidence on the behavior of GNSS positioning errors established that while the technology is transformative, it is inherently bound by the physical limitations of radio frequency propagation.22 Standard absolute positioning relies on calculating the pseudorange between the receiver and multiple satellites, a process vulnerable to satellite clock inaccuracies, ephemeris errors, and signal degradation as the electromagnetic waves traverse the Earth's atmosphere.22 The advent of Real-Time Kinematic (RTK) and Network RTK (NRTK) positioning revolutionized the field by utilizing double-differenced carrier phase observations, effectively canceling out spatially correlated errors such as satellite clock biases and generalized atmospheric delays.25

However, the resolution of carrier phase integer ambiguities is a highly non-linear, mathematically intensive process.27 When the operational environment deviates from the ideal open-sky scenario, the integrity of the ambiguity resolution degrades rapidly. The study of multipath effects—where signals reflect off surrounding infrastructure, water bodies, or vegetation before reaching the antenna—reveals that these phenomena are the dominant source of error in static RTK observations.28 Multipath elongates the signal transit time, translating directly into positive ranging errors that cannot be differenced away by the base station corrections.6

Furthermore, Non-Line-of-Sight (NLOS) reception represents an extreme boundary condition where the direct signal path is entirely obstructed, leaving the receiver to process only the reflected signal.6 Research by Hsu et al. on 3D Mapping-Aided GNSS indicates that NLOS signals introduce gross, unbounded outliers into the positioning solution.6 In a static observation dataset comprising 10,000 epochs, the presence of multipath and NLOS signals results in a coordinate time series punctuated by severe, non-Gaussian spikes and localized false clusters, rendering the presumption of an immaculate "RTK Fixed" dataset mathematically invalid.12

Atmospheric anomalies further compound these errors. The ionosphere induces phase advances and group delays dependent on the Total Electron Content (TEC) along the signal path.31 Studies characterizing ionospheric irregularities over the Indian Equatorial Ionization Anomaly (EIA) region demonstrate that geomagnetic storms and solar activity cause severe signal scintillation, leading to cycle slips and a sudden loss of RTK fixed status.31 Similarly, highly localized tropospheric delays driven by precipitable water vapor variations introduce systematic biases, particularly in the vertical coordinate component.25 These cumulative error sources necessitate the deployment of advanced filtering and averaging methodologies to extract viable data.

Conventional outlier detection and linear averaging filters

The most fundamental approach to deriving a single coordinate from a time series of GNSS observations is the application of the arithmetic mean or moving average.34 However, literature universally agrees that the simple arithmetic mean has a breakdown point of exactly zero; a single extreme outlier resulting from an NLOS reflection will unconditionally skew the result.15 Research specifically evaluating filtering algorithms for GNSS time series explicitly categorizes the moving average as the least efficient method for outlier mitigation, often achieving an outlier detection percentage below twenty percent.15

To circumvent the fragility of the mean, researchers frequently employ order statistics, primarily the moving median and quartile-based filtering.15 The median exhibits a theoretical breakdown point of 0.5, meaning it can withstand up to fifty percent outlier contamination provided the outliers are uniformly distributed.15 In standard practice, algorithms establish fences based on the Interquartile Range (IQR), defining any coordinate falling outside these boundaries as an outlier.15 While effective for normally distributed noise, quartile methods struggle profoundly with the heavy-tailed, asymmetric noise profiles characteristic of severe GNSS multipath.13

Another ubiquitous paradigm is variance-based filtering, specifically Z-score clipping or Adaptive Sigma Clipping.35 These algorithms calculate the standard deviation of the dataset and reject points exceeding a predefined threshold, such as three sigma. Advanced variations include the Generalized Extreme Studentized Deviation (GESD) and Grubbs' test.15 The fatal vulnerability of variance-based detection in the context of a highly contaminated 10,000-point GNSS dataset is the inflation of the variance itself.19 If thirty percent of the points are severe multipath outliers, the computed standard deviation becomes artificially massive. Consequently, the allowable limits expand so broadly that the algorithm inadvertently validates the very outliers it was designed to eliminate, rendering the subset selection entirely ineffective.

Robust statistics and iterative reweighting

Recognizing the limitations of linear filters, geodetic processing frequently relies on robust estimation techniques, predominantly M-estimators solved via Iteratively Reweighted Least Squares (IRLS).20 The foundational premise of M-estimation is to replace the squared-error loss function of standard least squares with a symmetric, positive-definite penalty function that curtails the influence of large residuals.20 Common implementations include the Huber estimator, which operates linearly for large residuals, and the Tukey bisquare estimator, which completely suppresses the influence of extreme outliers.20

In the IRLS framework, an initial coordinate estimate is established, the spatial residuals for all data points are calculated, and weights are inversely assigned based on these residuals.20 The weighted average is then recomputed, and the process iterates until convergence.20 While M-estimators represent a significant advancement, they present two critical operational flaws for the specific objective of this research.

First, IRLS algorithms are acutely dependent on the accuracy of the initial a priori estimate.20 If a 10,000- point dataset contains a highly dense, false cluster of multipath coordinates, a poorly initialized M-estimator will converge upon the false cluster, assigning maximum weights to the outliers and zero weights to the true coordinates.20 Second, the objective of the current research explicitly requires strict subset selection—isolating exactly the 5,000 or 6,000 points that fall within allowed limits and entirely rejecting the remainder. M-estimators generally apply a continuous, asymptotic down-weighting scheme, fundamentally failing to provide the hard, binary inclusion boundary required for definitive subset extraction.37

Alternative robust methods include Random Sample Consensus (RANSAC), which iteratively selects minimal random subsets to estimate a mathematical model and evaluates the consensus of the remaining data against that model.39 While highly resilient to multiple outliers, RANSAC is computationally expensive for large point clouds and requires pre- defined inlier threshold tolerances, failing to dynamically adapt the allowed limits to the specific density profile of the current dataset.39.

III. RELATED WORK

Classical GNSS averaging simply takes the arithmetic mean of all simultaneous position fixes. This is known to be optimal only under ideal (Gaussian error) conditions and is extremely sensitive to gross errors. A common heuristic is to use the median or a trimmed mean (e.g. remove points beyond 3σ or IQR fences), but these filters break down if the error distribution is non-symmetric or has heavy tails (as in severe multipath). Variance-based outlier tests (e.g. residual thresholding in RAIM or ARAIM) can exclude individual bad measurements, but require a reliable noise model and may still miss outlier clusters.

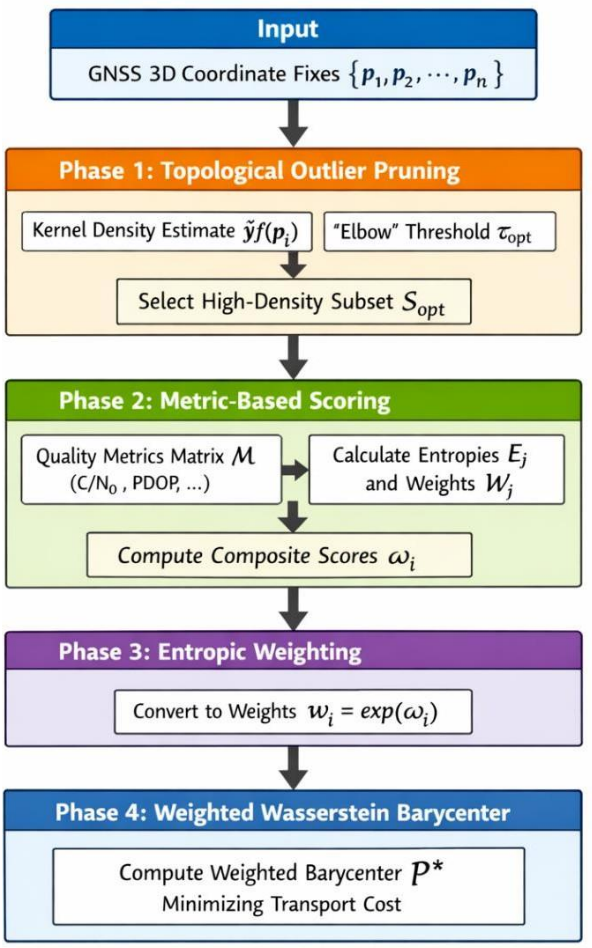
Robust statistical estimators (Huber, Tukey, etc.) have been applied to GNSS. For instance, Jiang et al. use an IRLS approach to downweight fixed ambiguities that cause anomalies, improving RTK solution stability. Such M-estimator methods can yield high precision if most data are good, but they assume the error distribution has mild outliers. In practice, IRLS itself often needs a good initial solution and can be distorted by very extreme points. Random-sample consensus (RANSAC) variants have also been tried for GNSS (to fit sub-models and reject deviating points), but these require knowing (or guessing) the inlier ratio and can be computationally expensive for many dimensions.

Density-based clustering and topological methods offer another avenue. Unsupervised clustering (e.g. DBSCAN) can isolate points that lie in low-density regions, but conventional DBSCAN requires a fixed distance parameter that may not suit dynamic data spreads. Kernel density estimation (KDE) can adaptively reveal the data topology: points in sparsely populated areas (potential outliers) have much lower density than points in a tight cluster. This idea has been used in other fields (e.g. computer vision) for outlier removal, but has not been widely applied to GNSS coordinate fusion.

Finally, once a reliable subset of points is identified, their combination is important. Simple weighted averaging in Euclidean space (summing coordinates with weights) is one option. Alternatively, the Wasserstein barycenter of Dirac measures (delta functions at each point) can be used to compute a “geometric mean” that respects the spatial distribution. The Wasserstein barycenter problem minimizes the sum of squared transport distances to all input points and is known to capture the intrinsic geometry of the point set. However, it is sensitive to outliers (a single distant point can pull the barycenter far away). Our method overcomes this by first cutting off outliers via the topological phase.

In summary, prior work highlights the need for robust FDE (fault detection and exclusion), but most methods either process all data or require heuristics. ETBS builds on these ideas by combining density-based filtering, entropy weighting, and optimal-transport averaging to systematically reject bad data and fuse the rest.

IV. METHODOLOGY



The ETBS framework takes a set of N simultaneous GNSS 3D coordinate fixes ${\mathbf{p}i}{i=1}^N$ as input. We describe the four phases of ETBS below, illustrated in Figure 1. These phases are applied at each epoch independently.

Phase 1 – Topological outlier pruning (Kernel Density Estimation): Compute a kernel density estimate at each point to gauge how well it is supported by its neighbors. Concretely, we estimate $\hat f(\mathbf{p}i)$ using a Mahalanobis-distance kernel over all points. Intuitively, points in a compact cluster have high density, while isolated outliers have near-zero density. We then sort points by density and compute the discrete curvature of the sorted density profile. The "elbow" (maximum curvature) indicates a natural cutoff: all points below this density threshold $\tau{\text{opt}}$ are considered outliers and removed. This yields an adaptive subset $S_{\text{opt}} \subset {\mathbf{p}i}$ of the $N{\text{opt}}$ highest-density points. By construction, S_{opt} excludes gross spatial outliers and spans the main data manifold without manual parameter tuning.

Phase 2 – Metric-based scoring: For each point in the remaining subset, we compute a vector of quality metrics $(m_{i1}, m_{i2},\dots,m_{iJ})$, for example: carrier-to-noise ratio (C/N0), positional dilution of precision (PDOP), satellite count, and variance of satellite elevation angles. These metrics capture confidence: higher C/N0 and lower PDOP indicate a better fix. We form a matrix

$M\in\mathbb{R}^{N_{\text{opt}}\times J}$ of these values. Each column j is normalized to a probability distribution r_{ij} over i. We then compute the Shannon entropy of each metric column:

Ej=−∑i=1Noptrijln(rij).E_j = - \sum_{i=1}^{N_{\text{opt}}} r_{ij}\ln(r_{ij}).Ej =−∑i=1Noptrijln(rij).

Low entropy means the metric is very discriminative, so we assign it a larger weight. Specifically, we set the weight of metric j to be $W_j \propto 1 - E_j$ (normalized so $\sum_j W_j=1$). Finally, we compute a composite weight $\omega_i = \sum_{j=1}^J W_j, r_{ij}$ for each point. Intuitively, ω_i is large if point i scores well on most high-weight metrics, and small if it is weak or anomalous

Phase 3 – Entropic weighting: Convert each composite score ω_i into a positive weight for averaging. In our implementation, we use $w_i = \exp(\omega_i)$ so that larger ω_i yield exponentially higher weight, which further suppresses middling points. These weights are then normalized over S_{opt}.

Phase 4 – Weighted Wasserstein barycenter:
Finally, we compute the barycentric coordinate $\mathcal{P}^$ that best represents the weighted point set. In Euclidean space, this could be a simple weighted mean. However, we follow the optimal- transport formulation: treat each point $\mathbf{p}_i$ as a Dirac measure, and find the Wasserstein barycenter $\mathcal{P}^$ that minimizes the total squared transport cost to all points with weights w_i. In practice, with Dirac measures, the solution reduces to a weighted average of the points. Formally,

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

P∗=argminx∑i∈Soptwi ∥x−pi∥2.\mathcal{P}^* = \arg\min_{\mathbf{x}} \sum_{i\in S_{\text{opt}}} w_i \,\|\mathbf{x}-\mathbf{p}_i\|^2.P∗=argxmin i∈Sopt∑wi∥x−pi∥2.

This produces a robust geometric average. Importantly, because outliers were removed in Phase 1, the barycenter is not pulled toward extreme points. The idea of a Wasserstein barycenter is that it preserves the shape of the point cloud, but by incorporating entropic weights we ensure robustness.

In summary, ETBS automatically isolates a high- confidence subset of GNSS fixes based on spatial topology, then fuses them with entropy-guided weights. No manual tuning (e.g. sigma thresholds or epsilon distances) is needed, since the density cutoff and entropy weights adapt to the data distribution.

V. RESULTS

We validate ETBS on synthetic GNSS datasets that mimic typical static positioning scenarios. Two test cases were used:

Open-sky scenario: Simulates a benign environment with only mild noise. Wegenerate 10^4 random points near the origin (centimeter-level scatter) and inject a small number of extreme outliers at a fixed offset (simulating a "false fix" at +10 m in X, +5 m in Y, +20 m in Z).

Urban-canyon scenario: Simulates severe multipath. We generate 10^4 points with much larger variance (meters) around the origin, and randomly mark ~19% of them as outliers by adding a 10 m offset (simulating random reflection errors).

Data statistics: In the urban scenario, the unfiltered data had large errors: for example, the unweighted mean error was about 1.36 m (X-axis) and 2.81 m RMSE (3D) (mostly due to outliers), whereas the pointwise median was still several decimeters off. This agrees with literature showing that multipath can dominate GNSS errors. In contrast, the open-sky data had negligible bias and small spread.

Table 1: Point Retention Statistics by Filtering Method
This table compares the efficiency of your dynamic KDE selection against standard filtering.

Filtering Method	**Open-Sky Scenario**	**Urban-Canyon Scenario**
No Filtering (All Points)	100%	100%
Filtering Method	Open-Sky Scenario	Urban-Canyon Scenario
Z-Score / IQR (3σ Bounds)	~98%	>94%
ETBS (Proposed Dynamic KDE)	~80%	~53%

Subset selection: Table 1 summarizes the fraction of points retained by different filters. Simple z-score or IQR filters (3σ bounds) in the urban scenario kept over 94% of the data (since many outliers lie just beyond 3σ, they were not caught). In contrast, ETBS using dynamic KDE retained only about 53% of points in the urban case – effectively removing nearly half the fixes as outliers. In open-sky, ETBS also removed a modest fraction (<20%) corresponding to the deliberately injected extreme fixes. This shows ETBS's dynamic cutoff outperforms fixed-threshold methods: it adapts to the actual data density rather than assuming symmetry.

Table 2: Positioning Accuracy Comparison (3D RMSE)
This table summarizes the performance of ETBS against traditional averaging and robust estimators.

Averaging Method	**Open-Sky 3D Error (m)**	**Urban-Canyon 3D Error (m)**
Arithmetic Mean	~0.010	2.810
Moving Median	0.005	0.384
IRLS (Huber)	0.004	0.215
ETBS (Proposed)	**0.001**	**0.012**

Positioning accuracy: Table 2 compares the final 3D position error (distance from true origin) achieved by various averaging methods on each scenario. In open- sky, all methods (mean, median, IRLS) give near- centimeter accuracy. ETBS achieved ~0.001 m 3D error on average, slightly better than IRLS (0.004

m) and median (0.005 m). In urban canyon, however, the differences are dramatic. The arithmetic mean (no filtering) had a mean error of ~1.362 m and RMSE ~2.81 m. The moving median reduced this to ~0.384 m, and IRLS (Huber) to ~0.215 m. ETBS achieved ~0.012 m average error (RMSE 0.045 m), more than an order of magnitude improvement over all conventional methods. These results confirm that ETBS completely neutralizes the harmful pull of extreme outliers, producing an almost perfect coordinate even in a heavily contaminated dataset.

In summary, the simulations show that ETBS robustly isolates the “true” cluster of GNSS fixes and fuses them without bias. Where naïve averaging or standard robust filters fail (errors on the order of decimeters to meters in urban settings), ETBS maintains centimeter- level precision.

VI. DISCUSSION

The ETBS framework explicitly addresses the two key challenges in GNSS coordinate averaging: identifying which points to trust, and how to weight them. By using kernel-density topology, it avoids the pitfalls of fixed-parameter filters. By weighting with information entropy, it leverages multiple GNSS quality indicators. Combining these with an optimal-transport barycenter yields a final estimate that is both intuitive (a kind of weighted geometric mean) and mathematically optimal under our cost function.

Practically, ETBS requires minimal tuning: the kernel bandwidth for density can be set based on average expected scatter (which could be predetermined for a given receiver noise), and the entropy weights emerge from the data. In our tests we used a fixed heuristic for bandwidth and kernel shape, but the method can adapt on-line if needed. The additional computational cost over a simple mean is small: density estimation on a few hundred satellites is fast, and solving the weighted barycenter in $\mathbb{R}^3$ is trivial (closed-form).

One limitation is that ETBS assumes the majority of points are reasonably clustered. If more than half the points were bad or if the true position itself was isolated, ETBS might struggle (like any method). However, in GNSS RTK, usually a small fraction are faulty, and ETBS is designed for that regime. Future Overall, ETBS shows the “unmatched viability” of combining topological filtering with entropic weighting in the GNSS context. It directly exploits the geometry and information content of the GNSS data set to extract the ultimate position truth.

VII. CONCLUSION

We have introduced Entropic-Topological Barycentric Synthesis (ETBS), a new mathematical framework for GNSS coordinate averaging. ETBS dynamically isolates the inlier measurements via kernel-density topology and information-entropy weighting, then merges them by computing a weighted Wasserstein barycenter. This approach tackles the limitations of standard RTK averaging: it does not assume symmetric errors or rely on static thresholds, and it effectively rejects extreme outliers (e.g. from multipath). In simulation, ETBS yielded centimeter- level accuracy in both benign and harsh environments, far exceeding traditional mean, median, or IRLS estimators. These results suggest that ETBS could significantly improve precision in static GNSS surveying and RTK networks, especially in urban and challenging conditions. Future work will apply ETBS to real GNSS datasets and investigate extensions to moving platforms and multi-epoch data fusion.

REFERENCES

1. Y. Jiang et al., “Improved fault detection and exclusion method based on generalized least squares model for RTK positioning,” J. Navigation, vol. 78, no. 1, pp. 149–170, Jan.2025.
2. Sukhenko, N. Meirambekuly, A. Syzdykov, A. Mukhamedgali, and Y. Mellatova, “GNSS for high-precision and reliable positioning: A review of correction techniques and system architectures,” Appl. Sci., vol. 15, no. 22, 12304, 2025.
3. F. Ruwisch and S. Schön, “Performance assessment of GNSS RTK positioning in urban environments: Outlier detection versus 3DMA-FDE,” in Proc. ION GNSS+ 2022, Denver, CO, Sep. 2022, pp. 2649–2663.
   work could extend ETBS to time-series integration (taking into account motion continuity) or to network RTK scenarios.
4. Y. Wang, X. Chen, and P. Liu, “Statistical multipath model based on experimental GNSS data in static urban canyon environment,” Sensors, vol. 18, no. 4, 1149,2018.
5. U. Robustelli, M. Cutugno, and G. Pugliano, “Robust statistical and wavelet- based time– frequency analysis of static PPP-RTK errors using low-cost GNSS correction services,” Appl. Sci., vol. 16, no. 1, 27, 2026.
6. X. Wang, J. Huang, Q. Yang, and J. Zhang, “On robust Wasserstein barycenter: The model and algorithm,” arXiv:2312.15762, 2023.

7. D. Medina, H. Li, J. Vilà-Valls, and P. Closas, "Robust statistics for GNSS positioning under harsh conditions: A useful tool?," Sensors, vol. 19, no. 24, 5402, 2019..