GNSS AUGMENTATION

Introduction

In dense urban environments, utilizing the Global Navigation Satellite System (GNSS) for localization presents numerous challenges due to:

- Signal obstructions caused by tall buildings and vehicles
- Multipath interference from surrounding structures.

This study addresses these challenges, particularly in the context of data collected from smartphones equipped with GNSS receivers.

- Investigating outlier detection methods to enhance the reliability of GNSS data in urban settings
- Highlight the differences between each technique to give insight about advantages and disadvantages

Methodology

Our data consists of **132 satellites** and **1631 epochs**, collected by driving a vehicle equipped with Xiaomi and Samsung smartphones around urban areas, simulating real-world usage scenarios.

Outlier detection methods were implemented to identify and mitigate localization errors caused by non-line-of-sight and multipath anomalies. These methods included:

- Global Test:
    - Chi-square test
- Local Test:
    - Parametric Test:
        - T-test
    - Non-Parametric Test:
        - Local Outlier Factor (LOF)
        - Isolation Forest (IF)
        - Isolation Forest - Local Outlier Factor (IF-LOF)

Implementing these methods aimed to provide a comprehensive approach to outlier detection, balancing computational efficiency with accuracy.

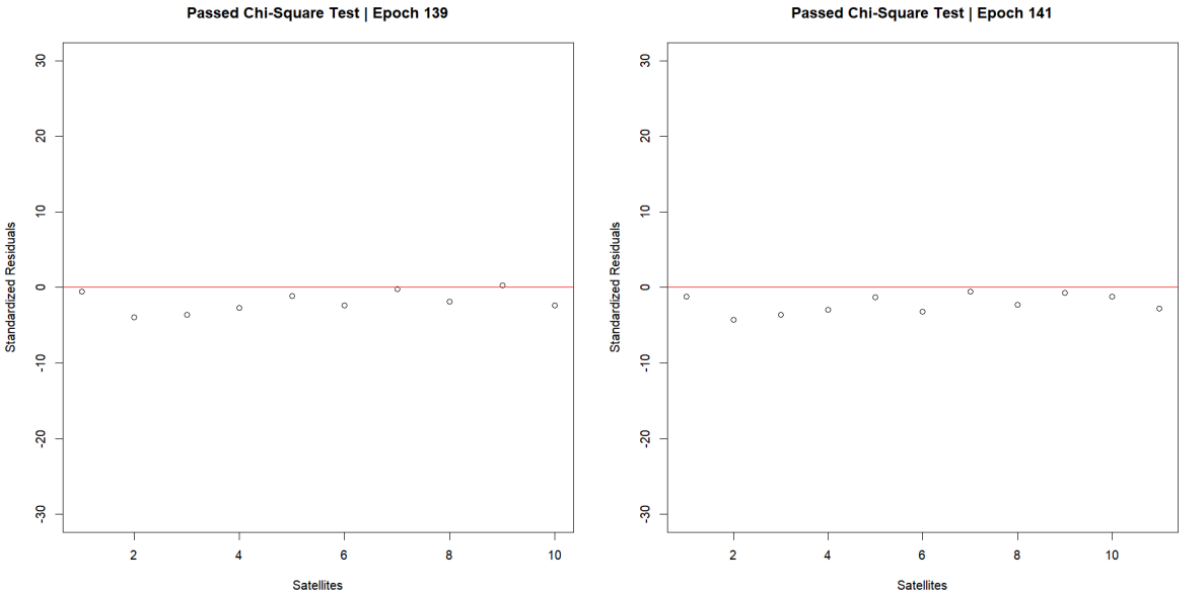
Analysis

The analysis phase consists of two-stage processes, global and local steps.

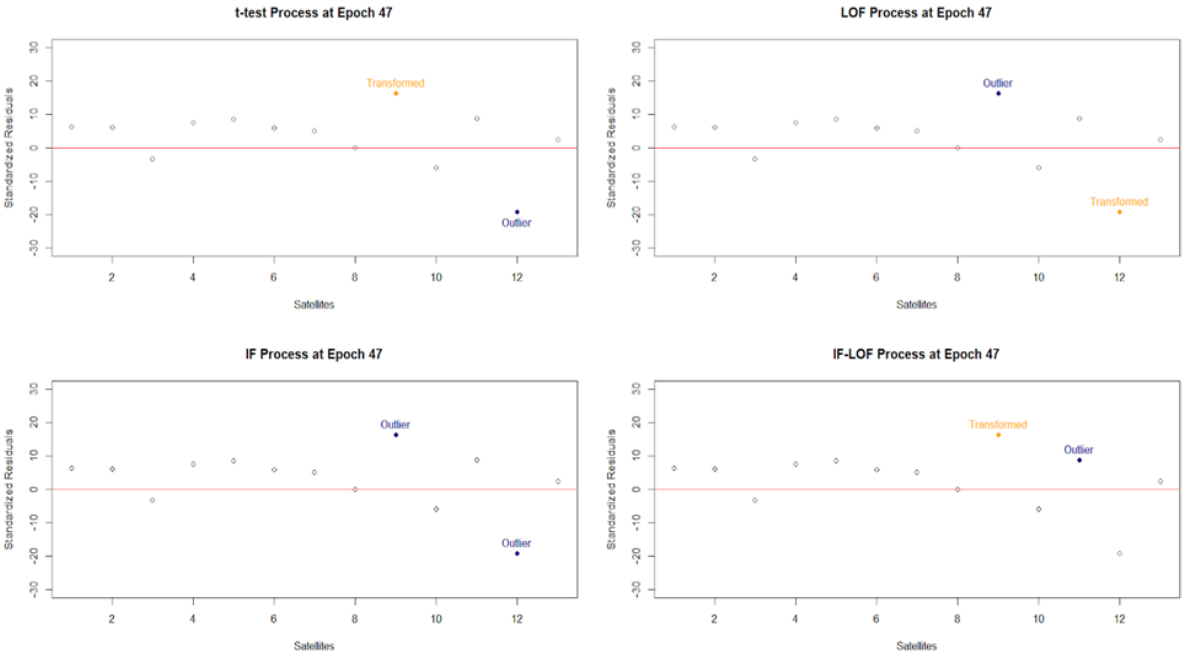
- Global chi-square test was conducted on the covariance matrix. The epochs that failed the global test signify that they require more investigation.

- Local tests were implemented as a second stage for the failed epochs. The t-test, LOF, and IF-LOF methods used predetermined threshold values of **90%** and **97%** confidence levels

Furthermore, any test statistics that fall under the **90-97** region were transformed by a 3-segment adaptive factor technique and anything beyond the **97** confidence level was considered an outlier.



Stage 1: Random sample of epochs that passed the global chi-square test



Stage 2: Random sample of epoch that failed the global chi-square test

Delving into the reason why contradiction about transformed and outlier points happens among the **t-test** and **LOF**.

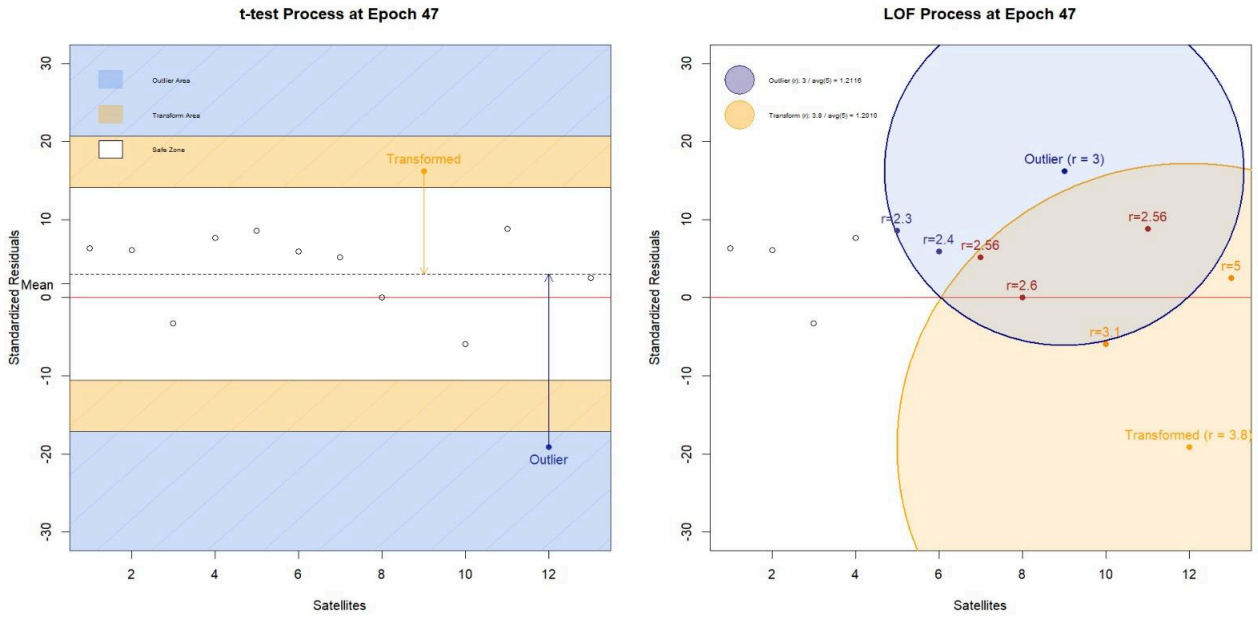


Figure: Root cause of the discrepancy between the t-test and LOF

Despite the obvious differences in methods due to their underlying structures, it is analyzed whether or not they were suspicious of the similar satellites in the duration of **1631 epochs**. Methods are compared with initial t-test applications.

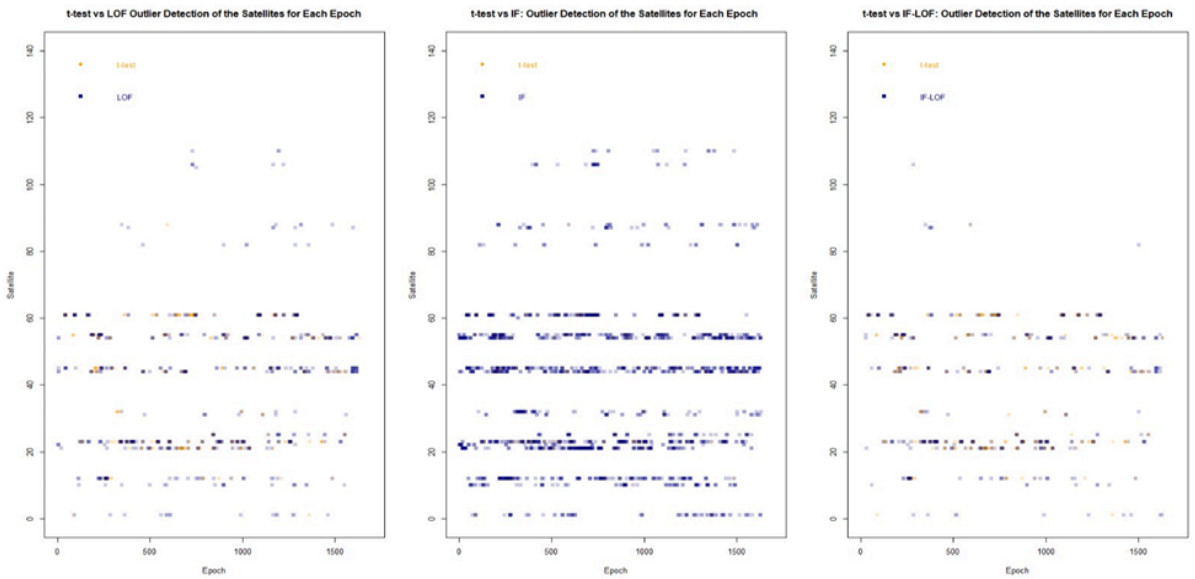


Figure: Outlier detection of the Satellites for each Epochs

	Total Observation	Outlier %	Transformed %
t-test	19188	2.47%	10.89%
LOF	19188	2.98%	4.77%
IF	19188	12.26%	NA
IF-LOF	19188	2.13%	3.38%

Xiaomi Outlier Detection Table

	Total Observation	Outlier %	Transformed %
t-test	17878	2.20%	11.63%
LOF	17878	3.89%	6.30%
IF	17878	15.25%	NA
IF-LOF	17878	1.92%	3.81%

Samsung Outlier Detection Table

Figure: outlier and transformed percentage for samsung and Xiaomi

Results

The overall distribution of the standardized residuals has improved with minor changes for Xiaomi in regards to the Normal Q-Q plot.

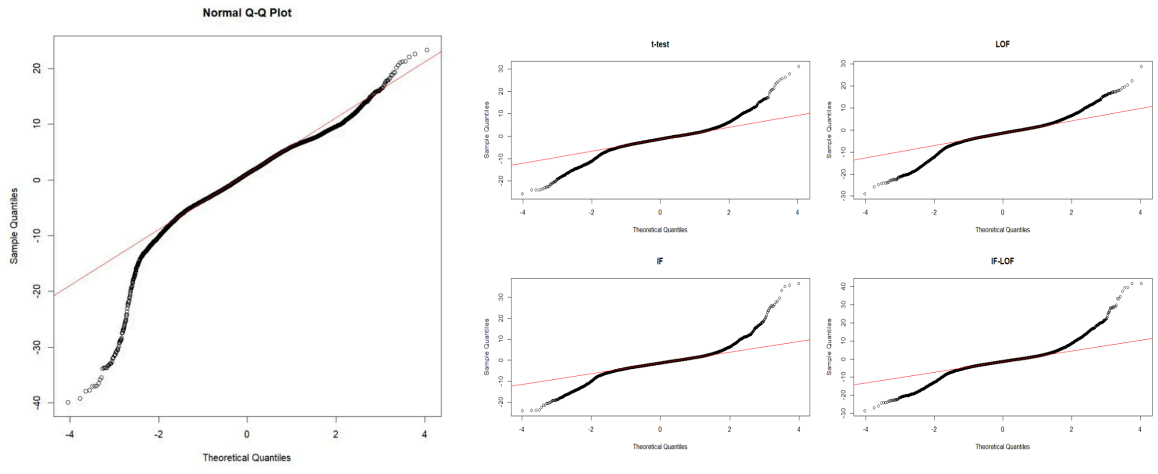


Figure: Xiaomi Outlier Detection - Before and After

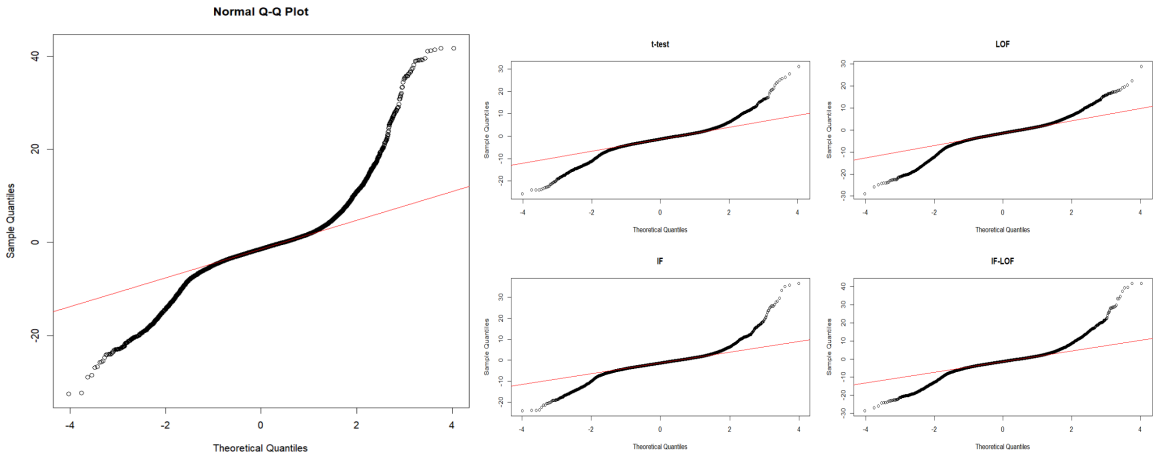


Figure: Samsung Outlier Detection - Before and After

Conclusion

In the study, we investigated outlier detection techniques to enhance the reliability of GNSS data in dense urban environments.

- While global statistical tests like the **Chi-Square** test offer computational advantages, they may overlook epochs with group outliers
- For the local tests, the t-test, local outlier factor, and the isolation forest have the most robust results, **LOF** being the most successful one under large quantile values.
- While **isolation forest (IF)** performed successfully as t-test and LOF, it removed significantly more observations, which resulted in the deduction of degrees of freedom for the model.
- Within the **t-test** and **LOF** comparison, the t-test has substantially more transformation, and LOF has some unreliable results for reliable GPS satellites.

In conclusion, methods with transformation technique, t-test and LOF have the most optimized results compared to others. Thus, t-test or LOF methods can be implemented depending on the assumption of data behavior. These findings can contribute to advancing the field of urban localization and have practical implications for improving location-based services and navigation systems on smartphones. As a result, it can open up new possibilities for the application of these methods in various fields, including autonomous driving, delivery services, and more.