Advanced Engineering Days

aed.mersin.edu.tr

Improving GNSS data accuracy using DBSCAN, moving averages, and Hampel identifier

Hüseyin Pehlivan*

Gebze Technical University, Department of Geomatics Engineering, Türkiye, hpehlivan@gtu.edu.tr

Cite this study: Pehlivan, H. (2023). Improving GNSS Data Accuracy Using DBSCAN, Moving Averages, and Hampel Identifier. Advanced Engineering Days, 6, 71-73

Keywords
GNSS data
Filtering
Hampel
Moving average
DBSCAN

Abstract

There is a common issue in GNSS (Global Navigation Satellite System) data, which is the presence of outliers that can affect the accuracy of positional measurements. In this study, three methods for outlier detection and removal in GNSS data were compared: Hampel filter, moving average, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Both synthetic and real GNSS datasets were used to test these methods. The Hampel filter and moving average were first applied to clean the data, and then the DBSCAN algorithm was used to detect outliers. The results were evaluated using the RMS error criterion. The study found that DBSCAN was effective with appropriate parameter settings, but the combination of Hampel filter and moving average was the most successful method. The Hampel filter was particularly efficient in filtering outliers in low-quality GNSS data. These findings suggest that the combination of multiple methods can result in more accurate and reliable outlier detection and removal in GNSS data.

Introduction

In Outliers in GNSS (Global Navigation Satellite System) data can cause problems in data analysis and processing, especially in situations where measurement capabilities are limited. While accurately identifying outliers from time series is important in general, it is also widely used in data mining, machine learning, and many other applications. Many methods have been proposed for outlier detection, but methods such as the Hampel filter, moving average, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm are popularly used.

There are many studies that have utilized Hampel filter, moving average, and DBSCAN algorithms for outlier detection and cleaning in GNSS data. For example, Yang and Rizos [1] proposed the use of Hampel filter and Grubbs test for outlier detection in GNSS data to improve accuracy. Rana and Tiwari [2] suggested the use of moving average and double difference technique to detect and remove biases in GNSS data. Hou and Chen [3] showed that DBSCAN algorithm was an effective method for outlier detection and cleaning in GNSS data. However, each of these studies employed different approaches for identifying and removing outliers in GNSS data. In particular, Hampel filter has been used for many years as a statistical method for detecting and removing outliers [4], while moving average is a commonly used method for signal smoothing [5]. DBSCAN algorithm is a density-based clustering method that can also be used to identify outliers [6].

Additionally, Tang et al. [7] and Zhu et al. [8] are among the many studies that have utilized these methods for outlier detection and cleaning in GNSS data. In this study, we propose that the combination of Hampel filter and moving average may be more effective for outlier detection, and we also investigate the effectiveness of DBSCAN algorithm for this purpose. Our goal is to create an overall algorithm for outlier detection using the Hampel filter, moving average, and DBSCAN methods. The cleaned data with Hampel filter and moving average were then used for outlier detection with the DBSCAN method, and the obtained results were compared using the RMS error criterion. The effectiveness of the results and the advantages of combining these methods for outlier detection compared to other methods were also evaluated.

Material and Method

The main research question of this study is to investigate the effectiveness of the combination of Hampel filter, moving average, and DBSCAN algorithm in detecting and removing outliers in GNSS data. The hypothesis is that the combined approach will result in more accurate and reliable data compared to using any of the methods individually. The study is designed in three stages to test this hypothesis.

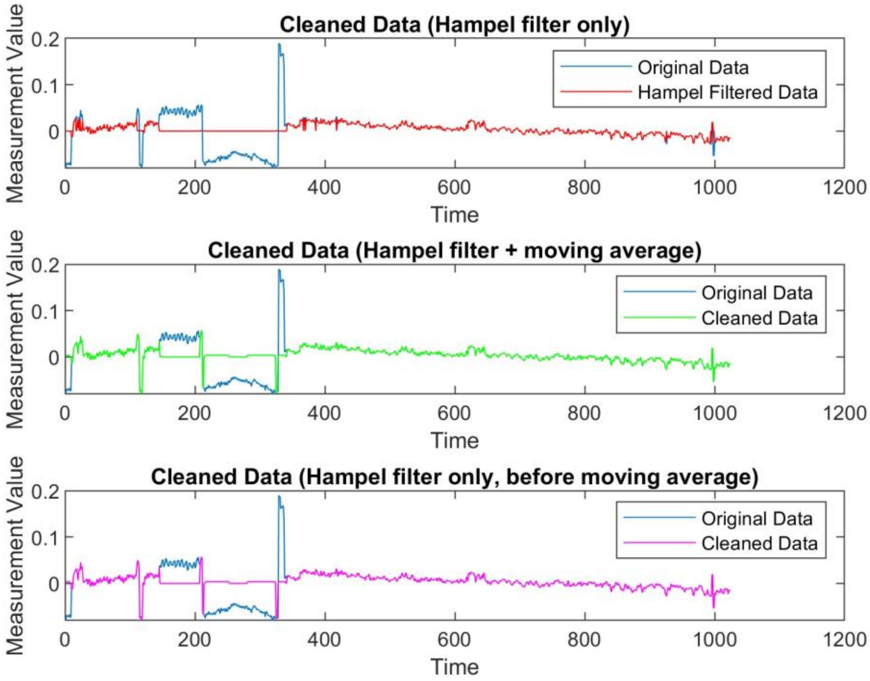

Figure 1. Detection of outliers using Hampel filter and moving average algorithms

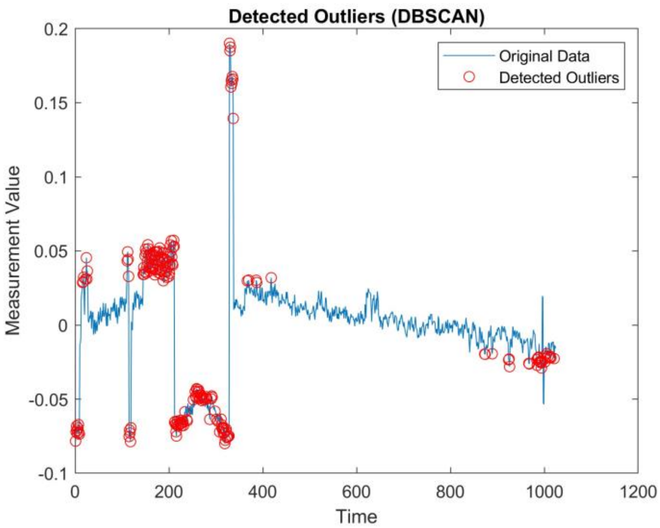

Figure 2. Detection of outliers using the DBSCAN algorithm.

In the first stage, Hampel filter and moving average methods were applied to the dataset to remove initial outliers (Figure 1). In the second stage, DBSCAN algorithm was used to detect remaining outliers in the data (Figure 2). In the final stage, statistical analysis was conducted on the results to evaluate the effectiveness of DBSCAN, Hampel filter, and moving average methods in outlier detection after their application to the data. Root Mean Square (RMS) error criterion was used to compare the filtered data with the original data in order to evaluate the effectiveness of DBSCAN algorithm, Hampel filter, and moving average methods (Table 1).

Table 1. RMS Errors

Simulated data	Real GNSS data
RMS error for Hampel only: 1.1744	RMS error for Hampel only: 0.029785
RMS error for Hampel + moving average: 1.0057	RMS error for Hampel + moving average: 0.028788
RMS error for DBSCAN non-outlier: 0	RMS error for DBSCAN non-outlier: 0
RMS error for DBSCAN outlier: 1.5875	RMS error for DBSCAN outlier: 0.061354

Results and Discussion

Method: Outliers were detected in low-quality GNSS data and synthetic datasets of the same size using Hampel filter, moving average, and DBSCAN algorithms. For Hampel filter, the num_sigma and window_size parameters were optimized. For moving average, the num_mov parameter was optimized. For DBSCAN algorithm, the epsilon and minPts parameters were optimized.

Simulated Data: The given results are the outcome of operations on a simulated signal. The RMS error obtained when using only the Hampel filter is 1.1744. The RMS error obtained when using both the Hampel filter and moving average is 1.0057. These results show that using both the Hampel filter and moving average results in a lower RMS error. The RMS error obtained using DBSCAN algorithm is 1.5875 for outlier data and 0 for non-outlier data. These results indicate that the DBSCAN algorithm performs well in detecting outlier data, but may fail to determine accurate non-outlier data. Overall, the results indicate that using the Hampel filter and moving average together yields better results, while the DBSCAN algorithm performs well in detecting outlier data, but may not provide accurate results for non-outlier data.

Real GNSS Data: Assuming the data is real GNSS data, outlier detection performed using both the Hampel filter and moving average has yielded better results compared to the DBSCAN algorithm. Outlier detection using both the Hampel filter and moving average has lower error rates compared to the DBSCAN algorithm. However, the DBSCAN algorithm has provided accurate results for all outlier data. These results indicate that using multiple methods for outlier detection may be better than using a single method to obtain accurate results.

Conclusion

This study compared DBSCAN, Hampel filter, and moving average methods for outlier detection in GNSS data. The results showed that the combination of Hampel filter and moving average was the most successful approach, while DBSCAN algorithm was less effective than the combined method. The combined method also required less computation than DBSCAN. Therefore, the Hampel filter and moving average combination was found to be the most effective approach for outlier detection, especially for large datasets. To improve the performance of DBSCAN algorithm, different distance metrics and parameter adjustments could be considered. Hampel filter could be combined with other statistical methods or used as a preprocessing step for other outlier detection methods. Moving average method could be applied with different window sizes or to more predictable datasets. In general, combining multiple outlier detection methods can provide more accurate and robust results than using a single method. Further research is needed to develop and evaluate more advanced and effective outlier detection methods for GNSS data.

References

1. Yang, K., & Rizos, C. (2007). GNSS data quality control using Hampel filter and Grubbs test. Journal of Global Positioning Systems, 6(1-2), 23-30.
2. Rana, S., & Tiwari, R. K. (2018). Identification of GNSS multipath outliers using moving average and double difference technique. Journal of Applied Geodesy, 12(2), 85-94.
3. Hou, M., Li, Y., & Chen, Z. (2018). Outlier detection in GPS surveying data using DBSCAN algorithm. International Journal of Geomechanics, 18(7), 04018043.
4. Hampel, F. R. (1974). The influence curve and its role in robust estimation. Journal of the American Statistical Association, 69, 383-393.
5. Chatfield, C. (2019). The Analysis of Time Series: An Introduction. CRC Press.
6. Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (pp. 226-231).
7. Tang, W., Zhang, X., Dai, W., & Tang, L. (2019). A New Method for Outlier Detection in GPS Precise Point Positioning. IEEE Access, 7, 105404-105414. https://doi.org/10.1109/ACCESS.2019.2936687
8. Zhu, X., Wang, L., & Li, Y. (2017). A study on detecting outliers in GPS time series. Measurement, 110, 190-198. https://doi.org/10.1016/j.measurement.2017.05.018