6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

A detection method of mismatched measures in GNSS coordinate time series: Fuzzy logic and IQR (Interquartile Range) based approach

Hüseyin Pehlivan*[1]

[1]Gebze Technical University, Engineering Faculty, Department of Surveying Engineering, Kocaeli, Türkiye

Keywords
GNSS data
Outlier
Fuzzy Logic
IQR
Filtering

Abstract

This study presents a new approach for outlier detection aiming to enhance the accuracy and reliability of GNSS data. Unlike traditional approaches, this approach combines fuzzy logic and the Interquartile Range (IQR) method to improve outlier detection. The fuzzy logic-based method is employed to flexibly model data characteristics. Individual outlier scores are calculated for each data point using fuzzy logic, and these scores are then utilized to identify outliers in the dataset. By determining a threshold value based on the spread of the data around its center using the IQR method, data points scoring above this threshold can be considered outliers. The combination of these two methods ensures a more reliable and accurate outlier detection. When applying the proposed approach to a test signal containing obvious outlier values, it is observed that the processed time series exhibits a better normal distribution and improves performance metrics, indicating enhanced signal quality. Experimental results demonstrate the effectiveness of the proposed approach in effectively detecting outliers in GNSS coordinate time series. Overall, the proposed approach offers a promising solution for outlier detection in GNSS data by integrating fuzzy logic and the IQR method. It provides improved accuracy and reliability, leading to enhanced data analysis and interpretation in GNSS applications.

1. Introduction

GNSS (Global Navigation Satellite System) coordinate time series are widely used in determining geographical locations and serve as important data sources in various fields. However, Data obtained from GNSS systems contain outliers that can arise from measurement errors, hardware malfunctions, multipath effects, station-related errors, orbital anomalies, and unknown reasons. In addition to outliers, GNSS position time series are affected by temporally correlated noise, which includes white noise and flicker noise (Mao et al., 1999). These outliers can bias estimates in both functional and stochastic models (Koch 1999; Khodabandeh et al. 2012). Consequently, the presence of outliers can negatively affect the analysis and interpretation of GNSS data, leading to inaccurate results.

Detecting outliers in GNSS position time series is crucial, and various approaches have been proposed for this purpose. There are several approaches for detecting outliers in the GNSS position time series, such as three sigma method (3σ) (Mao et al. 1999), Bayesian method (Zhang and Gui 2013). Besides these methods, the window-opening test algorithm based on the Interquartile Range (IQR) statistic is another commonly used approach for outlier detection in the GNSS position time series (Nikolaidis 2002; Li and Shen 2018). This algorithm is fast and robust since the median and IQR values of a time series are less affected by outliers. Due to its superior performance, the outlier detection approach based on IQR criterion has been widely applied in GNSS position time series analysis.

On the other hand, Aliosmanoğlu and Akyilmaz (2002) conducted a study comparing conventional outlier testing, robust estimation, and fuzzy techniques for detecting outliers within a geodetic network. Similarly, Gökalp ve Boz (2005), Outlier data in GPS networks was identified using both traditional statistical methods and fuzzy logic approach. Cateni et al. (2007) extracted abnormal data caused by various factors using a fuzzy logic-based method and compared it with a statistical technique. Overall, these studies compare fuzzy logic and other methods with various approaches.

In this study, a novel approach is proposed that combines fuzzy logic and IQR (Interquartile Range) methods as an alternative to traditional approaches. The

* **Corresponding Author**

*(hpehlivan@gtu.edu.tr) ORCID ID 0000 – 0002 – 0018 – 6912

Cite this study

Pehlivan, H. (2023). A detection method of mismatched measures in GNSS coordinate time series: Fuzzy logic and IQR (Interquartile Range) based approach. Intercontinental Geoinformation Days (IGD), 6, 220-223, Baku, Azerbaijan

fuzzy logic-based method is employed to flexibly model data characteristics and calculate individual outlier scores for each data point. On the other hand, the IQR method identifies outliers by considering the distribution of the data around its center. By combining these two methods, the proposed approach offers a promising solution for effective outlier detection in GNSS data analysis

The method ensures effective detection of outliers in GNSS coordinate time series. Experimental studies demonstrate its superiority in terms of reliability and accuracy compared to other methods.

In summary, it is essential to accurately detect and differentiate outliers in GNSS coordinate time series. The proposed approach combining fuzzy logic and IQR methods offers an effective solution to this challenge, aiming to improve the accuracy and reliability of GNSS data analysis.

2. Method

2.1. Mamdani Fuzzy Inference System (FIS)

Mamdani Fuzzy Inference System (FIS) is a model of an inference system based on fuzzy logic. Fuzzy logic is an artificial intelligence technique used to solve problems with uncertainty and complexity. Unlike traditional binary (true/false) logic, fuzzy logic operates with fuzzy sets and fuzzy rules.

FIS aims to obtain one or more output values by performing fuzzy inference on input data. It consists of three main components: input variables, output variables, and a rule set.

Input variables are the variables that provide data to the system and define the fuzzy sets associated with them. Each input variable has membership functions that represent different value ranges and fuzzy sets.

Output variables represent the results obtained from the system. Each output variable also has membership functions that represent different value ranges and fuzzy sets.

The rule set defines the relationship between input variables and output variables through if-then rules. Each rule specifies a combination of input variables and the corresponding output variable value. These rules are based on the "if-then" logic of fuzzy reasoning.

Mamdani FIS fuzzifies the input data, applies the rules, and calculates the output values. These output values can be used to solve a specific problem or make a decision. The System is a model used to deal with uncertainty and complexity in the real world. It finds applications in various fields such as control systems, decision support systems, data analysis, artificial intelligence, and many others.

2.2. The Interquartile Range (IQR)

The method is a statistical technique used to assess the spread of a dataset and detect outliers. The IQR represents the difference between the lower quartile (25th percentile) and the upper quartile (75th percentile) of the data.

The IQR method determines outliers by measuring the dispersion around the center of the dataset. In the first step, the data is sorted, and quartiles are calculated. The lower quartile corresponds to the 25th percentile of the data and is also known as the first quartile. The upper quartile corresponds to the 75th percentile of the data.

Next, the IQR is calculated by subtracting the lower quartile value from the upper quartile value. This value represents the spread of the middle 50% of the dataset.

To detect outliers, a threshold value is established using the IQR method. Typically, the lower bound is determined by subtracting 1.5 times the IQR from the lower quartile, and the upper bound is determined by adding 1.5 times the IQR to the upper quartile. Data points outside these bounds are considered outliers.

The IQR method takes into account the density and distribution of the dataset. As a robust and statistically reliable approach based on the center of the data, it is used to identify extreme values in a dataset.

2.3. Fuzzy Logic and IQR based outlier detection

Mathematical Model and Process Steps:

1- The input data is defined.

2- Detection of outliers using Fuzzy Logic:

a. A Mamdani Fuzzy Inference System (mamfis) is created for fuzzy logic-based outlier detection.

b. Input variable (data) and output variable (outlier) are defined.

c. Three membership functions ("low", "medium", "high") are defined for the input variable "data".

d. Two membership functions ("no_outlier", "outlier") are defined for the output variable "outlier".

e. Rules are defined to establish the relationships between the input variable (data) and output variable (outlier).

f. The fuzzy logic system is constructed.

3- Calculation of the output signal:

a. The evalfis function is used on the fuzzy logic system to calculate the output value (outlier_scores) for each point in the dataset.

4- Outlier detection using the Interquartile Range (IQR) method:

a. Q1 is calculated: It represents the lower quartile (25th percentile) of the dataset.

b. Q3 is calculated: It represents the upper quartile (75th percentile) of the dataset.

c. IQR is calculated: It is the difference between Q3 and Q1.

d. The threshold value is calculated as 1.5 times the IQR: threshold = 1.5 * IQR.

e. Outliers are determined as outliers = (outlier_scores < Q1 - threshold | outlier_scores > Q3 + threshold).

5- Examination of outliers:

a. The outliers are presented on the graph.

These steps encompass the procedures and mathematical explanations used to detect outliers in the dataset and generate the output signal through fuzzy logic-based outlier detection.

Figure 1 represents the process steps of fuzzy logic-based outlier detection mentioned in the 2nd step in a more visual manner. In this representation, the input variable "data" passes through membership functions (e.g., "low", "medium", "high"). Then, connections are established using rules (IF-THEN based statements). These rules result in the creation of a Mamdani Fuzzy Inference System (fis). This system produces the "outlier" output based on the given "data" input. Finally, the output values, referred to as "outlier_scores," are obtained and the values will be within the range of [0, 1].

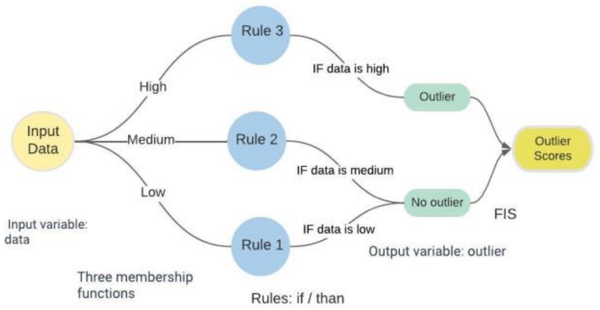

Figure 1. Fuzzy Logic-based outlier detection process.

The obtained 'outlier_scores' variable in this context is calculated for each data point in the dataset as a result of fuzzy logic-based outlier detection. These scores indicate the likelihood of data points being outliers and represent the outcome of the outlier detection process. These scores can be used for outlier detection and help identify outliers in the dataset.

In the 3rd step, the IQR method is used for outlier detection. By comparing the outlier scores obtained with the threshold values determined using IQR, we can determine the outliers. Data points in the input data that fall outside the range of (outlier_scores < Q1 - threshold) or (outlier_scores > Q3 + threshold) are identified as outliers.

3. Results

The input data used in the test consists of 1024 data points collected using the RTK GNSS method with a sampling frequency of 1 Hz over a duration of 17 seconds. This dataset contains evident outliers (Figure 2).

The 'outlier scores' obtained from applying the 2nd step of the Fuzzy Logic-based outlier detection process to the input signal can be seen in Figure 3.

The 'outlier scores' obtained from the Fuzzy Logic-based outlier detection process were used to perform outlier detection using the IQR method as described in the 3rd step. The threshold values for IQR were calculated as follows: Q1 = prctile(data, 25), Q3 = prctile(data, 75), IQR = Q3 - Q1, and threshold = 0.5 * IQR. The points that fell outside the threshold, determined as Outlier = (outlier_scores < Q1 - threshold or outlier_scores > Q3 + threshold), were identified as outliers.

Calculated index numbers of identified outliers:

Outlier indexs = [36, 116, 117, 118, 136, 137, 145, 146, 156, 157, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 182, 183, 184, 188, 189, 190, 194, 195].

Figure 4 shows the input data, outlier values, and corrected output data.

When examining the histograms of the input and output signals, the high and narrow peak observed in the histogram graph of the input signal indicates the potential outliers in the data. It can be seen that the output signal exhibits a better normal distribution (Gaussian distribution) (Figure 5).

Additionally, to evaluate the performance of the output signal, metrics such as RMS (Root Mean Square), signal power, noise power, and SNR (Signal-to-Noise Ratio) values can provide information about the quality of the output signal.

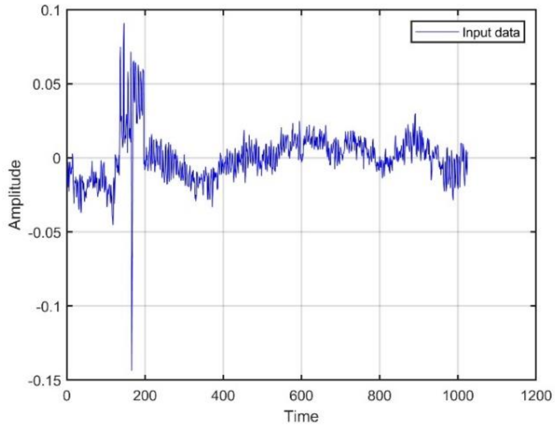

Figure 2. The input data

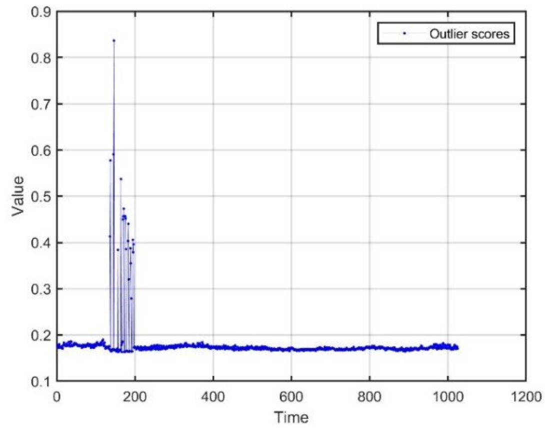

Figure 3. The "outlier scores" of the input data obtained from the fuzzy Logic-based outlier detection process

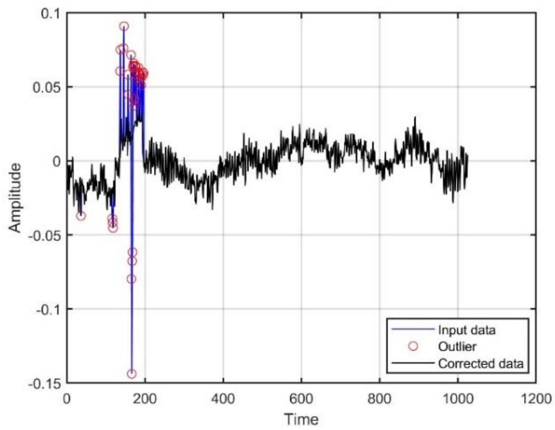

Figure 4. Fuzzy Logic and IQR based method applied signal outputs

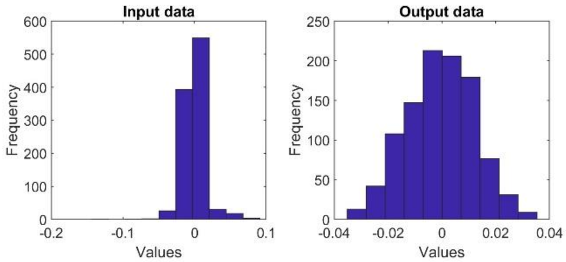

Figure 5. The histograms of input and output signals.

After applying the proposed model, the accuracy measures of the input and output signals are presented in Table 1.

Table 1. Accuracy metrics for input and output signals

Items	Input Data	Output Data
RMS	0.1670	0.1260
Signal Power	0.0028	0.0016
Noise Power	0.0027	0.0001
SNR	-0.0001	2.2182

4. Discussion

RMS value of the input dataset is calculated as 0.1670, while it is 0.1260 for the output dataset. This indicates that the output data is more stable and exhibits less variation compared to the input data. The signal power of the input dataset is 0.0028, whereas it decreases to 0.001 in the output dataset. This suggests that the output dataset has a lower signal power. The noise power of the input dataset is 0.0027, whereas it is calculated as 0.0001 for the output dataset. The significantly lower noise power in the output dataset compared to the input data indicates that the output signal contains less noise. The SNR value of the input dataset is -0.0001, while the SNR value of the output dataset is 2.2182. The positive and high SNR value of the output dataset indicates that it has a better signal-to-noise ratio compared to the input signal. These evaluations demonstrate that the output signal has lower RMS, signal power, and noise power values compared to the input signal, and it has a higher SNR value. This suggests that the output signal is cleaner, less noisy, and more stable.

5. Conclusion

In this study, a novel approach for outlier detection in GNSS coordinate time series was developed using fuzzy logic. The proposed method calculates a fuzzy logic-based outlier score for each data point, indicating its probability of being an outlier. To detect outliers, the IQR method was employed to obtain outlier scores. These outlier scores were then compared with threshold values determined by the IQR. Data points in the input data that exceeded the threshold range of outlier scores were identified as outliers. The results demonstrate that this approach is effective for outlier detection in GNSS coordinate time series.

Accurate detection of outliers enhances the overall accuracy and reliability of GNSS data. Moreover, the proposed method provides a general framework that can be applied to other domains of data analysis.

In conclusion, this study showcases the successful application of a combined fuzzy logic and IQR approach for outlier detection in GNSS coordinate time series. The proposed method offers a valuable tool for obtaining more reliable results in data analysis.

References

Aliosmanoğlu, Ş., & Akyilmaz, O. (2002). A comparison between statistical & fuzzy techniques in outlier detection. In Vistas for Geodesy in the New Millennium: IAG 2001 Scientific Assembly, Budapest, Hungary September 2–7, 2001 (pp. 382-387). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-04709-5_63

Cateni, S., Colla, V., & Vannucci, M. (2007, February). A fuzzy logic-based method for outliers detection. In Artificial Intelligence and Applications (pp. 605-610).

Gökap, E., & Boz, Y. (2005). Outlier detection in GPS networks with fuzzy logic and conventional methods. From Pharaohs to Geoinformatics. FIG Working Week.

Khodabandeh, A., Amiri-Simkooei, A. R., & Sharifi, M. A. (2012). GPS position time-series analysis based on asymptotic normality of M-estimation. Journal of Geodesy, 86, 15-33.

Koch, K. R. (1999). Parameter estimation and hypothesis testing in linear models. Springer Science & Business Media.

Li, W., & Shen, Y. (2018). The consideration of formal errors in spatiotemporal filtering using principal component analysis for regional GNSS position time series. Remote Sensing, 10(4), 534.

Mao, A., Harrison, C. G., & Dixon, T. H. (1999). Noise in GPS coordinate time series. Journal of Geophysical Research: Solid Earth, 104(B2), 2797-2816.

Nikolaidis, R. (2002). Observation of geodetic and seismic deformation with the Global Positioning System. University of California, San Diego.

Qianqian, Z., & Qingming, G. (2013). Bayesian methods for outliers detection in GNSS time series. Journal of Geodesy, 87(7), 609-627.