J. Int. Environmental Application & Science, Vol. **19**(2): 57-68 (2024)
Research Paper

Design and Implementation of a Graphical User Interface for Outlier Data Analysis: A Case Study on the Yeşilırmak River**

Eda Göz[1, *], Zübeyde Zengül[2], Erdal Karadurmuş[2], Mehmet Yüceer[3]

[1]*Ankara University Engineering Faculty, Chemical Engineering Department, 06100, Ankara, Türkiye;* [2] *Hitit University Engineering Faculty, Chemical Engineering Department, 19200, Çorum, Türkiye;*[3] *İnönü University Engineering Faculty, Chemical Engineering Department,44280, Malatya, Türkiye;*

Received December 19;2023; Accepted June 13, 2024

Abstract: Water quality control, especially in large-scale monitoring regions or networks, requires easy and automatic processes for detecting potential outliers in a reproducible manner. This study focuses on removing outlier values from a dataset collected by an online monitoring station on the Yeşilırmak River between 2007 and 2009. Seven different parameters were evaluated: dissolved oxygen (luminescence dissolved oxygen, LDO), temperature, pH, conductivity, total organic carbon (TOC), nitrate nitrogen (NO_3-N), and ammonium nitrogen (NH_4-N). Five methods – median, mean, Grubbs', generalized extreme studentized deviate (GESD), and interquartile range (IQR) – were used for outlier removal. The developed models were integrated into a graphical user interface (GUI) in the MATLAB environment, facilitating practical and easy access. This study enables users to input any dataset into the software and remove outlier values using various methods in a few steps, thus preparing the data for modeling studies. It was observed that the median algorithm removed the most data points among the outlier data-removal methods.

Keywords: *River water quality data, outlier detection, statistical methods, graphical user interface (GUI)*

Introduction

Water is an indispensable resource for all living organisms, playing a critical role in sustaining life and maintaining environmental balance. In recent years, the rapid escalation in population growth, coupled with intensified industrial activities and unsustainable agricultural practices, has significantly heightened the strain on global water resources. This situation has prompted a surge in the establishment of governmental standards for water quality monitoring, underscoring the necessity of robust and efficient monitoring systems.

As delineated by Berendrecht *et al*. (2023), the process of water quality monitoring is multifaceted, encompassing the definition of objectives, the selection of appropriate methodologies, and culminating in the detailed evaluation of results. A critical component of this process is the development of an effective monitoring network, which is particularly crucial in continuous monitoring scenarios. Selecting the most suitable installation, sampling, and measurement techniques is vital for ensuring accurate and reliable data collection.

The advent of online and real-time monitoring systems has marked a significant advancement in the field. These systems provide comprehensive district-level control and enable detailed trend analysis, thereby accumulating substantial volumes of data. Alongside these technological advancements, the evolution of data mining and data logging technologies has greatly facilitated the processing and interpretation of this data. However, a major challenge in data interpretation arises from the presence of outliers, which may result from environmental fluctuations, human errors, or sensor inaccuracies. Outlier detection is thus a critical preliminary step in data preprocessing, essential for enhancing data reliability and facilitating accurate analysis. The preprocessing typically involves two main steps: standardizing data to a mean of zero and a standard deviation of one and applying statistical methods to render time-series data stationary (Jamshidi *et al*., 2022).

Various methods for outlier detection in surface water have been extensively documented in the literature. For instance, Muñiz *et al*. (2012) employed functional data analysis to detect outliers in water quality data within the San Esteban estuary. Cho *et al*. (2013) introduced a two-step process for

*Corresponding: E-Mail: esemizer@eng.ankara.edu.tr, Tel: +903122033526. Fax: +903122121546

**This study was presented as an oral presentation in the International Symposium for Environmental Science and Engineering Research 2023, which will be held on 18-21 October 2023

removing outliers in ocean temperature data, which involved using approximate and complex components coupled with harmonic analysis. Saberi (2015) enhanced the univariate method by implementing an auto-regressive moving average model within a moving data window. Furthermore, Di Blasi et al. (2015) analyzed several water quality parameters for outlier detection, utilizing a functional data analysis approach. Additional contributions to the field have been made by researchers such as Plazas-Nossa et al. (2016), Jingang et al. (2017), Dogo et al. (2019), Bae et al. (2019), and Sun et al. (2019). Each of these studies introduced varying methodologies for the effective detection of outliers, thereby enriching the existing body of knowledge.

Despite the abundance of research in outlier detection, a notable gap exists in the literature concerning the design of graphical user interfaces (GUIs) specifically for this purpose. This study addresses this gap by developing a user-friendly interface that applies five distinct statistical outlier detection methods to seven crucial water quality parameters: Luminescent Dissolved Oxygen (LDO), temperature, conductivity, pH, Total Organic Carbon (TOC), Nitrate Nitrogen (NO_3-N), and Ammonium Nitrogen (NH_4-N). Integrating statistics-based models, this interface significantly enhances the practicality and efficiency of data preprocessing for subsequent modeling studies.

Materials and Methods

Study Area and Data Collection

Data for this research were obtained from two real-time in-situ monitoring stations. The first station was located in the Aynalı Cave area, downstream from the sewage system and Tersakan stream. The second station was positioned at the Administration of Hydraulic Works' Durucasu site, approximately 26.876 kilometers from the Aynalı Cave station and further downstream from the yeast factory. A monitoring office was established at Ankara University to facilitate central monitoring and data management. Data from these stations were transmitted to the university via GPRS technology. Figure 1 depicts the geographical positions of the monitoring stations and outlines the data collection and transmission process.

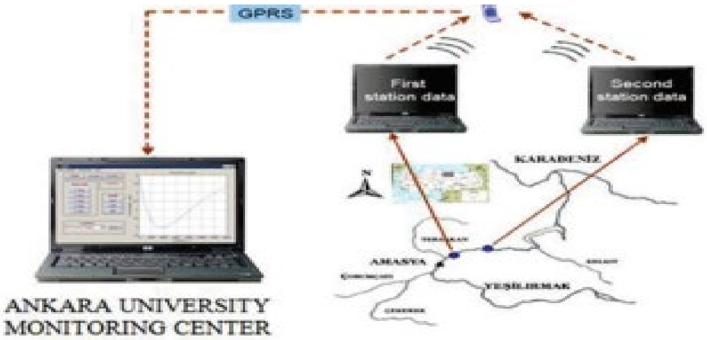

Figure 1. Locations of the online measurement stations and the data collection & transmission process

The selected study area encounters diverse pollution sources. The stations were programmed to record real-time data at five-minute intervals. This data encompassed various parameters, including luminescent dissolved oxygen (LDO), pH, conductivity, nitrate nitrogen (NO_3-N), ammonium nitrogen (NH_4-N), total organic carbon (TOC), orthophosphate, chloride, temperature, turbidity, suspended solids, and flow rate. This study focused on seven parameters with 18,815 data points: LDO, pH, conductivity, temperature, TOC, NO_3-N, and NH_4-N.

Outlier Detection Methods

Outlier detection in statistical analysis can be broadly categorized into parametric and non-parametric methods, depending on the data distribution. Parametric tests, such as Grubbs' and Generalized Extreme Studentized Deviate (GESD) tests, are typically applied to datasets with uniform distributions. It is important to recognize that there is no universally applicable method for automatically cleaning outlier data; the selection of a test often varies based on the specific requirements of the study. In this research, we employed five distinct methods for outlier detection, each suited for different data characteristics:

Mean Method for Outlier Detection: The mean, also known as the arithmetic mean, is a fundamental statistical measure calculated as the sum of all values in a dataset divided by the number of values. While straightforward, this method can be less robust than the median method, especially in the presence of extreme outliers. The mean is calculated using the following formula (Eq.1):

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

Where:
μ represents the mean value.
n is the total number of values in the dataset.
x_i represents the ith value in the dataset.
Standard deviation (σ) measures the spread of data values around the mean. It is calculated as (Eq.2):

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad (2)$$

Where: σ represents the standard deviation.

Outliers are typically identified using the Z-score, which indicates how many standard deviations a value is from the mean (Eq.3):

$$Z = \frac{z_i - \mu}{\sigma} \qquad (3)$$

A value is generally considered an outlier if $|Z|>3$, meaning it is more than three standard deviations away from the mean.

This method presupposes that the data follow a normal distribution and is sensitive to extreme values, which can markedly affect both the mean and standard deviation. Hence, it is the most effective in datasets that are normally distributed and do not contain extreme outliers.

Median Method for Outlier Detection: The median method for outlier detection is a non-parametric approach that is particularly useful when dealing with skewed data distributions or when the dataset contains extreme values (outliers) that could significantly influence the results of parametric methods, such as those based on the mean.

In this method, the data is sorted from smallest to largest or smallest when it is the middle value of the distribution. Outliers are more than three scales from the median or as MAD (mean absolute deviations) element, which is the absolute deviation of the arithmetic mean.

Outliers are data points that lie a significant distance from the median. A common threshold is data points that are more than 1.5 times the MAD above or below the median. This threshold can be adjusted based on the specific context or distribution of the data. The median method is more robust to outliers than the mean. Since it uses the median and MAD, it is not as affected by extreme values.

Suitability for Non-Normal Distributions: This method does not assume a normal distribution of data, making it suitable for a broader range of datasets, particularly those that are skewed or have heavy tails. However, the method might be less sensitive in detecting subtle outliers, especially in symmetric distributions. The scaled MAD is calculated by using Equation 4.

$$\mathrm{MAD} = \mathrm{median}(|x_i - \mathrm{median}(x)|) \qquad (4)$$

The Median method is renowned for its wide applicability across various fields, including finance, environmental science, and engineering. These disciplines often encounter data that are skewed or contain extreme values, making the Median method particularly useful. Its robustness against outliers renders it an invaluable tool in these areas, where accurate data representation is crucial.

In addition to its practical applications, the Median method plays a pivotal role in the initial stages of data cleaning. Before delving into more complex statistical analyses, this method provides an efficient way to cleanse the dataset, ensuring that subsequent analyses are not skewed by outliers.

The significance of the Median method is further underscored by its prominent presence in statistical education and literature. It is a well-established technique, extensively covered in numerous statistical textbooks and academic papers. This widespread coverage is a testament to its effectiveness in managing datasets with outliers, highlighting its importance as a foundational tool in statistical analysis.

Grubbs' Test for Outlier Detection: Grubbs' Test, introduced by Frank E. Grubbs in 1969, is a statistical method designed to detect outliers in datasets that are assumed to follow a normal distribution. The test is based on two hypotheses:

Null Hypothesis (H_0): There are no outliers in the dataset.
Alternative Hypothesis (H_1): The dataset has at least one outlier.
One-sided Left-tailed Test: This version of the test assesses the probability of the minimum value being an outlier. It is calculated using Equation 5:

$$\mathrm{Gmin} = \frac{\bar{x} - x_{min}}{\sigma} \qquad (5)$$

One-sided Right-tailed Test: This test evaluates the probability of the maximum value being an outlier, as given in Equation 6:

$$\mathrm{Gmax} = \frac{x_{max} - \bar{x}}{\sigma} \qquad (6)$$

Double-sided Test: This test investigates both the minimum and maximum values for potential outliers. The calculation of the critical G value is outlined in Equation 7:

$$G = \max(\mathrm{Gmax}, \mathrm{Gmin})\,;\ G_{critic} \approx \frac{(N-1)t_{\frac{\alpha}{2N}N(N-2)}}{\sqrt{n(n-2+t^2_{\frac{\alpha}{2N}N(N-2)})}} \qquad (7)$$

Here, σ is the standard deviation, $t_{\frac{\alpha}{2N}N(N-2)}$ is critical table value of t distribution, N-2 is degrees of freedom.

A notable limitation of Grubbs' Test, as identified by Chelishchev et al. (2018), is the masking effect. This effect arises in datasets containing multiple outliers, where the presence of one outlier can conceal or 'mask' another. The result is that once the most obvious outliers are removed, others may appear that were not initially detectable. Ben-Gal (2005) discusses how this effect can complicate the outlier detection process, especially in datasets with numerous outliers. Similarly, Vera et al. (2013) describes the 'swamping effect', where non-outliers may be incorrectly identified as outliers in the presence of actual outliers. These phenomena highlight the need for careful analysis when using Grubbs' Test, particularly in datasets with potential multiple outliers.

Despite this limitation, Grubbs' Test is extensively utilized in various scientific and engineering disciplines, notably in laboratory quality control settings. In such environments, the accurate detection of single aberrant measurements is often critical. The test is also valuable in other research areas where the integrity of data is paramount and the assumption of a normal distribution is valid.

Grubbs' Test is recognized as a standard method in statistical outlier detection and is extensively covered in various statistical textbooks and academic papers. Its importance is also reflected in its widespread implementation in numerous statistical software packages and programming languages, making it an accessible and practical tool for data analysts and researchers.

Generalized Extreme Studentized Deviate (GESD) Test: The GESD Test, commonly referred to as the Rosner test, is a parametric outlier detection method designed to identify outliers in datasets where the number of outliers is not precisely known, and the data is assumed to follow a normal distribution. This test is particularly adept at calculating both single and multiple two-tailed outliers. Similar to Grubbs' Test, the GESD Test operates on two fundamental hypotheses:
Null Hypothesis (H_0): The dataset contains no outliers.
Alternative Hypothesis (H_1): The dataset contains at least one outlier.
In the GESD Test, Equation 8 is calculated for each data point x_i, where $\bar{x}$ is the arithmetic mean and σ is the standard deviation of the dataset.

$$R_i = \frac{x_i - \bar{x}}{\sigma} \qquad (8)$$

The test sequentially identifies the most extreme observation (outlier) and then recalculates the test statistics for the remaining data. This process continues, iteratively identifying and removing outliers.

Unlike Grubbs' Test, the GESD Test can adjust the critical values based on the number of data points remaining after each iteration, allowing for more dynamic outlier detection. This feature makes the GESD Test particularly valuable in scenarios where multiple outliers are suspected, requiring only an upper limit on the number of potential outliers (Grubbs, 1969).

The GESD Test is highly useful in fields where the precise identification of multiple outliers is crucial, such as in quality control, laboratory data analysis, and environmental data assessment. Its implementation in various statistical software packages makes it a practical tool for rigorous outlier detection in statistical analyses. The GESD Test is well-documented in statistical literature and represents a significant enhancement of the Grubbs' Test. It offers greater flexibility in handling complex datasets that may contain multiple outliers, making it a valuable extension in the domain of statistical outlier detection.

Interquartile Range (IQR) for Outlier Detection: The Interquartile Range (IQR) is a critical statistical measure used to identify outliers, particularly in datasets that are not normally distributed or contain a significant number of extreme values. The IQR is the difference between the 75th percentile (Q_3) and the 25th percentile (Q_1) of the dataset, representing the range of the middle 50% of the data.

Outliers in a dataset are typically identified as data points that fall above or below 1.5 times the IQR from Q_1 and Q_3, respectively. Values outside this range are considered outliers. This method provides a robust way to determine outliers, particularly in data with skewed distributions or extreme values (Bonakdari et al., 2022).

Grubbs' Test: Highly effective for datasets with suspected single outliers and a normal distribution. GESD Test: More suitable for identifying multiple outliers in normally distributed data. Median Method: Offers robustness in skewed distributions or datasets with extreme values. Mean Method: Generally effective but can be unreliable when extreme outliers are present. The selection of an appropriate outlier detection method should be tailored to the specific nature of the data and the objectives of the analysis. It is often advantageous to employ multiple methods to cross-validate results, particularly in complex datasets. This approach ensures a comprehensive understanding and accurate identification of outliers.

Graphical User Interface Design

In this study, we developed a user-interactive graphical interface within the MATLAB (Matrix Laboratory) environment. This interface facilitates the outlier removal process, enhancing usability and efficiency for data analysts. The interface's general layout is depicted in Figure 2.a. Upon uploading a dataset to the software, the interface automatically calculates and displays the number of variables present in the dataset, as illustrated in Figure 2.b.

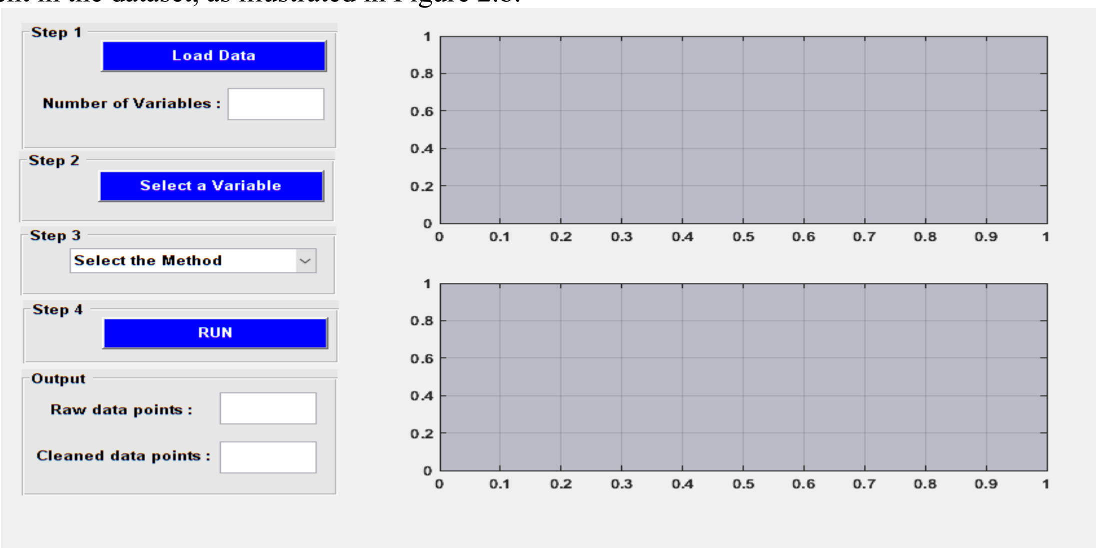

Figure 2.a. Overview of the Graphical User Interface

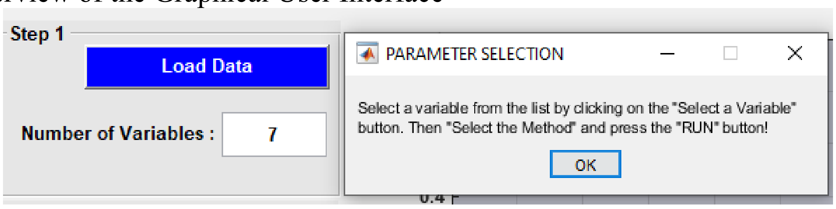

Figure 2.b. Display of the Number of Variables in the Dataset

As indicated in Figure 2.c, the user selects the specific parameter to undergo the outlier cleaning process. Subsequently, the user selects the desired outlier removal method, and the program initiates the process (Figure 2.d). Post-processing, the program provides a comparative display of the original (raw) and cleaned data counts. Additionally, graphical representations of both datasets are showcased on the interface's right side, as seen in Figure 2.e. This interface offers flexibility, allowing users to alter parameters and methods to observe varied results. The cleaned data, post-processing, are saved by the software, enabling users to utilize this refined dataset for further analysis or application as required.

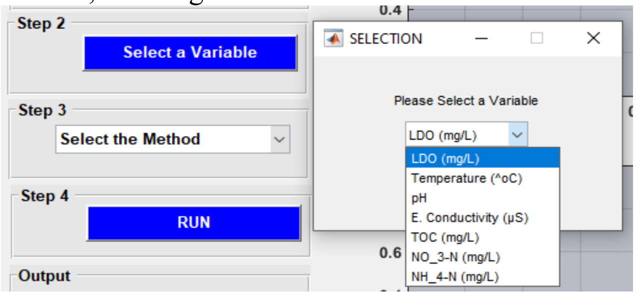

Figure 2.c. Parameter Selection for Outlier Removal

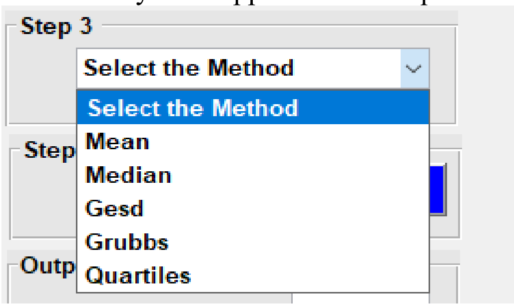

Figure 2.d. Method Selection for Outlier Removal

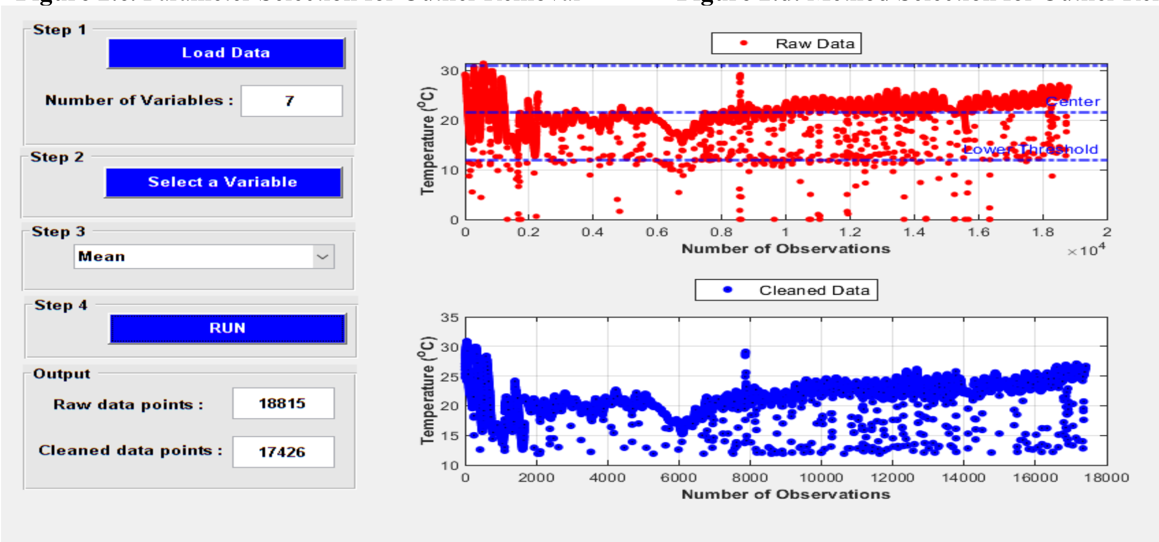

Figure 2.e. Results Display Post-Processing

Results and Discussion

The results of the outlier detection methods, as applied to various parameters through the developed GUI, are presented in this section. It's important to note that employing the same method across different parameters in the matrix-format dataset results in the same number of cleaned data points. This is because outlier values in one parameter necessitate the removal of corresponding values across all parameters.

Outlier Detection Results for LDO Parameter: The outcomes of different outlier detection methods for the LDO parameter are illustrated in Figure 3.

Analyzing Figure 3, which uses 18,815 original data points, reveals that the median method detected the highest number of outliers, whereas the Grubbs method detected the fewest.

Outlier Detection Results for Temperature Parameter: The results for the temperature parameter, as displayed by the GUI, are depicted in Figure 4.

In this graph, data points outside the defined lower and upper limits were classified as outliers. The graph with blue dots represents the data remaining post-cleaning.

Outlier Detection Results for pH Parameter: Figure 5 presents the outlier detection results for the pH parameter. Outliers are visibly identifiable as data points exceeding the upper or lower limits in the graph.

Outlier Detection Results for Conductivity Parameter: The results for the conductivity parameter are provided in Figure 6.

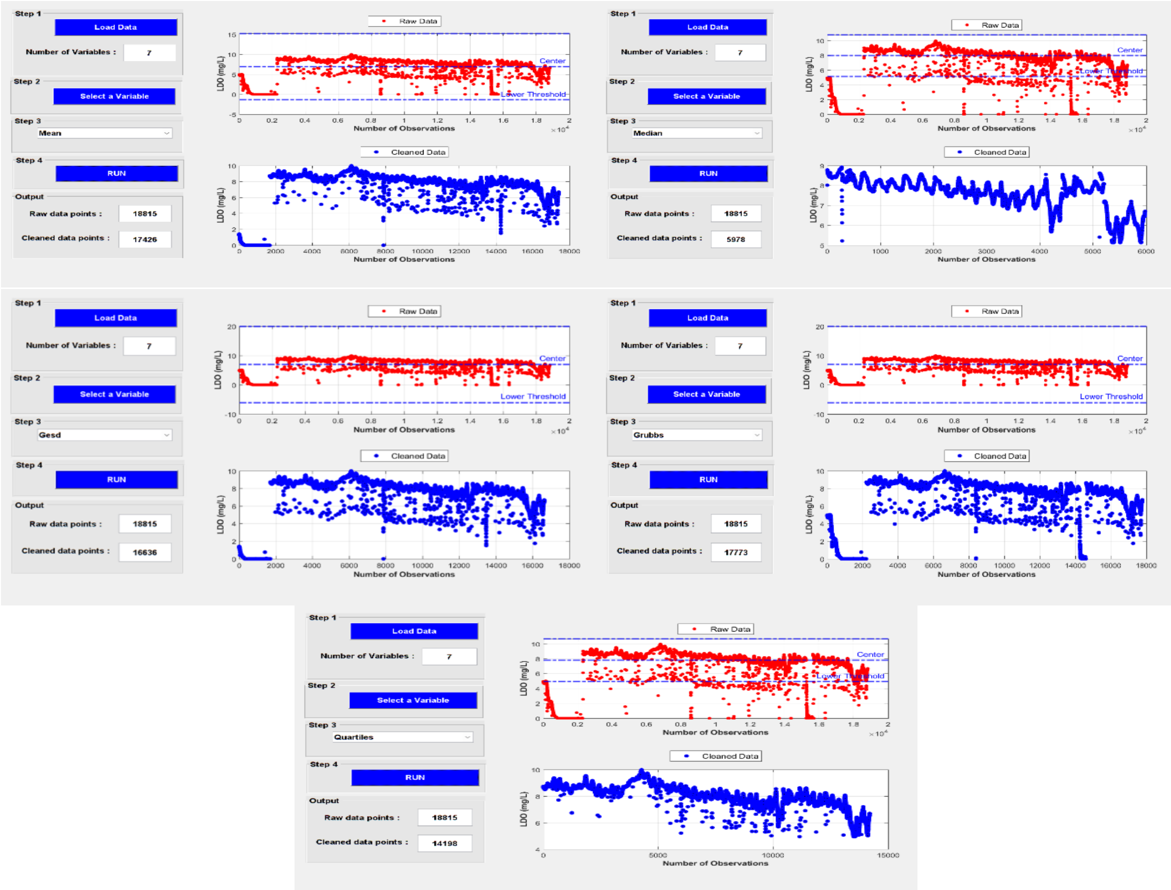

Figure 3. Outlier Detection Results for LDO Parameter (Mean, Median, GESD, Grubbs, Quartiles)

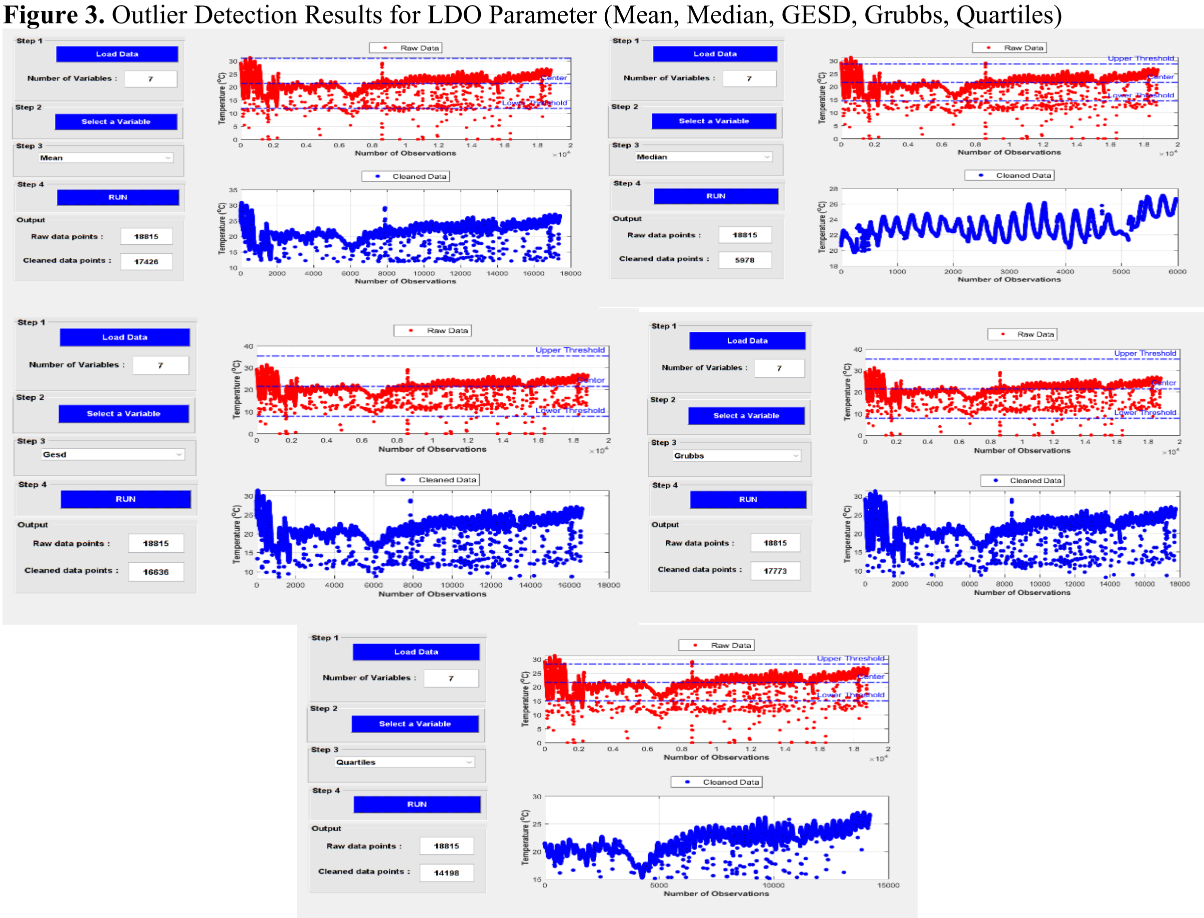

Figure 4. Outlier Detection Results for Temperature Parameter (Mean, Median, GESD, Grubbs, Quartiles)

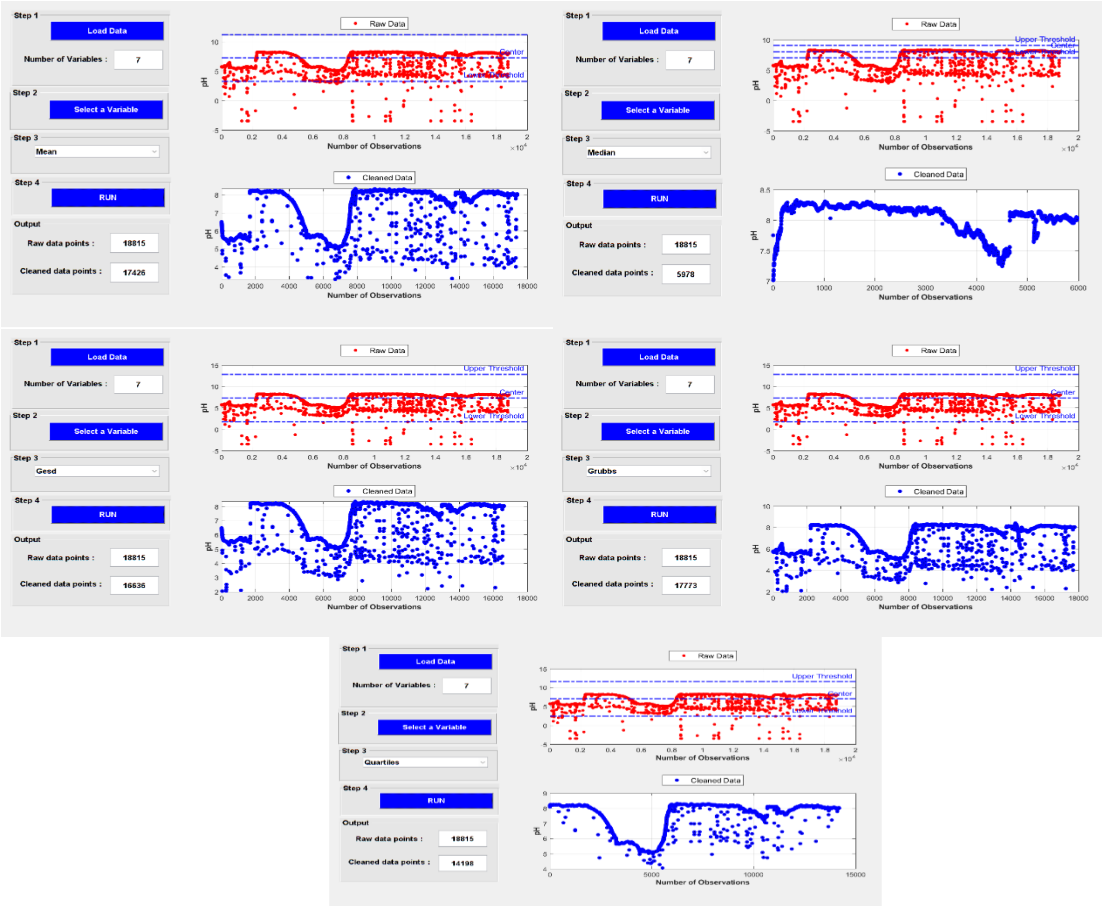

Figure 5. Outlier Detection Results for pH Parameter (Mean, Median, GESD, Grubbs, Quartiles)

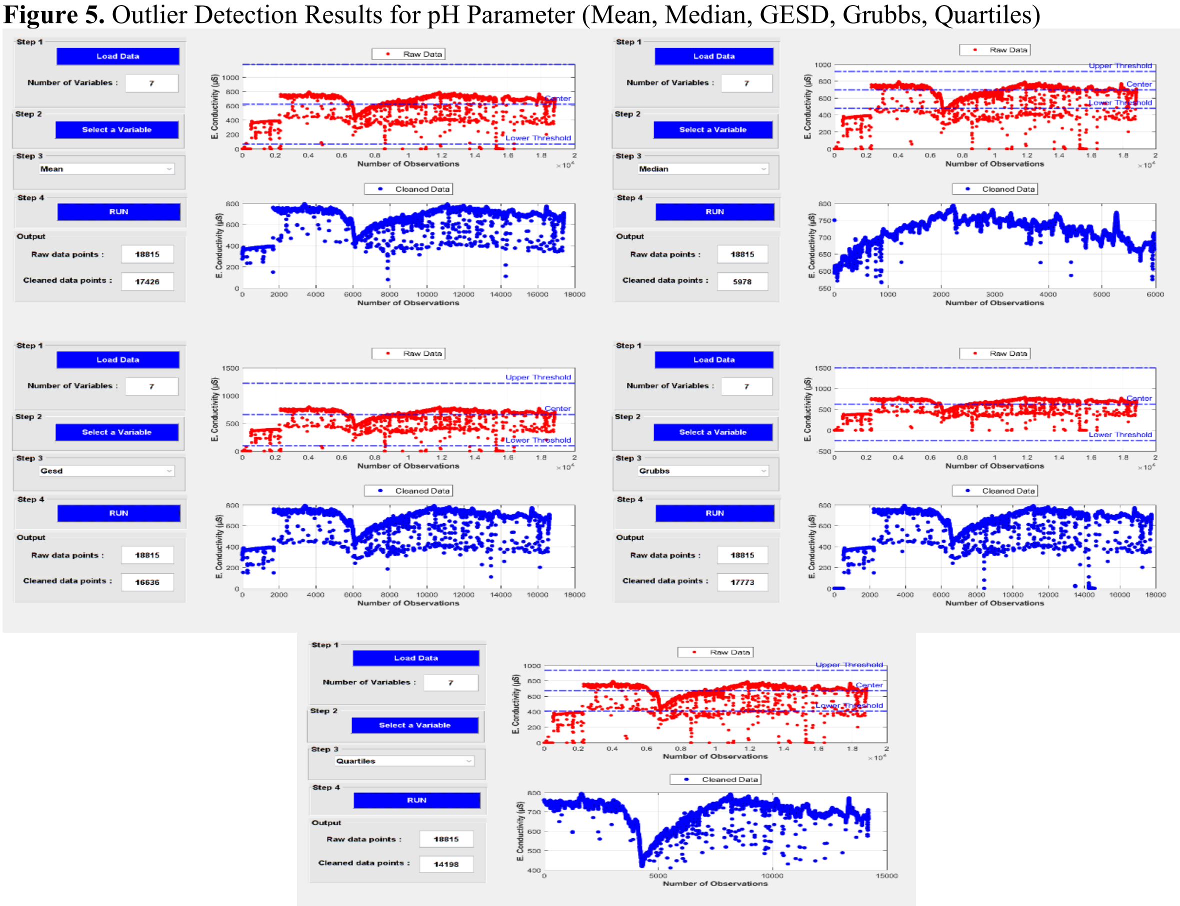

Figure 6. Outlier Detection Results for Conductivity Parameter (Mean, Median, GESD, Grubbs, Quartiles)

Outlier Detection Results for TOC Parameter: The outcomes for the TOC parameter are shown in Figure 7.

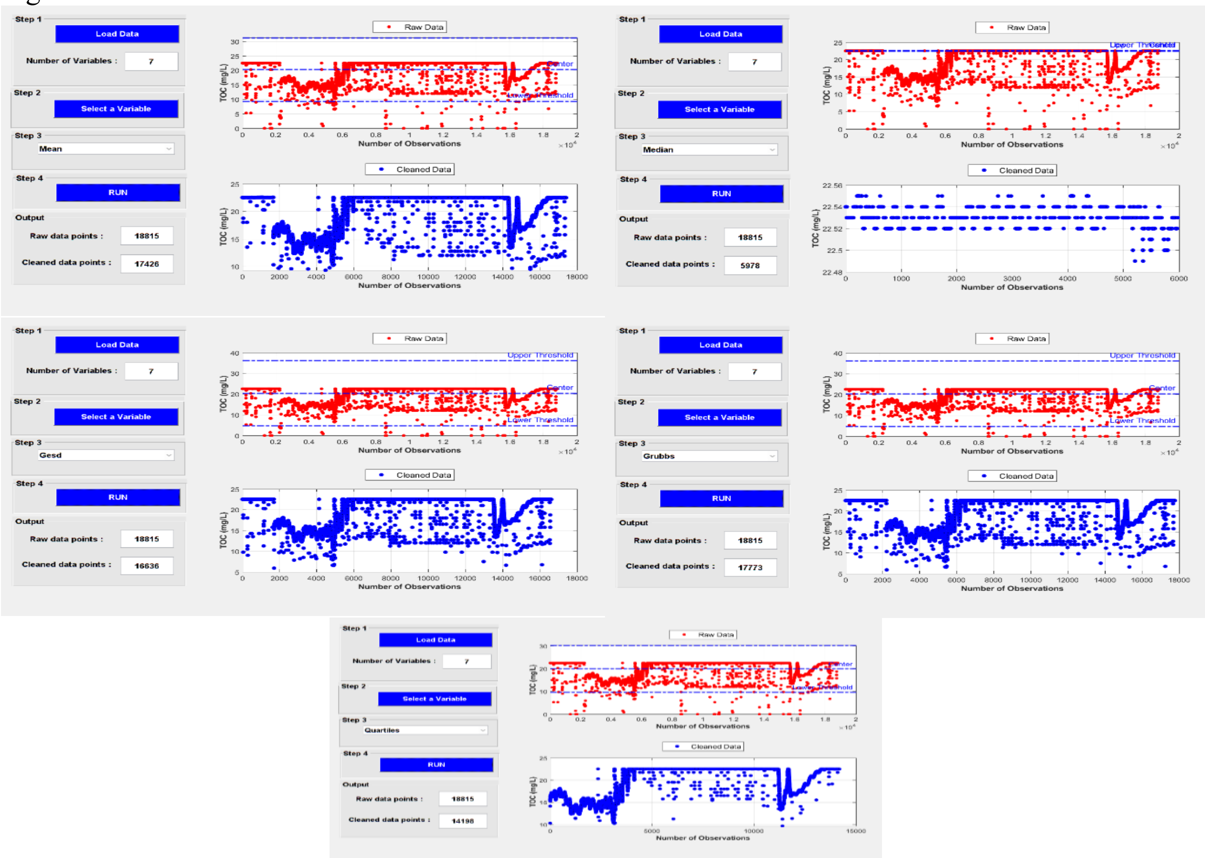

Figure 7. Outlier Detection Results for TOC Parameter (Mean, Median, GESD, Grubbs, Quartiles)

Outlier Detection Results for NO_3-N and NH_4-N Parameters: The detection of outliers for the nitrogen parameters (NO_3-N and NH_4-N) is displayed in Figures 8 and 9.

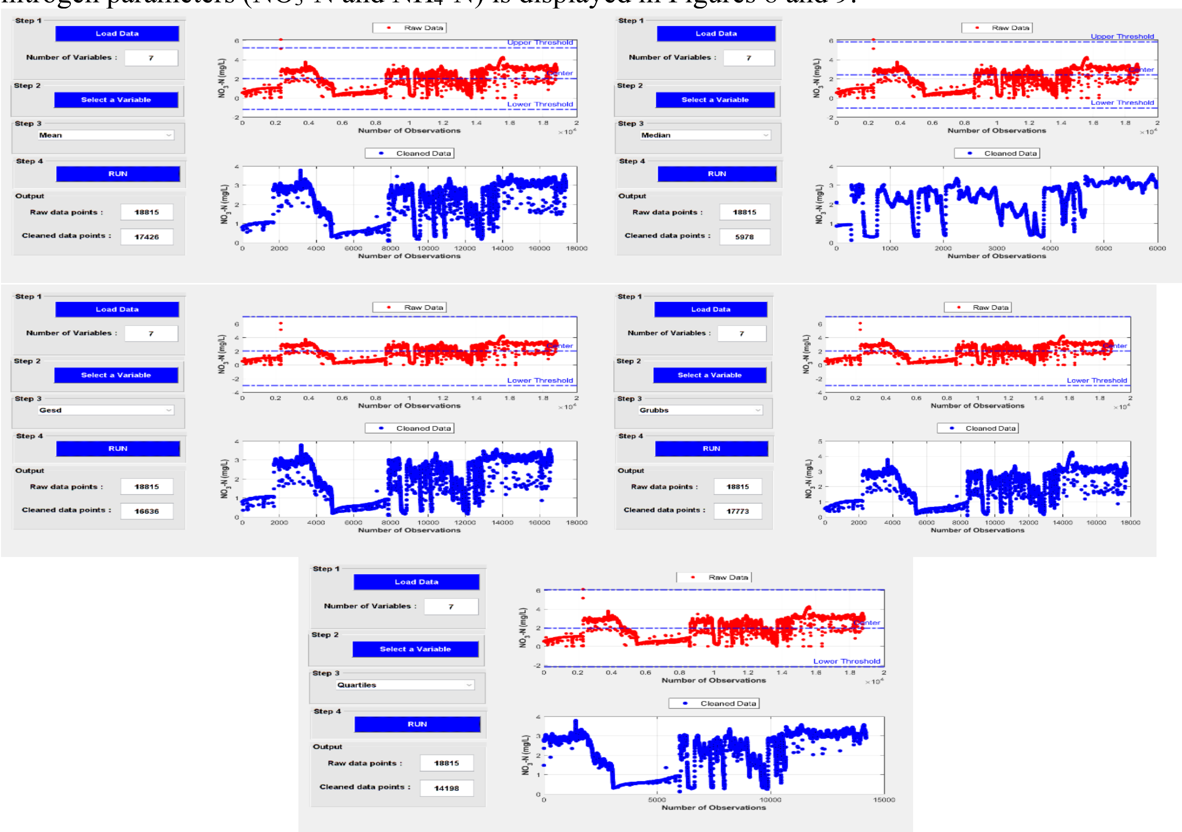

Figure 8. Outlier Detection Results for NO_3-N Parameter (Mean, Median, GESD, Grubbs, Quartiles)

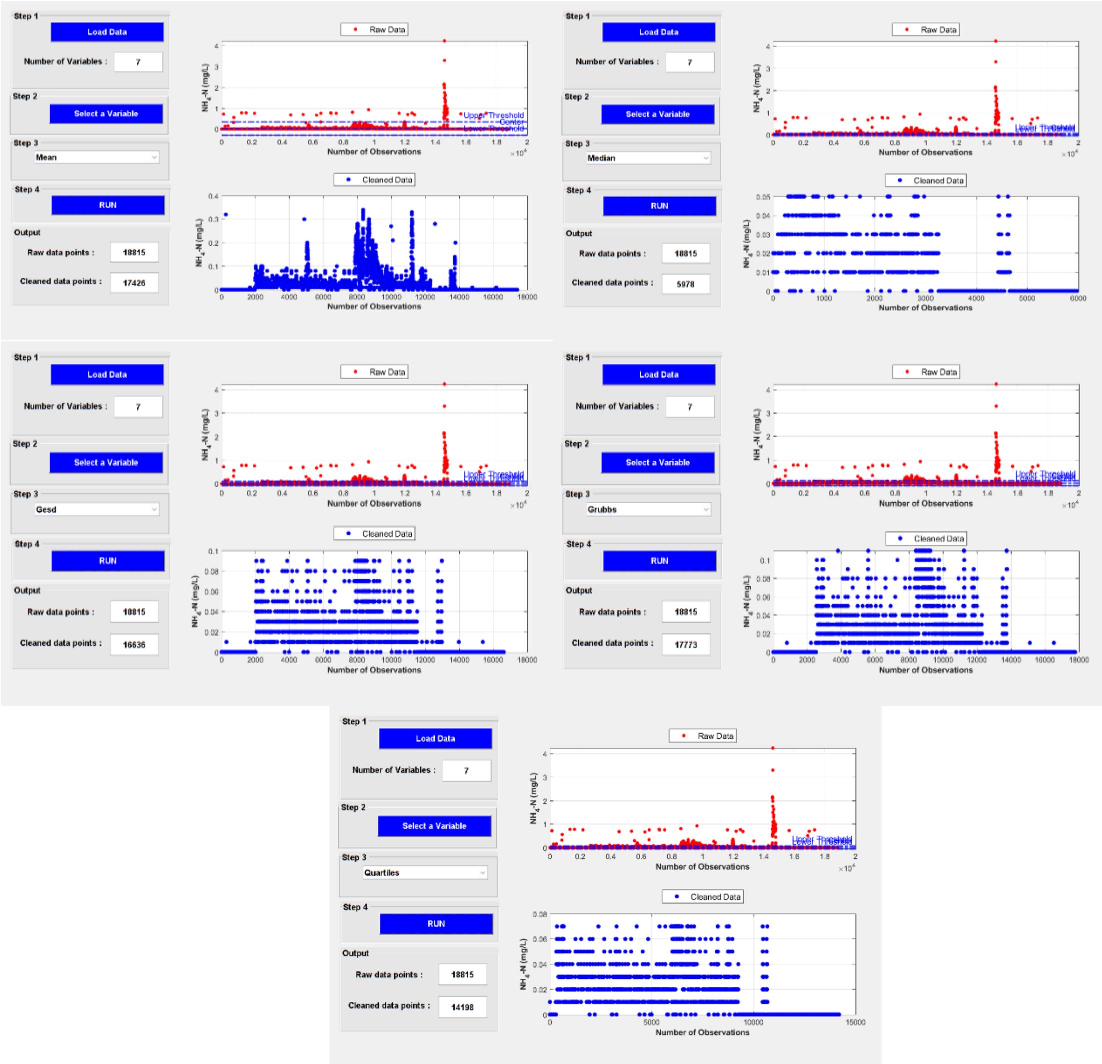

Figure 9. Outlier Detection Results for NH_4-N Parameter (Mean, Median, GESD, Grubbs, Quartiles)

The comparative analysis of original and cleaned data is succinctly summarized in Table 1. According to the results presented in Table 1, Grubbs' Test identified the least number of outliers in the dataset, indicating its conservative nature in outlier detection. Conversely, the Median method detected the highest number of outliers, reflecting its sensitivity to extreme values. These findings underscore the varying effectiveness of different outlier detection methods depending on the nature of the data.

Table 1. Original and Remaining Data Numbers by Method

Methods	Original Data Number	Remaining Data Number
Mean	18,815	17,426
Median	18,815	5,978
GESD	18,815	16,636
Grubbs	18,815	17,773
Interquartile	18,815	14,198

This study on the design and implementation of a graphical user interface (GUI) for outlier data analysis on the Yeşilırmak River differs from existing literature, such as TODS (Lai et al., 2021), which focuses on time series data outlier detection, and other studies involving MATLAB-based GUIs for IoT sensor measurements (Bashir et al., 2022) and patient monitor event logs (Friel et al., 2004), by specifically targeting environmental data from water quality monitoring, emphasizing the integration of multiple statistical methods for robust outlier detection in river data.

The next phase of this study will involve comprehensive modeling studies. These will provide a deeper insight into the performance of the various outlier detection algorithms used. In the modeling stage, the impact of each outlier detection method on the accuracy and reliability of the models will be critically assessed. This step is crucial, as it will help determine the most effective pre-processing technique for enhancing the quality of the dataset, thereby ensuring robust and accurate modeling outcomes.

Conclusion

This study successfully developed a comprehensive and user-friendly graphical user interface (GUI) for outlier detection, integrated with advanced algorithms. Utilizing a dataset of 18,815 data points, we focused on six critical river quality parameters: Luminescent Dissolved Oxygen (LDO), temperature, pH, conductivity, Total Organic Carbon (TOC), Nitrate Nitrogen (NO_3-N), and Ammonium Nitrogen (NH_4-N). These parameters were meticulously cleaned and preprocessed for future modeling studies.

In our analysis, five different outlier detection methods were employed: Mean, Median, Generalized Extreme Studentized Deviate (GESD), Grubbs, and Interquartile Range (IQR). Among these, the Median method was found to detect the highest number of outliers, indicating its sensitivity and effectiveness in outlier identification.

The GUI developed in this study significantly simplifies the process of data loading and outlier detection. It allows users to efficiently process data and obtain results with just a few clicks. This feature is particularly advantageous for researchers and practitioners in environmental science and data analysis, streamlining their workflow and enhancing productivity.

The cleaned datasets, obtained after the outlier removal process, are primed for use in various predictive and analytical modeling applications. The effectiveness of these subsequent models will further validate the efficiency of the pre-processing methods used in this study. This approach not only contributes to the field of environmental data analysis but also sets a precedent for future research in data preprocessing and model optimization.

Acknowledgment: *We thank the Hitit University for encourage given to this study. The authors would like to thank the Editor-in-Chief for editorial suggestions. A special thanks go to the reviewers*

Compliance with Ethical Standards Ethical responsibilities of Authors: *The author has read, understood, and complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors".*

Funding: *No funding was received by the authors.*

Conflict of Interest: *The authors declare that they do not have any conflict of interest.*

Change of Authorship: *The author has read, understood, and complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors and is aware that with minor exceptions, no changes can be made to authorship once the paper is submitted.*

References

Bae I, Ji U, (2019) Outlier detection and smoothing process for water level data measured by ultrasonic sensor in stream flows. Water 11, 1-16. doi: 10.3390/w11050951

Bashir A, Awawdeh M, Faisal T, Flower Queen MP, (2022) Matlab-based Graphical User Interface for IoT Sensor Measurements Subject to Outlier. In: Proceeding of the Advances in Science and Engineering Technology International Conferences (ASET), pp. 1-6, Dubai, United Arab Emirates. https://www.scribd.com/document/716118362/C-2022

Ben-Gal I, (2005) Outlier detection in data mining and knowledge discovery handbook: A complete guide for practitioners and researchers (Ed. by O. Maimon and L. Rockach) pp. 131–146. https://is.muni.cz/el/fi/jaro2013/PV056/um/outlier-HandbookOfKD

Berendrecht W, van Vliet M, Griffioen J, (2023) Combining statistical methods for detecting potential outliers in groundwater quality time series. Environ. Modell. Asses. 195, 1-14. doi: 10.1007/s10661-022-10661-0

Bonakdari H, Zeynoddin M, (2022) Stochastic Modeling: A Thorough Guide to Evaluate, *Pre-Process, Model and compare time series with MATLAB software* (Ed. by Aera Gariguez) pp. 1-353. https://www.scribd.com/document/722552043/Full-download-book-Stochastic-Modeling-A-

Thorough-Guide-To-Evaluate-Pre-Process-Model-And-Compare-Time-Series-With-Matlab-Software-Pdf-pdf

Chelishchev P, Popov A, Sorby K, (2018) An investigation of outlier detection procedures for CMM measurement data. Matec Web of Conferences 220, 1-6. doi: 10.1051/matecconf/201822004002

Cho H., Oh JH, Kim KO, Shim JS, (2013) Outlier detection and missing data filling methods for coastal water temperature data. Journal of Coastal Research 165,1898–1903. doi: 10.2112/SI65-321.1

Di Blasi JIP, JM, Torres, Nieto PJG, Fernández JRA, Muñiz CD, Taboada J., (2015) Analysis and detection of functional outliers in water quality parameters from different automated monitoring stations in the Nalón River Basin (Northern Spain). Environmental Science and Pollution Research 22, 387–396. doi: 10.1007/s11356-014-3318-5

Dogo EM, Nwulu NI, Twala B, Aigbavboa C, (2019) A survey of machine learning methods applied to anomaly detection on drinking-water quality data. Urban Water Journal 16, 235-248. doi: 10.1080/1573062X.2019.1637002

Friel M, Enderle J, DeFrancesco V, Rosow E, Greenshields I, (2004) Design, implementation and validation of a graphical user interface for analysis of patient monitor event logs. In: Proceeding of the IEEE 30th Annual Northeast Bioengineering Conference, pp. 182-183, Springfield, MA, USA. doi: 10.1109/NEBC.2004.1300055

Grubbs FE, (1969) Procedures for detecting outlying observations in samples, Technometrics, 11, 1–21. doi: 10.1080/00401706.1969.10490657

Jamshidi EJ, Yusup Y, Kayode JS, Kamaruddin MA, (2022) Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature. Ecological Informatics 69, 1-12. doi: 10.1016/j.ecoinf.2022.101672

Jingang J, Lu S, Zhongya F, Jiaguo Q, (2017) Outlier detection and sequence reconstruction in continuous timeseries of ocean observation data based on difference analysis and the Dixon criterion. Limnology and Oceanography: Methods 15, 916-927. doi: 10.1002/lom3.10212

Lai KH, Zha D, Wang G, Xu J, Zhao Y, Kumar D, Chen Y, Zumkhawaka P, Wan M, Martinez D, Hu X, (2021) TODS: An Automated Time Series Outlier Detection System. In: Proceedings of the AAAI Conference on Artificial Intelligence, 35(18), pp. 16060-16062. doi: 10.1609/aaai.v35i18.18012

Muñiz CD, Nieto PJG, Fernández JRA, Torres JM, Taboada J, (2012) Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain). Science of the Total Environment 439, 54-61. doi: 10.1016/j.scitotenv.2012.08.083

Plazas-Nossa L, Angulo MAA, Torres A, (2016) Detection of outliers and imputing of missing values for water quality UV-VIS Absorbance Time Series. INGENIERÍA Journal: Computational Intelligence Section 22, 111-124. doi: 10.14483/UDISTRITAL.JOUR.REVING.2017.1.A01

Saberi A, (2015) Automatic outlier detection in automated water quality measurement stations. M.Sc. Thesis, University Laval, 1-112. https://modeleau.fsg.ulaval.ca/fileadmin/modeleau/documents/Publications/MSc_s/saberiatefeh_msc.pdf

Sun G, Jiang P, Xu H, Yu S,1 Guo D, Lin G, Wu H, (2019) Outlier detection and correction for monitoring data of water quality based on improved VMD and LSSVM. Complexity 2019, 9643921, 1-12. doi: 10.1155/2019/9643921

Vera MJ, Dubravka B, Nikola J, Vojin I, Bojana PB, (2013) Detecting and removing outlier(s) in electromyographic gait-related patterns. J. Appl. Stat. 40, 1319-1332. doi: 10.1080/02664763.2013.785495