Name: Rohit Kumar

Fourth Year Undergraduate, Mechanical Engineering, IIT Goa

E-mail: **rohit.kumar.21063@iitgoa.ac.in**

LinkedIn: **rohit-kumar**

Duration: 26 May - 30 June (5 weeks)

Mentor: Prof. Hanumant Singh Shekhawat

Water Quality Monitoring System

Introduction

Water quality is essential for the health of ecosystems, human consumption, and industrial processes. Monitoring water quality involves evaluating the physical, chemical, and biological characteristics of water to ensure it meets established standards. This is crucial for safeguarding public health, protecting natural habitats, and ensuring the sustainability of water resources.

A Water Quality Monitoring System (WQMS) is a comprehensive framework that includes the collection, analysis, and interpretation of data related to water quality. These systems are implemented using various technologies and methodologies to detect and respond to changes in water conditions promptly.

Key Components

1. Sensors and Instruments: These are deployed in water bodies to measure parameters such as temperature, pH, dissolved oxygen, conductivity, and the presence of specific contaminants like heavy metals or pathogens.
2. Data Acquisition: The data collected by sensors is transmitted to a central system. This can be achieved through wired or wireless communication networks.
3. Data Management: Collected data is stored, processed, and analyzed. Advanced data management systems can handle large datasets, ensuring data integrity and availability.
4. Data Analysis and Interpretation: Using statistical and computational methods, the data is analyzed to identify trends, anomalies, and potential issues. Machine learning algorithms and predictive models can enhance the accuracy and reliability of the analysis.
5. Reporting and Visualization: Results are presented in user-friendly formats such as graphs, dashboards, and reports. This aids in decision-making by providing clear insights into water quality conditions.
6. Alert Systems: Automated alerts and notifications are generated when water quality parameters exceed predefined thresholds, enabling timely interventions to mitigate potential risks.

Benefits

- Environmental Protection: Continuous monitoring helps detect pollution sources, prevent ecological damage, and maintain biodiversity in aquatic ecosystems.
- Public Health Safety: Ensures that drinking water is safe, reducing the risk of waterborne diseases.

- Regulatory Compliance: Helps organizations comply with environmental regulations and standards, avoiding legal penalties.
- Resource Management: Facilitates efficient water resource management, ensuring sustainable use and conservation.
- Industrial Applications: Ensures the quality of water used in industrial processes, which can affect product quality and equipment longevity.

Applications

- ➢ Drinking Water Supply: Ensuring that water distributed to households meets safety standards.
- ➢ Wastewater Treatment Plants: Monitoring effluent quality to ensure compliance with discharge regulations.
- ➢ Natural Water Bodies: Tracking the health of rivers, lakes, and coastal areas to protect wildlife and recreational activities.
- ➢ Agriculture: Monitoring irrigation water to prevent soil salinity and crop contamination.
- ➢ Industrial Effluents: Controlling the quality of water discharged from industrial activities to prevent environmental contamination.

Apparatus Used:

pH-W218

Description →

- It can easily measure PH/EC/SALT/ORP/TDS/S. G/TEMP/CF in the water
- WIFI connection and data transfer, realize real time monitoring on the phone
- A mobile app connects 20 devices at the same time, convenient to use Simply immerse the electrode into water, and get the data on the LCD screen
- High precision replaceable electrode, wide range of application

Temperature - Current Temperature (℃/°F) The best temperature range for water to be absorbed and rehydrate effectively is between 10-22 °C (50 - 72 °F) <wisewell.ae>, <www.bisleri.com>

pH - pH value (pH) The pH of most drinking-water lies within the range 6.5–8.5. Natural waters can be of lower pH, as a result of, for example, acid rain or higher pH in limestone areas.

ORP - ORP value (mV) Safe drinking water, for example, is an oxidizing agent with an ORP range between +200 and +600 mV. Chlorinated pool water, on the other hand, should have a much higher ORP between +650 and +750 mV.

EC - EC value(μs/cm) Electrical conductivity is a measure of the “total salts” concentration in the nutrient solution (drip, slab or drain). It is expressed in milliSiemens per linear centimeter (mS/cm) or microSiemens per linear centimeter (mS/cm) where 1mS = 1000μS.

Types of water	**Conductivity Value**
Pure distilled and Deionized water	0.05 μS/cm
Seawater	50 mS/cm
Drinking water	200 to 800 μS/cm.
Rain or Snow water	2 to 100 μS/cm

(conductivity-of-water)

TDS - TDS value (ppm) <Total dissolved solids> According to the BIS, the permissible limit for TDS in drinking water is 500 mg/L. However, the World Health Organization recommends a TDS level of less than 300 mg/L for drinking water. Minimum TDS of drinking water should not go below 50 ppm.

TDS Level (mg/L)	Water Quality	Health Implications
50-300	Excellent	Safe for drinking
300-600	Good	Safe for drinking
600-900	Fair	Safe for drinking
900-1200	Poor	May cause laxative effect in some people
1200-2000	Very Poor	Not suitable for drinking on a regular basis
Above 2000	Unacceptable	Not suitable for drinking

Drinking Water TDS Chart (www.livpuresmart.com)

Salts - Salinity value (ppm)

Table 1: Water Salinity

Salinity Status	Salinity (%)	Salinity (ppt)	Use
Fresh	< 0.05	< 0.5	Drinking and all irrigation
Marginal	0.05 – 0.1	0.5 – 1	Most irrigation, adverse effects on ecosystems become apparent
Brackish	0.1 – 0.2	1 – 2	Irrigation certain crops only; useful for most stock
Saline	0.2 – 1.0	2 – 10	Useful for most livestock
Highly Saline	1.0 – 3.5	10 – 35	Very saline groundwater, limited use for certain livestock
Brine	> 3.5	> 35	Seawater; some mining and industrial uses exist

Source: Department of Water. Government of Western Australia.

(www.horiba.com/ind/water-quality)

SG - Proportion value (S.G) <specific gravity> the specific gravity of water at about 4°C is 1.000.

CF - CF(CF) <Conductivity Factor> The conductivity factor of dissolved salts in a given solution is a measurement of conductivity. Using the electrical conductivity between two electrodes in a water solution, the level of dissolved solids in that solution can be measured. (en.wikipedia.org/wiki/Conductivity_factor)

Work

Collected values of all 8 parameters through exporting xlsx file from Tuya Smart App:

a. Fresh Drinking Water
b. Fresh Drinking Water
c. Normal Tap Water
d. Normal Tap Water
e. Normal Lab Water after adding pH 6.86 powder
f. Normal Lab Water after adding pH 4 powder

g. Normal Lab Water after adding pH 9.18 powder

Data file -

https://docs.google.com/spreadsheets/d/1D85inHHH68qjMXMPlr5uXc2Dmc4UB307/edit?usp=sharing&ouid=112888976496580693154&rtpof=true&sd=true

Code Explanation:

Loading and Preprocessing: Load data from an Excel file, handle any missing or non-numeric values, and select relevant columns.

Standardization: Standardize the data using Standard Scaler to ensure all features have the same scale. Normalizing the data to have a mean of 0 and standard deviation of 1. After scaling the data using StandardScaler, the feature names are set back to the scaled data using pd.DataFrame(data_scaled, columns=data.columns).

Elbow Method: Plot the inertia (sum of squared distances) for different numbers of clusters to determine the optimal number of clusters.

Silhouette Method: Plot the silhouette scores for different numbers of clusters to further validate the optimal number of clusters.

K-means Clustering: Apply K-means clustering with the optimal number of clusters determined from the methods above.

Visualization: Use PCA for dimensionality reduction to 2D and visualize each data point colored by its assigned cluster. Plot centroids of each cluster as well. After performing PCA for dimensionality reduction, the feature names are set back to the reduced data (data_reduced) using pd.DataFrame(data_reduced, columns=['PC1', 'PC2']).

Centroids: Print and display the centroids of each cluster in the original feature space. The centroids are transformed back to the original scale using scaler.inverse_transform(kmeans.cluster_centers_) for interpretability.

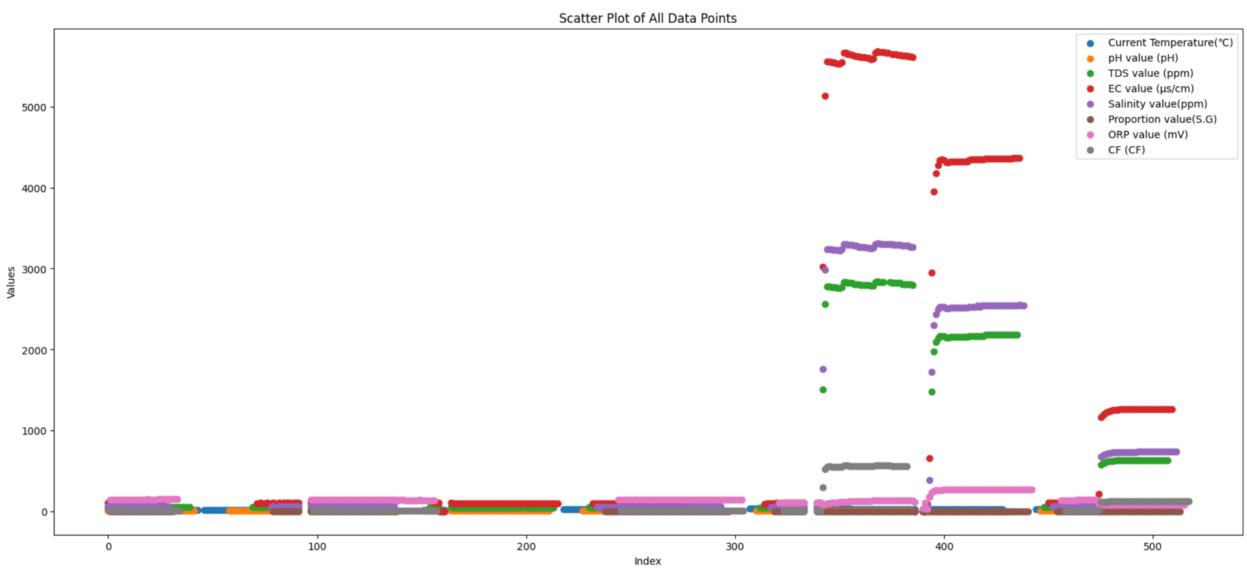
Notes:

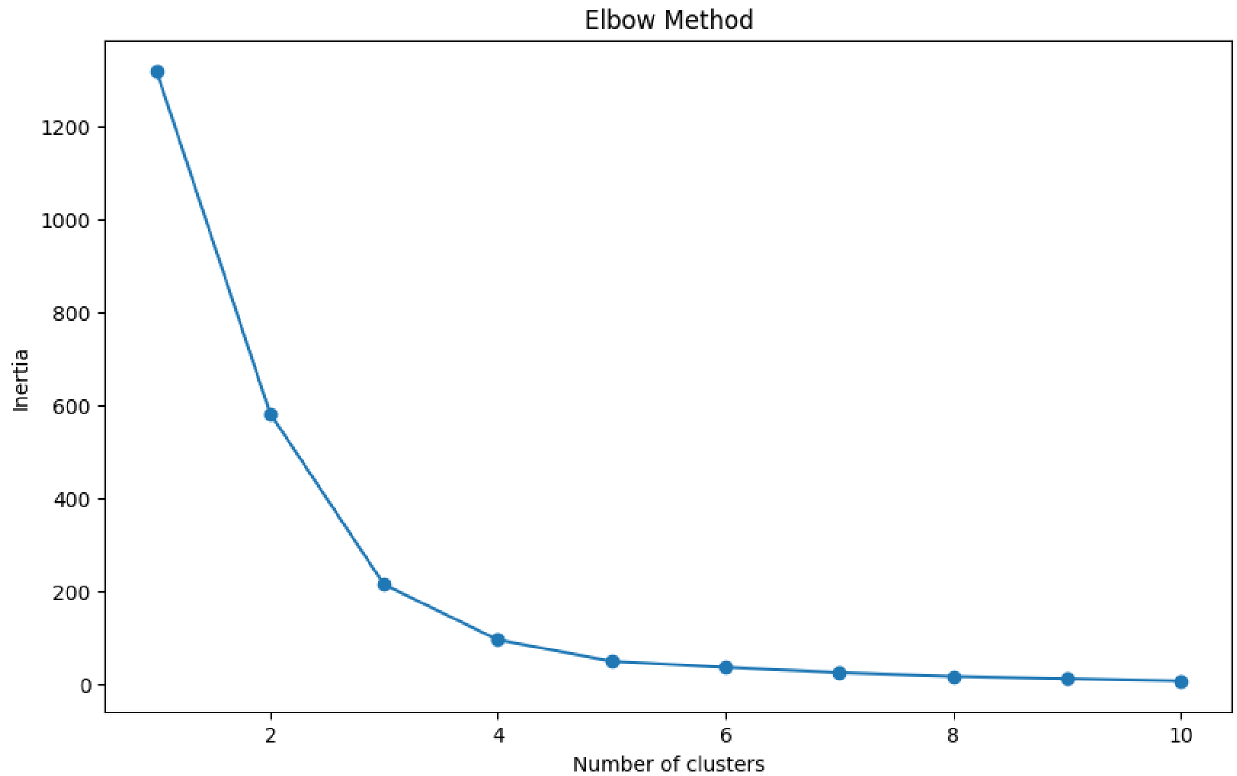
- Ensure 'your_file.xlsx' is replaced with the actual path to your Excel file.
- Adjust n_clusters based on the optimal number determined from the Elbow or Silhouette method plots.
- PCA is used here for visualization purposes; adjust as needed for higher-dimensional data.
- Centroids are transformed back to the original scale using scaler.inverse_transform() for interpretability.

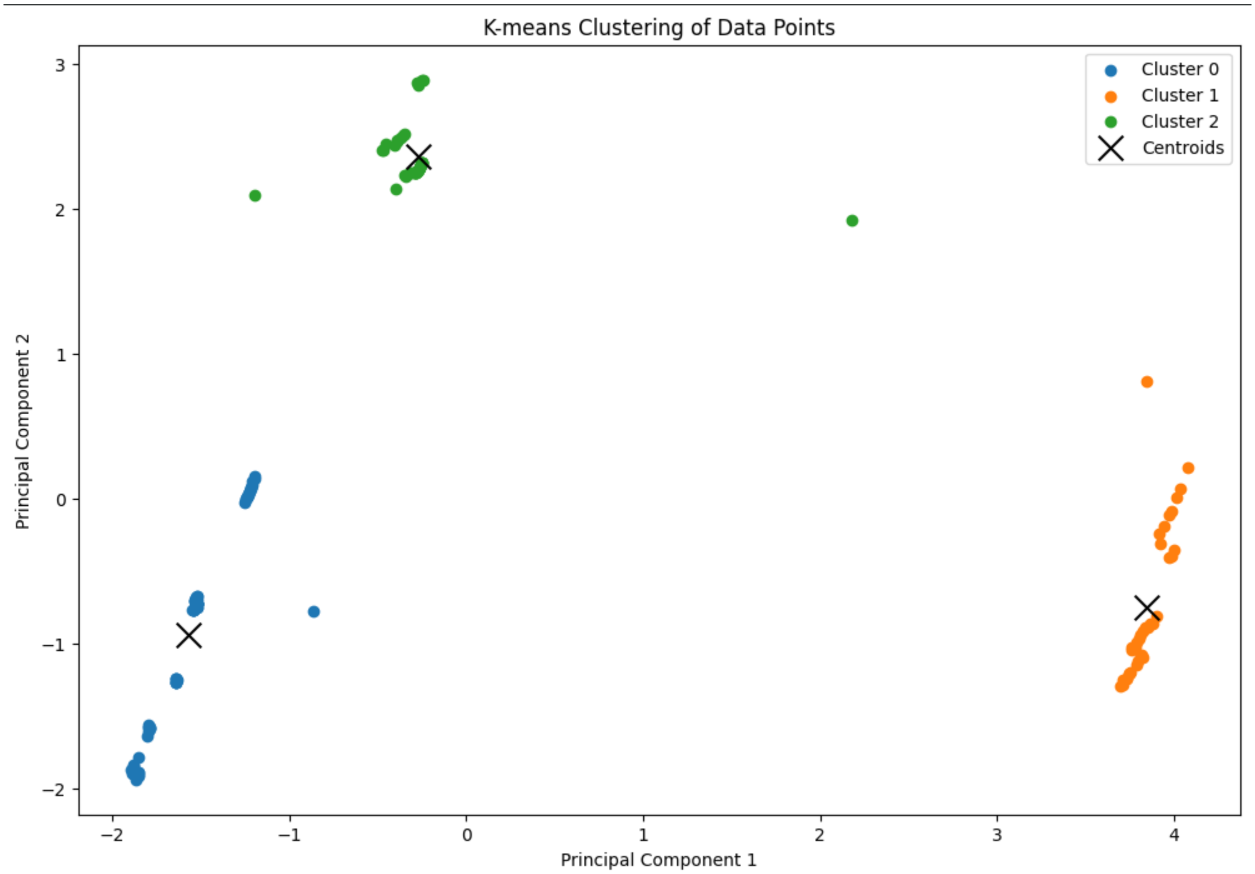
Code Summary:

Load Data: Read the data from an Excel file.
Preprocess Data: Handle missing values and convert necessary columns to numeric.
Standardize Data: Normalize the data to have a mean of 0 and standard deviation of 1.
Determine Optimal Clusters: Use the Elbow and Silhouette Methods to find the optimal number of clusters.
Apply K-means: Perform K-means clustering with the optimal number of clusters.
Visualize Clusters: Use PCA for dimensionality reduction and visualize the clusters.

Code - https://colab.research.google.com/drive/1erbYYJE07cYVroKDOOXY2fJe5cDsawUQ#scrollTo=d81cab1d-b5bf-491f-8536-8387e3401002




Elbow Method
Inertia
1200
1000
800
600
400
200
0
2
4
6
8
10
Number of clusters


K-means Clustering of Data Points
Cluster 0
Cluster 1
Cluster 2
Centroids
Principal Component 2
3
2
1
0
−1
−2
−2
−1
0
1
2
3
4
Principal Component 1

Google Drive:
https://drive.google.com/drive/folders/1sj3jEUk8HoelCwrdG9mpSNwWmGAgawXJ?usp=sharing

Interpretation

The centroids obtained in a K-means clustering analysis represent the center points of each cluster in the multidimensional feature space. When analyzing water quality data, each centroid can be interpreted as a representative or "average" point for all the water samples that belong to that particular cluster. The significance of these centroids in the context of water quality analysis can be understood through several key aspects:

Understanding Centroids in Water Quality Clustering

1. Representative Water Quality Profiles:
    a. Centroid Values: Each centroid's coordinates correspond to the average values of the water quality parameters (e.g., temperature, pH, TDS, etc.) for that cluster. These average values can give insights into the typical water quality profile of samples within that cluster.
    b. Cluster Characteristics: By examining the values of each parameter at the centroid, you can infer the general characteristics of the water quality in that cluster. For example, a cluster with a centroid indicating high salinity and TDS might represent samples from a brackish or polluted water source.
2. Water Quality Categorization:
    a. Quality Segmentation: The centroids can help categorize the water samples into different quality segments. For instance, one cluster might represent potable water with acceptable pH and low contaminant levels, while another cluster might represent non-potable water with high contaminant levels.
    b. Actionable Insights: Based on the centroids, water management authorities can prioritize actions. For example, clusters with centroids indicating poor water quality can be targeted for remedial measures or further investigation.
3. Comparative Analysis:
    a. Inter-cluster Comparison: By comparing the centroids of different clusters, one can identify how different clusters (and therefore different water quality profiles) vary from each other. This can be useful for understanding the diversity of water quality in the dataset.
    b. Trend Analysis: Changes in the centroids over time (if analyzing temporal data) can help identify trends or shifts in water quality, indicating improving or worsening conditions.
4. Identification of Anomalies:
    a. Outlier Detection: Clusters with centroids far from the rest might indicate outliers or anomalies in the dataset. These could be due to unusual contamination events or rare water sources that require special attention.

Water quality parameters for Interpretation: Current Temperature, pH value, TDS value, EC value, Salinity value, Proportion value, ORP value, and CF value. After performing K-means clustering, three centroids corresponding to three clusters has been obtained.

Centroid 1:

- Temperature: 24.28 °C, pH: 7.06, TDS: 54.33 ppm, EC: 108.65 µS/cm, Salinity: 62.93 ppm, Proportion: 0.997 S.G, ORP: 146.24 mV, CF: 10.87 CF

  Interpretation:

  - Temperature: Slightly below room temperature, indicating a relatively stable thermal environment.
  - pH: Near neutral (7.06), which is ideal for most drinking water standards.
  - TDS and EC: Both values are low (54.33 ppm TDS and 108.65 µS/cm EC), suggesting minimal dissolved solids and low conductivity, which generally indicate good water quality.
  - Salinity: Very low salinity (62.93 ppm), suggesting minimal salt content.
  - Proportion (Specific Gravity): Close to pure water (0.997 S.G), indicating no significant impurities affecting the density.
  - ORP: Moderate oxidation-reduction potential (146.24 mV), indicating a balanced redox state.
  - CF: Low contamination factor (10.87 CF), suggesting low levels of contaminants.

Overall, Centroid 1 represents high-quality, potable water with minimal impurities and balanced pH.

Centroid 2:

- Temperature: 33.29 °C, pH: 6.53, TDS: 2798.38 ppm, EC: 5602.70 µS/cm, Salinity: 3265.41 ppm, Proportion: 0.997 S.G, ORP: 126.05 mV, CF: 561.04 CF

  Interpretation:

  - Temperature: High (33.29 °C), indicating a warmer environment, which could be due to industrial discharge or hot climate.
  - pH: Slightly acidic (6.53), which may suggest influence from natural organic acids or pollutants.
  - TDS and EC: Very high (2798.38 ppm TDS and 5602.70 µS/cm EC), indicating a significant presence of dissolved solids and high conductivity, often associated with pollution or industrial runoff.
  - Salinity: Very high (3265.41 ppm), suggesting substantial salt content, which may be from seawater intrusion or industrial effluents.
  - Proportion (Specific Gravity): Close to pure water (0.997 S.G), which is normal.
  - ORP: Lower oxidation-reduction potential (126.05 mV), indicating a more reducing environment, possibly due to organic matter or pollutants.
  - CF: Very high contamination factor (561.04 CF), suggesting high levels of contaminants.

Overall, Centroid 2 represents poor water quality with high levels of dissolved solids, salinity, and contaminants, making it unsuitable for drinking without significant treatment.

Centroid 3:

- Temperature: 30.88 °C pH: 8.35 TDS: 456.38 ppm EC: 912.76 µS/cm Salinity: 532.09 ppm Proportion: 0.995 S.G ORP: 92.6 mV CF: 91.28 CF
  Interpretation:
    - Temperature: Relatively high (30.88 °C), indicating a warmer environment.
    - pH: Moderately alkaline (8.35), which might be due to alkaline minerals or contamination.
    - TDS and EC: Moderate levels (456.38 ppm TDS and 912.76 µS/cm EC), suggesting a noticeable presence of dissolved solids but not excessively high. Salinity: Moderate (532.09 ppm), indicating some salt content.
    - Proportion (Specific Gravity): Slightly lower than pure water (0.995 S.G), indicating some impurities affecting the density.
    - ORP: Low oxidation-reduction potential (92.6 mV), suggesting a more reducing environment.
    - CF: Moderate contamination factor (91.28 CF), indicating noticeable contamination.

Overall, Centroid 3 represents water of moderate quality. It has higher temperature, pH, and contamination compared to Centroid 1, but significantly better quality compared to Centroid 2. This water might be suitable for some uses but would need treatment for drinking.

Conclusion

In summary, Water Quality Monitoring Systems are vital tools in the quest to maintain and improve water quality. By leveraging advanced technologies and robust methodologies, these systems provide critical insights that drive effective water management and protection strategies.

The centroids from K-means clustering provide a powerful way to summarize and interpret the complex multidimensional data associated with water quality. By analyzing these centroids, water quality managers and researchers can gain insights into typical water conditions, identify areas of concern, and develop targeted strategies to improve water quality and ensure safe drinking water.

The centroids provide a comprehensive view of the different water quality profiles present in the dataset. Each centroid's parameters give insights into the general water quality within that cluster, highlighting areas of good quality and potential contamination. This information is crucial for water quality management, helping to identify clusters needing attention and informing decisions for treatment and remediation.

References

- https://www.tvcmall.com/details/ph-w218-8-in-1-ph-ec-salt-orp-tds-s-g-temp-cf-monitor-wifi-water-quality-tester-eu-plug-skuc0015246b.html
- https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf
- https://www.krinstruments.com/water-quality-monitoring-system/
- https://www.researchgate.net/profile/Prateek-Goyal-2/post/What-are-GMP-for-drinking-water-and-its-bottles/attachment/59d61fa179197b807797e136/AS%3A285871606648832%401445168510385/download/spec.pdf
- https://www.intechopen.com/chapters/69568

Thank You