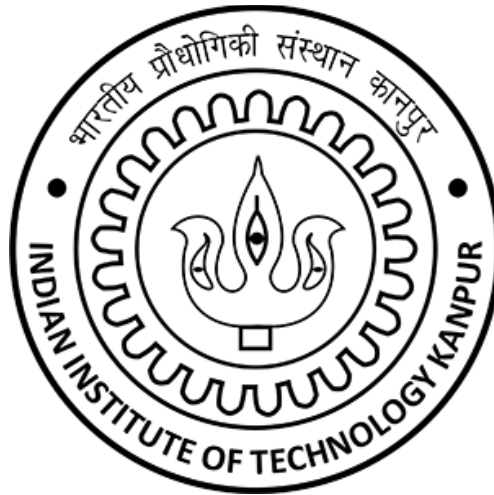


# INDIAN INSTITUTE OF TECHNOLOGY, KANPUR



Semester 23-24 II  
SURGE Project Report

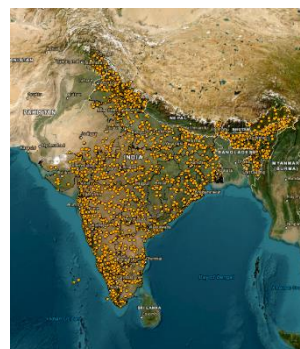
## **“Geospatial Visualization of Water Samples and Water Quality Indexes”**

*Submitted by*  
**Shivani Gupta**  
Application No.: 2430521  
Indian Institute of Technology  
Kanpur

*Under the Guidance of*  
**Prof. Shankar Prawesh**  
Indian Institute of Technology Kanpur

## A. INTRODUCTION

Assessing water quality is crucial for environmental management, especially in a large and diverse country like India. This project involves a thorough analysis of water quality across India, using a substantial dataset of over 61,677 entries from more than 30,000 water bodies and 84 physio-chemical parameters. Our objective is to analyze mineral concentrations from various monitoring stations across India and to preprocess and clean the data to make it usable for further research. By addressing data quality issues and geographical gaps, we aim to provide reliable insights into the status of mineral concentrations in Indian water bodies.



## B. DATA DESCRIPTION

The dataset includes 61,667 data points covering key minerals such as calcium, chloride, electrical conductivity, fluoride, bicarbonates, potassium, magnesium, sodium, and pH. Out of 84 parameters, only 18 have more than 45% data availability, with 9 key parameters having over 85% data coverage.

## C. PROBLEMS IDENTIFIED

### 1. Data Sparsity

Many minerals (e.g., silver, aluminum, aldrin, barium, etc.) have extremely sparse data, limiting comprehensive analysis.

### 2. Data Consistency

Discrepancies in longitude and latitude values due to new monitoring sites and spelling errors.

Lack of uniformity in data recording and entry.

### 3. Incomplete Dataset

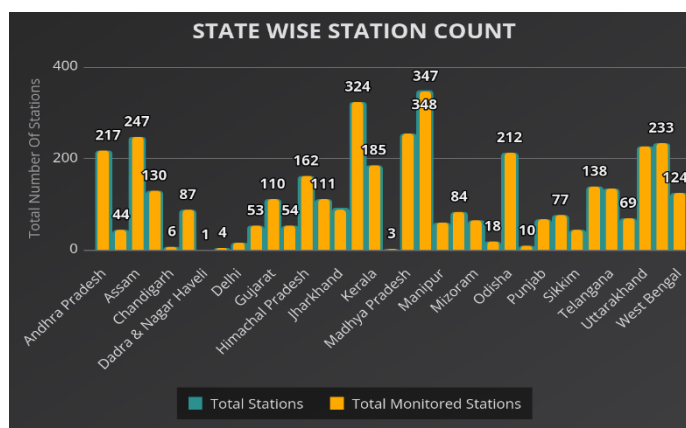
Missing values for important minerals on the India-WRIS site, affecting dataset comprehensiveness.

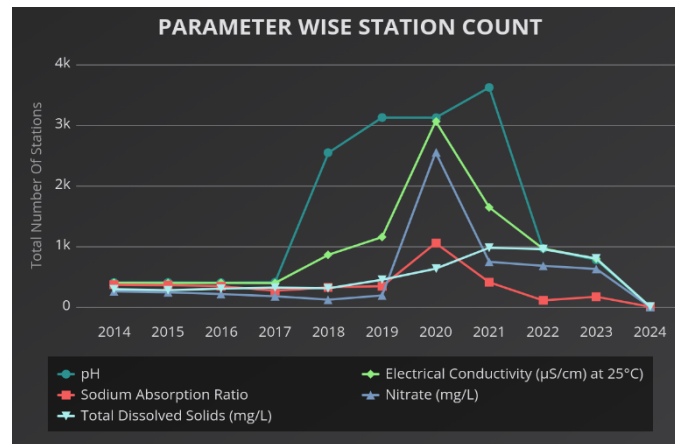
### 4. Manual Corrections Required

Errors identified through mismatches or error messages during the data integration process.

### 5. Geographic Gaps

Absence of monitoring sites in various regions of India, leading to incomplete geographic coverage and potential data gaps.





## D. SOLUTIONS AND METHODOLOGIES

### 1. Data Imputation

- Use multiple imputation techniques to handle uncertainties in imputed values.
- Employ statistical methods or machine learning techniques to impute missing values.

### 2. Data Standardization

- Standardize names and locations of monitoring sites to ensure consistency.
- Use geocoding and fuzzy matching techniques to correct spelling errors and align old and new site names.

### 3. Data Augmentation

- Supplement the dataset with additional data from reliable sources to fill in gaps.
- Use external databases and previous research to cross-validate and augment existing data.

### 4. Geospatial Data Integration

- Utilize pre-existing datasets containing latitude, longitude, and WRIS IDs up to the year 2018.
- Apply the Index-Match function in Microsoft Excel to cross-reference and retrieve corresponding geospatial data.
- Incorporate matched latitude, longitude, and WRIS IDs into the new dataset.

### 5. Manual Corrections

- Identify errors through mismatches during the Index-Match process.
- Manually locate and integrate correct WRIS IDs with the corresponding latitude and longitude.

### 6. Expanding Monitoring Coverage

- Advocate for the establishment of new monitoring sites in underrepresented regions.

- Use remote sensing and satellite data to estimate water quality in areas lacking direct monitoring.

## **E. CONCLUSION**

Preprocessing and cleaning the dataset is crucial for accurate and reliable analysis. Addressing data sparsity, consistency, completeness, and geographic gaps significantly enhances the quality of research. Implementing these solutions enables more robust and comprehensive studies of mineral concentrations across India.

## **F. ACKNOWLEDGMENTS**

Thanks to Indian Institute of Technology Kanpur, SURGE and Professor Shankar Pravesh for their support in this research.

## **G. REFERENCE**

- India-WRIS
- GEMStat - The global water quality database