# WATER TO CLOUD

## REVOLUTIONIZING WATER QUALITY MONITORING OF RIVERS

**Learnings from Godavari**

**University of Chicago**

PRITZKER SCHOOL OF MOLECULAR ENGINEERING
THE UNIVERSITY OF CHICAGO

Water-To-Cloud

THE UNIVERSITY OF CHICAGO

in partnership with

asci
Leadership through Learning

INTERNATIONAL INNOVATION CORPS

This research was supported by a subward from the Administrative Staff College of India via grant funding from the Bill & Melinda Gates Foundation [Grant Agreement OPP1163021].

Disclaimer

This document has been prepared by the Water-to-Cloud (hereinafter referred to as "W2C") team at the University of Chicago. This document shares the experience of this team in real time water quality monitoring solutions and data collection.

The status presented in this report is based on information collated through interactions and regular follow-up with stakeholders, government-recognized laboratories and sensor-based data collected by W2C team. Due care has been taken to validate the authenticity and correctness of the information, however, no representations or warranty, expressed or implied, is given by W2C or any of its respective partners, officers, employees or agents as to the accuracy or completeness of the information, data or opinions provided by various parties.

## Contributing Authors

**Professor Supratik Guha**
Professor at the Pritzker School of Molecular Engineering, The University of Chicago

**Priyank Hirani**
Program Director

**Himank Sharma**
Project Lead

**Srinivas Balivada, PhD**
Water Research Lead

**Ritambhara Chauhan**
Technical Lead

**Gopi Shaik**
Field Assistant

**Lova Murthy**
Field Assistant

**William Kent**
Web Developer

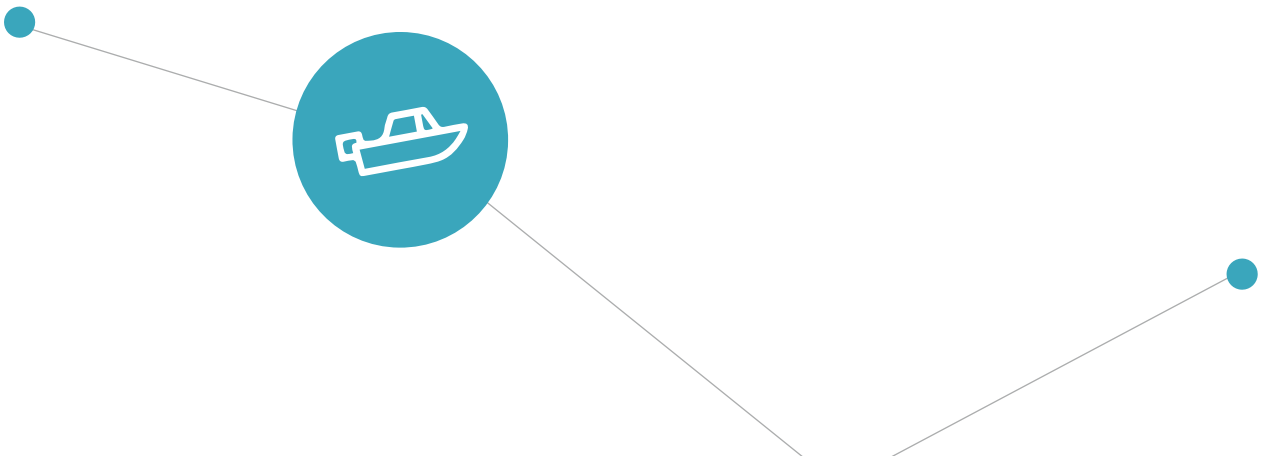
**Ashu Balhara**
Research Associate

**Anand Kumar, PhD**
Water Researcher

**Utkarsh Rai**
Web Developer

**Ziqi Ma**
Data Analyst

**Ramesh ChAndhra Ponduru**
Senior Research Advisor



# Table of Contents


Abbreviations 3
List of Figures 4
List of Tables 8
Acknowledgement 9
Executive Summary 11

Unit 1: Introduction 14
- 1.1 Existing Water Quality Monitoring Scenario 14
- 1.2 The Need for High Resolution Mapping 15

Unit 2: Measurement and Analysis Methodology for the Water-to-Cloud System 16
- 2.1 Overview 16
- 2.2 Data Collection Methodology 17
  - 2.2.1 Site and Route Selection 17
  - 2.2.2 Parameter Selection 17
  - 2.2.3 Sensor Identification and Calibration 19
  - 2.2.4 Monitoring Frequency 21
  - 2.2.5 Operationalization 22
- 2.3 Data Storage and Curation 23
  - 2.3.1 Data Retrieval 23
  - 2.3.2 Data Processing 24
  - 2.3.3 Digital Access 24
- 2.4 Analysis 25
  - 2.4.1 Visualization 25
  - 2.4.2 Laboratory Analysis 27
  - 2.4.3 Statistical Analysis 28
  - 2.4.4 Machine Learning 28

Unit 3: Godavari Monitoring Study 29
- 3.1 Project Context 29
- 3.2 Study Area 29
  - 3.2.1 Narsapur drainage system 33
  - 3.2.2 Palacole drainage system 36
  - 3.2.3 Kovvur drainage system 37
- 3.3 Work Carried Out 37
  - 3.3.1 Phase I: Scoping Study (October 2016 to March 2017) 37
  - 3.3.2 Phase II: Piloting Water-to-Cloud System (April 2017 to September 2017) 38
  - 3.3.3 Phase III: Data Curation and On-ground Implementation (October 2017 to March 2018) 39
  - 3.3.4 Phase IV: Data Analysis (April 2018 to September 2018) 40
  - 3.3.5 Phase V: Scaling Up, 24x7 Monitoring (October 2018 to March 2019) 41
  - 3.3.6 Phase VI: Wrapping up (April 2019 to September 2019) 42

Unit 4: Research Findings 43
4.1 Pollution Hotspots Identification 43
4.1.1 Major Drains in a Town 43
4.1.2 Industrial Discharge 45
4.1.3 Leaching Effects 46
4.2 Direct and Indirect Usage of Water 47
4.2.1 Suitable for Bathing 47
4.2.2 Unsafe for Drinking Sometimes 50
4.2.3 Toxicants Entering Food Chain 50
4.3 Evidence based WASH Interventions 51
4.3.1 Locating Insanitary Toilets & Slum Zones 51
4.3.2 Case for Treatment Plants 54
4.3.3 Regular Cleaning of Drains 62
4.4 Other Findings 67
4.4.1 Nature's Attempt at Revival 67
4.4.2 Fish Kills-Inverse Relation between EC & DO 69
4.4.3 Water Quality Prediction Models 71
Unit 5: Challenges 80
5.1 Sensor Technology 80
5.2 Calibration and Maintenance 80
5.3 Trained Technical Operators 80
5.4 Logistics 81
5.5 Ground Situation Analysis 81
5.6 Acceptable Monitoring Protocol 81
Unit 6: The Way Forward 83
Unit 7: Conclusion 84
APPENDIX I: Dissemination of Knowledge and Learnings 85
APPENDIX II: Using Cyber Physical Systems to Map Water Quality over Large Water Bodies 87
APPENDIX III: Team Profile 93
APPENDIX IV: Use based Classification of Surface Waters in India as per CPCB 96
APPENDIX V: Water Quality Parameters 97
APPENDIX VI: Calibration of Sensor Probes 99
APPENDIX VII: Statistical Tools 105
APPENDIX VIII: Heavy Metal Analysis 107

# Abbreviations

**ABS**-Absorbance
**ANN** -Artificial Neural Network
**ASCI**-Administrative Staff College of India
**BIS**-Bureau of Indian Standards
**BMGF**-Bill & Melinda Gates Foundation
**BOD**-Bio-chemical Oxygen Demand
**CDOM**-Colored Dissolved Organic Matter
**COD**-Chemical Oxygen Demand
**CPCB**-Central Pollution Control Board
**CSS**-Cascading Style Sheets
**CSV**-Comma Separated Values
**CWC**-Central Water Commission
**DO**-Dissolved Oxygen
**EC**-Electrical Conductivity
**ETP**-Effluent Treatment Plant
**FNU**-Formazin Nephelometric Unit
**FPS**-Feet Per Second
**GoAP**-Government of Andhra Pradesh
**GPS**-Global Positioning System
**GSM**-Global System for Mobile Communications
**ICPMS**-Inductively Coupled Plasma Mass Spectrometry
**MA**-Municipal Administration
**MLD**-Million Liters Per Day
**NaN**-Not-a-Number
**ODF**-Open Defecation Free
**ORP**-Oxygen Reduction Potential
**PME**-Pritzker School of Molecular Engineering
**RC**-Remote-controlled
**S3**-Sampoorna Swachhta Sankalp
**STP**-Sewage Treatment Plant
**SUVA**-Specific Ultraviolet Absorbance
**SVM**-Support Vector Machine
**TOC**-Total Organic Carbon
**TDS**-Total Dissolved Solid
**UDD**-Urban Development Department
**ULB**-Urban Local Bodies
**UVT**-Ultraviolet Transmission
**W2C**-Water-to-Cloud
**WQI**-Water Quality Index

# List of Figures

**Figure 1:** Illustration of the Water-to-Cloud system comprising a boat mounted with mobile sensors that record water quality data in real time, which is then hosted on a cloud server and disseminated through an open access digital platform in the form of high resolution heat maps 16

**Figure 2:** Pictures of different types of mobile and stationary real-time sensors used in the Water-to-Cloud system 19

**Figure 3:** Picture showing deployment of a mobile sensor mounted on a boat using cable ties, coir rope and a cane 22

**Figure 4:** Picture showing deployment of mobile sensor (in this case, a flowmeter) in a waste water drain by dipping it at a depth of 6-12 inches to measure water velocity 22

**Figure 5:** Picture showing deployment of solar-powered stationary sensors (in this case, Dissolved Oxygen and Turbidity sensors) in a waste water drain with the control panel and data logger enclosed in a steel box for protection 23

**Figure 6:** Diagram showing dataflow in the Water-to-Cloud system through the collection, processing and display stages 23

**Figure 7:** Screenshot depicting traditional tabular form of presenting data which is difficult to read on various water quality parameters at different CPCB monitoring stations on River Godavari 25

**Figure 8:** Screenshot of a 2-dimensional, interactive representation of water quality data collected through the Water-to-Cloud system hosted on the open access thoreau platform with the option to choose water body, parameter and the type of scale given to the user 26

**Figure 9:** Screenshot of an interactive heatmap showcasing water quality at various drain sampling points in Narspur with the options to select time range, parameters, drain and the type of analysis 27

**Figure 10:** Diagram depicting machine learning modelling to predict lab-measured parameters using sensor measured parameters by establishing correlation between them 28

**Figure 11:** Map of Rajahmundry in the East Godavari District of Andhra Pradesh depicting the boat ride route in River Godavari, waste water carrying drains, paper mill and sewage treatment plant 30

**Figure 12:** Map of Narsapur in the West Godavari District of Andhra Pradesh depicting the boat ride route, slums, dumping site and slaughterhouse 31

**Figure 13:** Map showing three sites - Kovvur, Palacole and Narsapur identified for monitoring of wastewater drains 32

**Figure 14:** Map of Narsapur depicting the drain pattern (in blue) and 13 sample collection points at the end of open drains (D1, D2, D3, D7, D8, D14 and D15), siphoned drains (S1, S2, S3, S4 and S6) and at Sand Ramp, where the waste water from sector 2 meets River Godavari 33

**Figure 15:** Map of Palacole depicting the drain pattern (in blue) and the six sample collection points along the two main drains 36
**Figure 16:** Map of Kovvur depicting the drain pattern (in blue) and the six sample collection points 37
**Figure 17:** Picture showing Water-to-Cloud team conducting a boat ride along the Narsapur stretch of River Godavari 38
**Figure 18:** Picture showing waste water released near a slaughterhouse in Narsapur into a drain that finally meets River Godavari 41
**Figure 19:** Heatmap tracing the boat trajectory along Rajahmundry stretch of River Godavari to depict an increase in electrical conductivity (μS/cm) around the area where Ava drain, Rajahmundry's biggest waste water drain, meets River Godavari 44
**Figure 20:** Heatmap tracing the boat trajectory along Narsapur stretch of River Godavari to depict a steep increase in CDOM (ppb) as the waste water from channel 2 meets River Godavari 44
**Figure 21:** Heatmap tracing the boat trajectory along Rajahmundry stretch of River Godavari shows a sudden increase in pH near the paper mill 45
**Figure 22:** Heatmap capturing the leaching effect next to a solid waste dumping site along Narsapur stretch of River Godavari depicted through a steep increase in electrical conductivity (μS/cm) 46
**Figure 23:** Water quality index plot of Rajahmundry stretch of River Godavari calculated using sensor parameter (EC, DO, pH, turbidity) values measured at four points with distinct water quality charachteristics: (i) meeting point of Ava waste water drain and river (AVA), (ii) middle of the river (GDVR), (iii) Near paper mill (APPM) and (iv) near Water Treatment Plant (WTP), noted on X-axis, between April 19, 2017 and December 8, 2017 48
**Figure 24:** Water quality index plot of Rajahmundry stretch of River Godavari calculated using sensor parameter (EC, DO, pH, turbidity) values measured at four points with distinct water quality charachteristics: (i) meeting point of Ava waste water drain and river (AVA), (ii) middle of the river (GDVR), and (iii) Near paper mill (APPM) noted on X-axis, between January 6, 2018 and April 4, 2019 49
**Figure 25:** Heatmap showing spatial distribution of households with toilets (in green) and without toilets (in red) in Narsapur 52
**Figure 26:** Heatmap showing spatial distribution of slums (in light mustard) and insanitary toilets (in pink) in Narsapur 52
**Figure 27:** Box plots showing logarithm of fecal coliform concentation (CFU/100ml) at the 13 sampling points (detailed in figure 14) on major drains in Narsapur 53
**Figure 28:** Box plots showing logarithm of fecal coliform concentation (CFU/100ml) at the starting, middle and ending of two major drains (MD1 and MD2 as described in figure 15) of Palacole 53
**Figure 29:** Box plots showing logarithm of fecal coliform concentration (CFU/100ml) at the five 54

**Figure 30:** Box plots showing BOD values (mg/l) at the 13 sampling points (D1, D2,...D15, S1, S2... and S6 as described in figure 14) on major drains of Narsapur 55
**Figure 31:** Box plots showing COD values (mg/l) at the 13 sampling points (D1, D2,...D15, S1, S2...and S6 as described in figure 14) on major drains of Narsapur 55
**Figure 32:** Box plots showing BOD (mg/l) values at the starting, middle and ending of two main drains (MD1 and MD2 as described in figure 15) of Palacole 56
**Figure 33:** Box plots showing COD (mg/l) values at the starting, middle and ending of two main drains (MD1 and MD2 as described in figure 15) of Palacole 56
**Figure 34:** Box plots showing BOD (mg/l) values at the five sampling points (D1, D2, D3, D4 and D5 as described in figure 16) on major drains of Kovvur 57
**Figure 35:** Box plots showing COD (mg/l) values at the five sampling points (D1, D2, D3, D4 and D5 as described in figure 16) on major drains of Kovvur 57
**Figure 36:** Bar chart showing BOD/COD ratio for drains in sector 1 (D1, D2, D3, D7 and D8 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur 59
**Figure 37:** Bar chart showing BOD/COD ratio for drains in sector 2 (D14 and D15 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur 60
**Figure 38:** Bar chart showing BOD/COD ratio for drains in sector 3 (S1, S2, S3, S4 and S6 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur 61
**Figure 39:** Timeplot of diurnal variation in dissolved oxygen concentration (mg/L) in Muriki Kaluva waste water canal across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval 62
**Figure 40:** Timeplot of diurnal variation in dissolved oxygen concentration (mg/L) in drain D1 (as described in figure 14) passing through Narsapur Waterworks Department across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval (data not recorded between January 1, 2019 and January 5, 2019) 64
**Figure 41:** Timeplot of diurnal variation in turbidity concentration (FNU) in drain D1 (as described in figure 14) passing through Narsapur Waterworks Department across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval 65
**Figure 42:** Timeplot of diurnal variation in turbidity concentration (FNU) in Muriki Kaluva waste water canal across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval 66
**Figure 43:** Heatmaps showing seasonal variation of electrical conductivity (µS/cm) in Narsapur stretch of River Godavari during different seasons in 2017 and 2018 67
**Figure 44:** Heatmaps showing seasonal variation of chlorophyll (mg/L) in Narsapur stretch of River Godavari during different seasons 68

**Figure 45:** Timeplots showing variation in dissolved oxygen concentration (mg/L) and electrical conductivity (µS/cm) between July 2017 and June 2019 (noted on X-axis) in Narsapur stretch of River Godavari 69
**Figure 46:** Timeplots showing variation in dissolved oxygen concentration (mg/L) and electrical conductivity (µS/cm) between May 2017 and May 2019 (noted on X-axis) in Rajahmundry stretch of River Godavari 70
**Figure 47:** Diagram depicting various machine learning approaches 71
**Figure 48:** a, b, c and d are the scatter plots between BOD measured and predicted BOD using SVM method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using SVM method with the four sets of input parameters 75
**Figure 49:** a,b,c and d are the scatter plots between BOD measured and predicted BOD using AdaBoost method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using AdaBoost method with the four sets of input parameters 77
**Figure 50:** Diagram depicting the architecture of Artificial Neural Network 78
**Figure 51:** a,b,c and d are the scatter plots between BOD measured and predicted BOD using ANN-MLP method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using ANN-MLP method with the four sets of input parameters 79

# List of Tables

**Table 1:** Parameters measured in W2C system 18
**Table 2:** Limits and resolution of Hanna HI 9829 19
**Table 3:** Limits and resolutions of C3 Fluorometer sensor 20
**Table 4:** Limits and resolutions of Optex sensors 20
**Table 5:** Limits and resolutions UV254 Go! (For 10 mm path length in mg/L) 21
**Table 6:** Town Profiles 32
**Table 7:** Details of sectors identified for Narsapur study 34
**Table 8:** Details of selected drains 35
**Table 9:** Summary of data collected for Water-to-Cloud 43
**Table 10:** Water Quality Classification 47
**Table 11:** Normalized factor (C) values 47
**Table 12:** CPCB limits for discharge into water bodies 51
**Table 13:** BOD and COD load per drain 58
**Table 14:** Waste type basis BOD/COD ratio 59
**Table 15:** Descriptive statistics of water quality data collected across various rivers in India from April 19, 2017 to September 24, 2018 72
**Table 16:** Correlation between COD, BOD and other parameters 73
**Table 17:** R, $R^2$, f and p-values resulted from multiple regression 73
**Table 18:** R, $R^2$, f and p-values resulted from SVM 76
**Table 19:** R, $R^2$, f and p-values resulted from SVM 76
**Table 20:** R, $R^2$, f and p-values resulted from ANN 78

# Acknowledgement

The Water-to-Cloud team would like to thank Professor Srinivas Chary for the opportunity to pilot the Water-to-Cloud system and work with grassroots team in Narsapur, Palacole and Kovvur. Our sincere gratitude to Dr Y Malini Reddy for ensuring coordination between Administrative Staff College of India, Ernst & Young and University of Chicago teams and for timely logistics support.

The team would like to thank the International Innovation Corps team, especially Professor Anup Malani, who co-founded the initiative and made this collaboration possible. Peggy Leung deserves a special mention for her untiring support in managing the operations and finance. We're thankful to Shriya Sethi for her strategic guidance to the project and reviewing this report.

We are grateful to the Bill & Melinda Gates Foundation for supporting us immensely and believing in our abilities. We would also like to express our gratitude to Professor Monisha Ghosh for mentoring the team on building machine learning capabilities. Additionally, Subhojit Goswami from the Tata Centre for Development has been extremely supportive in editing this report.

We are happy to have been supported by various laboratories, including Vimta Labs, Hyderabad and Andhra University Labs, Visakhapatnam for validation of our data. We are also thankful to Xooplabs for designing this report. The project would not have been possible without the on-ground support provided by the town municipalities in Narsapur, Palacole and Kovvur. We extend our gratitude to the municipal staff in all three towns, especially the municipal commissioners and chairpersons. Furthermore, we would like to thank all our colleagues from Administrative Staff College of India, Ernst & Young and University of Chicago for supporting our efforts in this project.

# Executive Summary

Declining water quality has become a global concern as human populations grow, industrial and agricultural activities expand, and climate change threatens to cause major alterations to the hydrological cycle[1]. Seventy percent of surface water in India is not fit for consumption[2]. A pressing need has emerged for comprehensive and accurate assessments of trends in water quality in order to provide a basis for action at all levels.

Effective river basin management and water policy formulation requires access to reliable and up-to-date data in easy-to-understand formats for timely decision making. Traditional laboratory-based testing is tedious, slow and expensive. It is also prone to inaccuracies in results due to error in sample collection, transportation or testing. By relying on a handful of data points from such sampling limits one's capacity to Analyze or generate insights. In contrast, spatial monitoring of river water quality at high resolution can identify pollution hotspots and provide data for putting in place relevant Water, Sanitation and Health (WASH) interventions. Automated, real-time sensors can facilitate measurements at this scale by providing quick results while being less prone to human errors. The incremental cost of data collection using such sensors (once procured) is low. Monitoring river stretches multiple times a month at different times of the day help understand pollution spread and identify trends in water quality over time.

Pritzker School of Molecular Engineering (PME) at University of Chicago developed the Water-to-Cloud (W2C) system for such high resolution spatial and temporal monitoring to gain insights that traditional approach may not capture. State-of-the-art sensors with a geospatial tag and time stamp are taken around a river on a boat to collect hundreds of data points in a few hours. Various physical (pH, electrical conductivity, turbidity), chemical (dissolved oxygen, nitrate) and bio-chemical (tryptophan, chlorophyll) parameters are measured several times a month in an automated manner. In drains, these sensors are taken around manually. Synthesizing this data as color-coded heatmaps by superimposing it on geospatial maps helps in pinpointing contaminant sources, analyzing spread of pollution over time and area as well as identifying trends.

1 UN Water. (2011), Water Quality Policy Brief. Retrieved from https://www.unwater.org/publications/un-water-policy-brief-water-quality/
2 NITI Aayog. (2018). Composite Water Management Index. Retrieved from https://www.niti.gov.in/sites/default/files/2019-08/CWMI-2.0-latest.pdf

W2C supported the Andhra Pradesh government in developing Narsapur, Palacole and Kovvur as model sanitation towns by evolving the Water-to- Cloud approach and monitoring Godavari waters at two sites, Kovvur-Rajahmundry and Narsapur between October 2016 and June 2019. Based on 197,037 data points gathered via 161 boat-based experiments, Godavari water seems safe for bathing (Water Quality Index shows suitability for class B usage[3]). However, water from these 11 sq.km. and 3.2 sq.km. stretches respectively was not always suitable for direct consumption due to fecal contamination (upto 330 CFU/100ml). Heavy metals were found in water which could have led to toxicants entering the food chain via irrigated crops or fish. We also showed how this mapping technique can be used effectively to pinpoint point sources of pollution such as drains and industrial outlets with high spatial precision (<50m range). W2C system also identified slow leaching effects in Narsapur due to non-point sources such as a solid waste dumping site which would otherwise be missed out. Through our studies, we also showed, through a two-month experiment, how stationary sensors can be installed at specific locations such as drains and provide continuous real-time data in a robust, unattended fashion. This can be adapted as a means for ensuring regulatory compliance in secure fashion without the need for human inspection or middlemen.

Studying 25 waste water drains in Narsapur, Kovvur and Palacole through 895 sensor-based measurements and 261 lab-based measurements, we located insanitary toilet zones and slum areas with inadequate sanitation facilities for authorities to set up interventions. Since all waste water in these towns is discharged into Godavari or nearby canals without treatment, regular cleaning of drains carrying it is recommended to avoid clogging and algal growth which may reduce dissolved oxygen levels thereby affecting aquatic life negatively. A need for setting up decentralized sewage treatment plants (STPs) for long-term safe sanitation planning for such towns was also recognized.

We provided digital access to all our data and visualizations through an open access platform (http://thoreau.uchicago.edu/water/home). This includes the 1,97,932 sensor data points for pH, dissolved oxygen (DO), electrical conductivity (EC), turbidity, colored dissolved organic matter (CDOM), tryptophan, etc. as well as 480 lab tested samples for bio-chemical oxygen demand (BOD), chemical oxygen demand (COD), fecal coliform. It ensured accountability and transparency while building trust on our approach. Stakeholders including researchers, practitioners, sanitation inspectors, government officers, media and communities could assess the impact of their own actions on changing water quality and use the latest data in any decision making through this portal. We explored machine learning-based water quality prediction models to reduce reliance on laboratory-based testing.

3 CPCB. (2017). Use based Classification of Surface Waters in India. Retrieved from https://cpcb.nic.in/water-quality-criteria-2/

Non-stationary sensor-based monitoring was thus found to complement and strengthen government's established system of laboratory-based monitoring to make it more effective in the near future. An urgent need for a universally acceptable protocol for sensor-based monitoring was felt. Availability of trained operators, ensuring accurate calibration and maintenance procedures for sensors and upgrading to unmanned vehicles would facilitate faster adoption of this approach at scale. Water-to-Cloud has, in essence, showcased its potential to fill the gaps in the current system, act as a decision making tool for policymakers, invoke public participation in regulation of river water quality and democratize access to river water quality data in the long run.

# 1 Introduction

In September 2015, the United Nations adopted 'Clean Water and Sanitation' as one of the Sustainable Development Goals for 2030. Access to clean and safe water resources is central to this vision. Millions of people residing in river basins across the world are dependent directly or indirectly on river ecosystems. Rivers may encounter pollutants entering them due to industrial, agricultural or anthropogenic activities. India's major rivers are known to be polluted with fecal matter, organic content, and chemical pollution. A 2018 Central Pollution Control Board (CPCB) report shows that between 2016 and 2018 the number of polluted stretches in India's rivers has increased to 351 from 302.[4]

Water quality and waste water monitoring are fundamental tools in freshwater resources management and provide essential information on determining the physical, chemical and/or biological characteristics of water resources, trends and changes over time, and emerging water quality issues. They also provide the means to identify policies and measures to enhance water quality, control water pollution from specific sources, evaluate the efficacy of pollution control interventions, and deal with water quality emergencies[5].

## *1.1 Existing Water Quality Monitoring Scenario*

The traditional water quality monitoring approach involves grab sampling, followed by transportation of samples to a laboratory where they are analyzed using scientific instruments and the results are produced in a couple of weeks. Critical dependence on laboratories and trained technical staff make it a resource intensive and time consuming approach. This approach is also prone to human error at each stage-sample collection, transportation and analysis. Lack of infrastructure, financial resources or institutional and technical capacities to implement this approach could lead to scarce, unreliable or unsystematic data on water quality. Hence this approach, although robust, may not provide holistic information on the health of a river. To illustrate, Central Pollution Control Board (CPCB) is the national agency responsible for monitoring river water quality in India. CPCB collects data once a month via State Pollution Control Boards' staff at 35 stations manually on Godavari, India's second longest river after Ganga. Godavari flows for 1465km from its starting point before emptying into the Bay of Bengal. This implies, currently, only one data point is being collected by the government every 42km approximately which restricts the capacity to draw any meaningful conclusions on water health. Additionally, when such data is not supplemented with metadata on the exact location of data collection, time of the day, etc., usability of such data for practical applications may be further restricted.

4 CPCB. (2018). River stretches for restoration of water quality. Retrieved from http://www.im4change.org/siteadmin/tinymce//uploaded/Report%20on%20River%20Stretches%20For%20Restoration%20Of%20Water%20Quality.pdf
5 UNESCO. (2015). Monitoring water quality and waste water. Retrieved from https://en.unesco.org/waterquality-IIWQ/activities-projects/monitoring-water-quality-waste water

## 1.2 The Need for High Resolution Mapping

High frequency spatial and temporal water quality data will help governments, health experts, and other decision makers make choices for water usage based on near real-time data and ensure regulatory compliance. This can also be used to develop new and innovative policies and aid in better enforcement of those. To study the effects of multiple polluting factors, it is essential to have access to data from continuous measurements.

Alternatives to laboratory testing involve stationary or mobile sensors providing data in real time, usage of satellite imagery or usage of water kits in the field. While satellites allow for high spatial resolution and easy scalability, they are restricted to measuring only a few optical parameters. Also, they need clear sky for generating accurate values. Governing equations have to be created and tested. Water kits are low-cost alternatives but do not provide accurate results as of now. Owing to each individual sample test, this method can be slow and cumbersome. Real-time sensors, on the other hand, provide quick and easy access to data. With the dipping costs of sensing technology, it has become an effective medium to collect continuous water quality data, minimizing the dependence on laboratories and human resources. Stationary sensors allow for continuous monitoring at a single location but regular maintenance and calibration may be a hassle due to accessibility issues. Since these can only be installed at a limited number of sites, spatial variation in pollution is also not monitored. Geo-tagged, time-stamped mobile sensors can be used to collect such data over a given area at high spatial resolutions. These can identify pollution hotspots as well as provide data on violators and efficacy of government sanitation interventions. Although one may miss out on some events that stationary sensors could have noticed, taking around these sensors multiple times ensures that regular events are not missed out. Availability of hundreds of data points in just a few hours makes the incremental cost of collecting each data point negligible. Sharing a set of sensors between various sites further keeps costs in check. Water-to-Cloud is one such approach leveraging sensor technology.

# 2 Measurement and analysis methodology for the Water-to-Cloud system

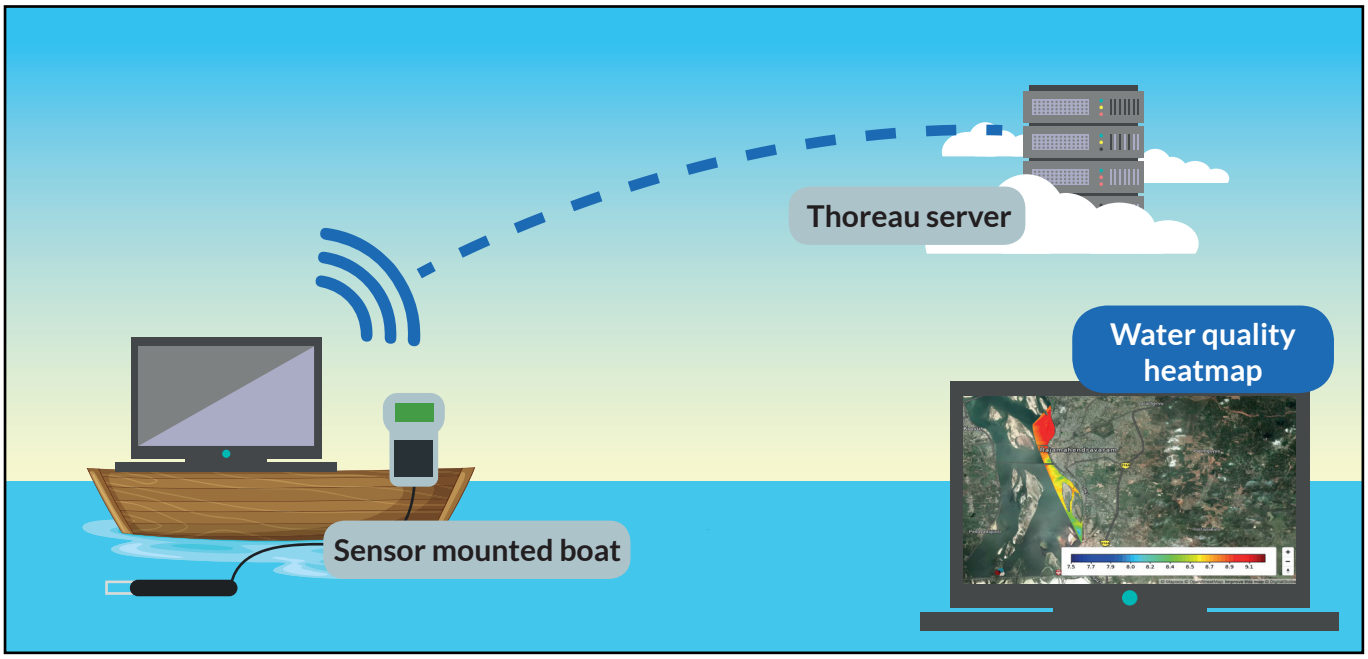



*Figure 1: Illustration of the Water-to-Cloud system comprising a boat mounted with mobile sensors that record water quality data in real time, which is then hosted on a cloud server and disseminated through an open access digital platform in the form of high resolution heat maps*

## 2.1 Overview

The Water-to-Cloud approach involves collecting water quality data at high geospatial resolution using automated, real-time, non-stationary, state-of-the-art cyberphysical sensor networks. These sensors can collect GPS-tagged and time/date-stamped data every few seconds on various water quality parameters. In case of large water bodies such as rivers, they are deployed on a boat which navigates a pre-defined route in the water body and collect hundreds of data points over space and time. In case of smaller non-navigable channels such as drains, they are dipped at pre-defined sampling locations to collect point data in real time. This high-resolution data is superimposed on geospatial maps using a color-coded scale to form visualizations which are easy to interpret and assess various aspects of river water health. In particular, they can be used to pinpoint pollution sources, analyze temporal and spatial variations in contaminant levels and identify trends in water quality.

## 2.2 Data Collection Methodology

### 2.2.1 Site and Route Selection

Based on the objectives of the monitoring exercise, the exact locations for measurement are chosen on a river. This may include an urbanized area to study the effect of domestic waste on the river or an industrial area to understand the impact of illegal discharge of industrial waste into the river or both. At each location on the river, the preferred route for data collection is selected in consultation with local authorities and ground staff. This route is marked to include all point and non-point sources of pollution entering the river including domestic waste water drain outlets, industrial outlet points, solid waste dumping sites and centres of domestic or religious activities such as washing clothes, offering flowers, fruits and grains. This provides an overall picture of total waste water being discharged into the river. A boat ride is then taken along the chosen route to look for practical concerns of accessibility associated with the identified route such as low depth and excess algae, which may hinder boat maneuvering or cause sensors to entangle. This feedback helps decide the final route which is usually a zigzag path rather than a circular one, for better mapping. More data points are collected around a drainage point or around an industry to identify the spread of pollution from a point source.

In case of drains, where sensors cannot be carried around in a boat, these mobile sensors are dipped at pre-identified sampling points to obtain in-situ data in real time. Critical locations to be monitored along the drain are identified in consultation with the local authorities based on key water management questions they wish to address. This may include identification of insanitary toilets, source apportionment, analyzing trends in waste water generation in a town, etc. Locations which require a 24x7 monitoring are monitored using stationary sensors.

### 2.2.2 Parameter Selection

Central Pollution Control Board has divided the major pollutants in six categories[6]:

- General pollutants including color, odor, temperature, pH, electrical conductivity (EC), dissolved oxygen (DO), turbidity and total dissolved solid (TDS)
- Nutrients: nitrogenous compounds and phosphate
- Demand Parameters: Bio-chemical Oxygen demand and Chemical oxygen demand
- Inorganic ions including heavy metals
- Other location specific parameters: Arsenic, Fluoride
- Microbiological parameters: Total coliform and fecal coliform

Out of all these categories of pollutants, general water quality parameters, demand parameters and microbiological parameters are often used to assess the overall health of the river system. Other parameters are used based on locations to assess specific pollutant. Details about each of these parameters and their impact on human and aquatic life are provided in appendix V. A combination of sensor and laboratory-based testing methods are employed to monitor parameters of interest. The table below outlines the various parameters we measure, their common sources and whether they are measured in a lab or through sensors.

6 Guidelines on Water Quality Monitoring. Retrieved from https://www.cpcb.nic.in/wqm/Guidelines_Water_Quality_Monitoring_2017.pdf

| Parameter | What it measures | Common source | Senor/ Lab-based |
|---|---|---|---|
| **Physical parameters** | | | |
| Temperature | Heat | Multiple | Sensor |
| Electrical Conductivity | Salinity, Total Dissolved Solids | Inorganic salts/ions | Sensor |
| pH | Concentration of hydrogen ions | Multiple | Sensor |
| Turbidity | Optical water clarity | Multiple | Sensor |
| Total Hardness | Amount of dissolved calcium and magnesium salt in water | Weathering of rocks and anthropogenic activities | Lab |
| Total Suspended Solids | Water clarity | Suspended particles can come from soil erosion, runoff, discharges, stirred bottom sediments or algal blooms | Lab |
| Dissolved Oxygen | Concentration of molecular oxygen | Photosynthesis, Atmospheric dissolution, | Sensor |
| Chemical Oxygen Demand | Amount of oxygen required to oxidize organic pollutants | Organic wastes from agricultural run offs and industries | Lab |
| Bio-chemical oxygen demand | Amount of oxygen required for bio degradable organic matter | Organic wastes from agricultural run offs and industries | Lab |
| **Major Ions** | | | |
| Nitrate | Concentration of nitrate-N | Agricultural runoff, industrial wastes and sewages | Sensor |
| Nitrite | Concentration of nitrite-N | Acid rains, agricultural runoff | Lab |
| Ammonium | Concentration of ammonia-N | Agricultural run-off, industrial wastes | Lab |
| Chloride | Concentration of chlorides | Erosion, weathering of rocks, industrial wastes and domestic wastes | Lab |
| Sodium | Concentration of sodium | Erosion, weathering of rocks, industrial wastes and domestic wastes | Lab |
| Potassium | Concentration of potassium | Erosion, weathering of rocks, agricultural/industrial wastes | Lab |
| Calcium | Concentration of calcium | Erosion, weathering of rocks and industrial wastes | Lab |
| Magnesium | Concentration of magnesium | Erosion, weathering of rocks and industrial wastes | Lab |
| **Microbiology** | | | |
| Total Coliforms-MPN | Indicate the presence of other pathogenic organisms | Sewage | Lab |
| Fecal coliforms | Indicate the presence of other pathogenic organisms | Sewage | Lab |
| **Bio-Chemical Parameters** | | | |
| Tryptophan | Fecal contamination | Sewage | Sensor |
| Chlorophyll a | Phytoplankton | Nutrient concentration; contamination from man-made activities | Sensor |
| Colored dissolved organic matter (CDOM) | Organic carbon | Agricultural/industrial runoff, sewage | Sensor |
| **Heavy Metals** | | | |
| More than 20 heavy metals are analyzed | Cadmium, lead, mercury, chromium, manganese etc., | Weathering, Industrial effluents and pesticides | Lab |

*Table 1: Parameters measured in W2C system*

### 2.2.3 Sensor Identification and Calibration

Based on the parameters of interest, real-time sensors which collect GPS-tagged and time/date stamped data are chosen. Currently, the system utilizes various kinds of sensors as detailed below.

*Figure 2: Pictures of different types of mobile and stationary real-time sensors used in the Water-to-Cloud system*

### Hanna HI9829

The HI9829 is a waterproof portable logging multi-parameter meter that monitors up to 14 different water quality parameters (7 measured, 7 calculated). Some of the parameters that the sensor measures include pH, oxygen reduction potential (ORP), conductivity, dissolved oxygen, turbidity, ammonium ions, chloride ions, nitrate ions, and temperature. It has a time stamp capability and consists of a GPS for location tracking of measurements. The system is built to withstand harsh environments and is ideal for field measurements of lakes, rivers and sea[7].

| Parameters | Range | Revolution | Accuracy | Calibration |
|---|---|---|---|---|
| Temperature | -5.00 to 55.00 °C | 0.01 °C | ± 0.15 °C | Automatic at 1 custom point |
| pH | 0.00 to 14.00 pH | 0.01 pH | ± 0.02 pH | Automatic 1, 2 or 3 points with automatic recognition of 5 standard buffers (pH 4.01, 6.86, 7.01, 9.18, 10.01) and 1 custom buffer. |
| Dissolved Oxygen | 0.0 to 500.0 % (0.00 to 50.00 ppm (mg/L) | 0.1% 0.01 ppm (mg/L) | 0.0 to 300.0 %: ± 1.5 % of reading or ± 1.0 % whichever is greater;300.0 to 500.0 %: ± 3 % of reading 0.00 to 30.00 ppm (mg/L): ± 1.5 % of reading or ±0.10 ppm (mg/L) whichever is greater;30.00 ppm (mg/L) to 50.00 ppm (mg/L): ± 3 % of reading | Automatic 1 or 2 points at 0, 100 % or 1 custom point |
| Conductivity | 0 to 200 mS/cm | 1 µS/cm | ±1 % of reading | Automatic single point, with 6 standard solutions (84 µS/cm, 1413 µS/cm, 5.00 mS/cm, 12.88 mS/cm,80.0 mS/cm, 111.8 mS/cm) or custom point |
| Turbidity | 0.0 to 99.9 FNU; 100 to 1000 FNU | 0.1 FNU from 0.0 to 99.9 FNU, 1 FNU from 100 to 1000 FNU | ±0.3 FNU or ±2 % of reading | Automatic 1, 2 or 3 points at 0, 20 and 200 FNU, or custom |

*Table 2: Limits and resolution of Hanna HI 9829*

7 Hanna Instruments (2012). HI9289 Multiparameter meter Instruction Manuall. Retrieved from https://hannainst.in/downloads/dl/file/id/1035/hi9829_2012_04_13.pdf

### Turner C3 Submersible Fluorometer

The C3 Submersible Fluorometer is used to measure tryptophan, chlorophyll, colored Dissolved Organic Matter (CDOM). It is based on optical sensing ranging from deep ultraviolet to the infrared spectrum. The sensor requires an external lithium ion battery for extended deployment. The sensor comes with a factory-installed temperature sensor as well. The sensor is cased robustly and is highly resistant to harsh environments. A C-Soft Windows based software can be downloaded into a computer which allows for intuitive calibration, datalogging set-up and file management capabilities as well as digital or analog data integration[8].

| Parameters and specification | Minimum Detection Limit | Linear Range |
|---|---|---|
| Chlorophyll-a (in vivo) Blue excitation | 0.03 µg/L | 0-500 µg/L |
| Colored dissolved organic matter | 1.15 ppb 0.5 ppb | 0-1500 ppb<br>0-3000 ppb |
| Tryptophan | 3 ppb | 0-5000 ppb |
| Temperature | 0.1 °C | -2 to 50 °C |

*Table 3: Limits and resolutions of C3 Fluorometer sensor*

### FP211 Flow Probe

Flow Probe is a highly accurate water velocity instrument for measuring flows in open channels and partially filled pipes. The water velocity probe consists of a protected water turbo prop positive displacement sensor coupled with an expandable probe handle ending in a digital readout display. The water flow meter incorporates true velocity averaging for the most accurate flow measurements. It has range of 0.3 -19.9 Feet Per Second (FPS) and an accuracy of 0.1 FPS. The Flow Probe is ideal for storm water runoff studies, sewer flow measurements, measuring flows in rivers and streams, and monitoring water velocity in ditches and canals. The flow velocity values obtained from this instrument are used to calculate the waste water load in drains by multiplying them with drain dimensions[9].

### Optex Turbidty and Dissolved Oxygen Sensor

The Optex Turbidty and DO sensors are stationary sensors that are used to measure water quality in drains. They measure turbidity, DO and temperature values every 5 minutes 24X7 and are powered by solar panels. The data from these sensors is wirelessly transmitted on to the cloud and can be accessed through a dashboard. The Optex Turbidity Checker TSC-10(E) uses a method of 90-degree scattered light as the measuring principle of the world common standard. It has a built-in wiper cleaning system and the optical windows are made of hard-to-scratch sapphire glass and therefore can be manually cleaned. The compact design makes it possible to install the sensor easily. The Optex DOS-20 DO sensor can measure both DO levels and the temperature. It is not influenced by the water flow rate[10].

| Parameters | Range | Accuracy |
|---|---|---|
| DO (mg/l) | 0-20 mg/l | ±0.1mg/L (0~8mg/L), ±0.2mg/L (8~20mg/L |
| Turbidity (FNU) | 0-50 FNU | 0.01(0.00-19.99), 0.1(20.0-500.0) FNU |
| Temperature (degree C) | 0-50 degree C | 0.1 degree C |

*Table 4: Limits and resolutions of Optex sensors*

8 Turner Designs (2018). C3 & C6p submersible fluorometers User's Manual. Retrieved from http://docs.turnerdesigns.com/t2/doc/manuals/998-2300.pdf
9 Xylem Products (2018). FP211 Water Flow Probe. Retrieved from http://www.globalw.com/products/flowprobe.html
10 Optex Sensing Innovation (2019). Optex Products Manual. Retrieved from https://www.optex.co.jp/e/products/

### UV254 Go!

Photonic Measurement's UV254 Dip Go! is a light-weight, portable device, which allows easy measurement of open channel water for parameters such as total organic carbon (TOC), dissolved organic carbon (DOC), COD and BOD. It has a simple icon-based user interface that makes it easy to navigate between measuring, calibration and data collection. It also has extensive storage capacity to preserve numerous measurements. A user can connect and transfer data to a computer system through a USB port. The system has the ability to generate graphical charts of data patterns[11].

| Parameters | Range | Accuracy |
|---|---|---|
| UVT and SUVA | 0-100% UVT 0-2.5 ABS, is extendable with shorter path length cuvettes | ±0.5% UVT |
| Total Organic Carbon | 0-120 ppm | 0.1 ppm |
| BOD | 0-140 mg/l | 0.1 mg/l |
| COD | 0-280 mg/l | 0.2 mg/l |
| Dissolved Organic Carbon | 0-100 mg/l | 0.1 mg/l |

***Table 5:*** *Limits and resolutions UV254 Go! (For 10 mm path length in mg/L)*

All sensors are calibrated before taking measurements. Before any instrument is calibrated or used to perform environmental measurements, the instrument is stabilized according to the manufacturer's instructions and it is ensured that there are no air bubbles lodged between the probe and probe guard. As a quick calibration measure, the sensor can be calibrated prior to use on the day the measurements are to be performed. A post calibration check at the end of the day is performed to determine if the instrument drifted out of calibration. A detailed calibration procedure for each instrument and its parameter is attached in appendix VI.

### 2.2.4 Monitoring Frequency

CPCB guidelines require collection of data once a month from identified manual sampling stations. This way it is able to collect, on an average, one data point a month from a stretch of 42 km in River Godavari. This provides scarce data and may not be useful for proper analysis. In contrast, sensors can collect data at sampling intervals as low as one second. We usually tune them to record a parameter value every 10 seconds so as to collect data every few metres. Water to-Cloud system, thus, collected 1000-1500 points over a single boat ride lasting 3-4 hours in a 4-6 km stretch of river. Generally, four boat rides are carried out in a month-two in the morning and two in the evening. In case of drain monitoring, sensors are dipped into the drains on a biweekly basis.

11 Photonic Measurements (2018). Photonic User's Manual. Retrieved from https://photonicmeasurements.com/products

### 2.2.5 Operationalization

There are multiple options to carry sensors around a water body such as remote-controlled (RC) float boats, manual or motor boats, unmanned automated drones, etc. While commercially-available drones or float boats provide a sophisticated means to carry sensors and ensure navigability in smaller streams, drains or channels with shallow waters, they are expensive, require trained staff for operation and would present limitations in carrying lithium-ion batteries and laptop on board along with the multiple sensors. Manual boats or large vessels would be slow and cumbersome in surveying large water bodies. A motor boat driven by a boat person was found to be the most viable and accessible means to carry sensors in rivers due to their easy availability at most sites at affordable rates. Boats are also able to accommodate a technical professional on board to ensure continuous, obstruction-free working of the sensors. In the future, the dependence on boatmen and availability of boats can be minimized by designing water float boats.

Sensors are tied to a rod and immersed continuously at a depth of 1-3 feet under the surface of water. The rod is attached firmly to the side of a boat and the sensors record readings as the boat moves around. We found that a boat speed of 5-6 km/hr is suitable such that the turbidity sensor is not affected by the motion of the boat as it records a value every 15-20 m. Hanna HI 9829 and C3 Submersible Fluorometer are deployed in this manner. UV254 Go! is also attached with these sensors. HI 9829 does not require an additional power source. C3 sensor can be powered by a Lithium-ion battery. FP211 flow probe is used on the boat to note downstream velocity from time to time. Optex Turbidity and Dissolved Oxygen Sensors are fixed at critical locations in a drain and are powered using solar energy.

***Figure 3:*** *Picure showing deployment of a mobile sensor mounted on boat using cable ties, coir rope and a cane*

***Figure 4:*** *Picture showing deployment of mobile sensor (in this case, a flowmeter) in a waste water drain by dipping it at a depth of 6-12 inches to measure water velocity*

*Figure 5: Picture showing deployment of solar-powered stationary sensors (in this case, Dissolved Oxygen and Turbidity sensors) in a waste water drain with the control panel and data logger enclosed in a steel box for protection*

## 2.3 Data Storage and Curation

### 2.3.1 Data Retrieval

Laboratories may take up to two weeks to analyze water quality samples and deliver results. This time period may be further increased depending on availability of equipment and staff capacity. This slows the access to data. A quick data retrieval and curation system is critical for the Water-to-Cloud approach. Data is retrieved from the sensors in the form of excel files after each boat ride by connecting the logging meter to a computer with a pre-installed setup. The excel files contain geospatial data for different water quality parameters tagged with date and time. These files are processed and converted to CSV formats for further processing. In case of stationary sensors, data is wirelessly transmitted onto the cloud through Global System for Mobile Communications (GSM)-based sim cards which are connected to the data logger.

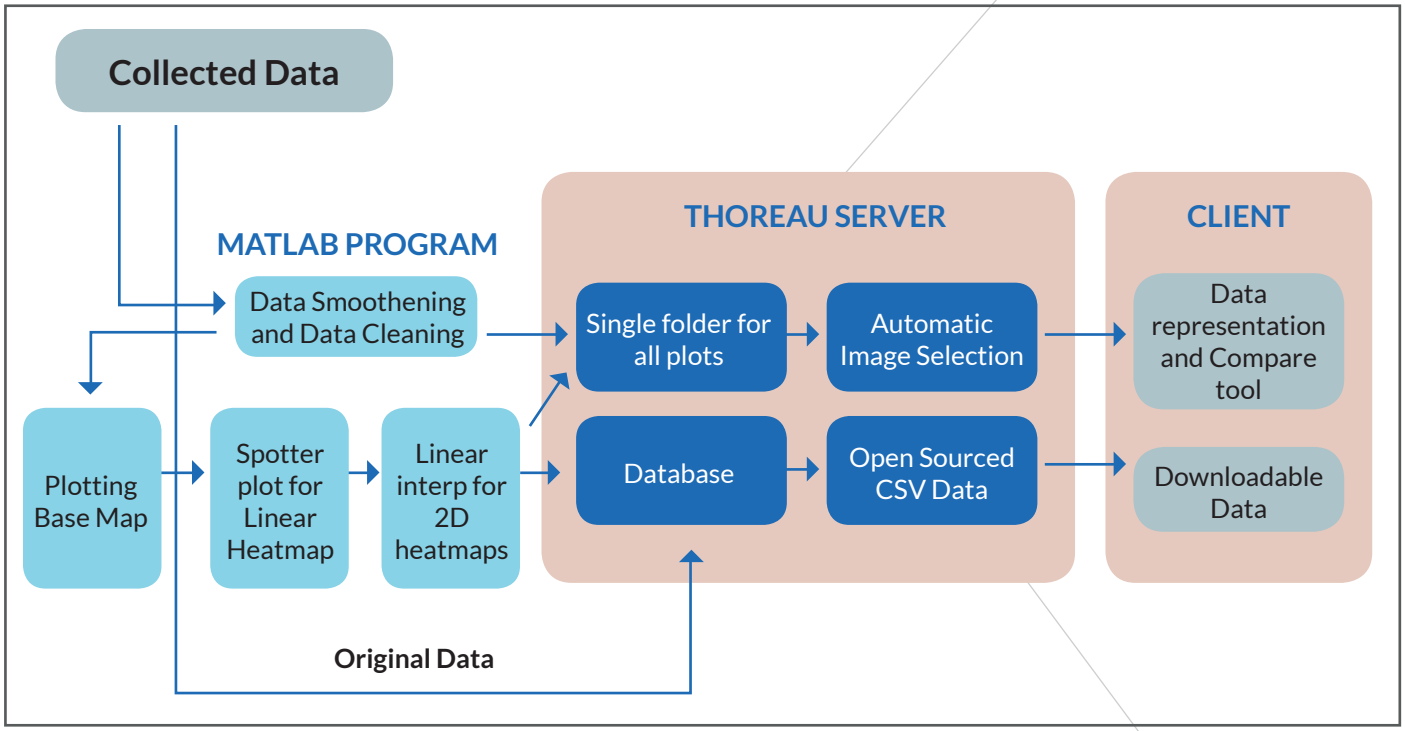



*Figure 6: Diagram showing dataflow in the Water-to-Cloud system through the collection, processing and display stages*

### 2.3.2 Data Processing

Data needs to be cleaned up before creating visualizations from it. A Matlab code reads the data into data frames to find the maximum and minimum value for each parameter. The data is cleaned by removing any undefined values, empty cells and empty rows from the frames. Noise in the data is reduced by using a moving average filter of five. This smoothens the data. For better resolution of images, MATLAB's image processing toolbox is used. The interpolation is done on a mesh grid generating heat-map image for any specific parameter. For non-water areas, mesh cell values are made to be Not-a-Number (NaN) resulting in processing of only water areas. A base-map is plotted using OpenStreet map API from MapBox similar to google maps and the geographical coordinates available in the data file. The data range is scaled with the color-bar of the heat-map. Scatter and Interp functions are used to plot the boat route and to plot the interpolated 2-D heatmap. The interpolated image is overlaid on the base map. This creates a heat map that represents spatial and temporal variation of a selected water quality parameter across the river.

### 2.3.3 Digital Access

To democratize access to water quality data, it is made available to everyone. These heat maps can be accessed at http://thoreau.uchicago.edu/. The website is hosted on the UChicago server. It is developed and designed mainly in Django virtual environment, JavaScript and Cascading Style Sheets (CSS). The plots are uploaded on the web-server, and are available for browsing over different dates. This process is automated by printing plots over the webpages through a python script. All data can also be downloaded from the website by clicking the download button. A dashboard provides an overview of all the experiments conducted by Water-to-Cloud across India so far. It provides a visual representation of the water quality parameters collected in different parts of India. This enables comparison of parameters among different sites and time trend Analyzes of parameters at a particular site.

## 2.4 Analysis

### 2.4.1 Visualization

Conventionally, water quality data is presented in tables or charts that are difficult to interpret or analyze. A snapshot from CPCB website shows the data from 2013 for Godavari. It was last updated in November 2018[12].

| STATION CODE | LOCATIONS | STATE | TEMPERATURE °C | | | D.O. (mg/l) | | |
|---|---|---|---|---|---|---|---|---|
| | | | Min | Max | Mean | Min | Max | Mean |
| WATER QUALITY CRITERIA | | | | | | > 4 mg/l | | |
| 1312 | GODAVARI AT JAYAKWADI DAM, AURNAGABAD,MAHARASHTRA | MAHARASHTRA | 28 | 31 | 29.2 | 5.9 | 7 | 6.4 |
| 2177 | GODAVARI RIVER NEAR SOMESHWAR TEMPLE. | MAHARASHTRA | 23 | 28 | 24.5 | 4.9 | 7 | 6 |
| 2182 | GODAVARI RIVER AT SAIKHEDA. | MAHARASHTRA | 23 | 28 | 25.8 | 4.5 | 6.5 | 5.5 |

*Figure 7: Screenshot depicting traditional tabular form of presenting data which is difficult to read on various water quality parameters at different CPCB monitoring stations on River Godavari*

12 Webpage can be accessed at https://www.cpcb.nic.in/wqm/2013/RIVERWATER%20DATA%202013_22.htm

Color-coded visualisations, on the other hand, are easy to understand for laypersons, sanitation staff working on ground and decision makers. Data is visualized as 1D heatmap which represents the variation of water quality along the boat route and as 2D heatmap which interpolates the data from the 1D maps to be spread over a larger surface area. Interpolating line maps allows us to create area maps for better visualizations. These provide a user-friendly way to carry out analysis without getting entangled with huge amounts of data or too many technicalities.

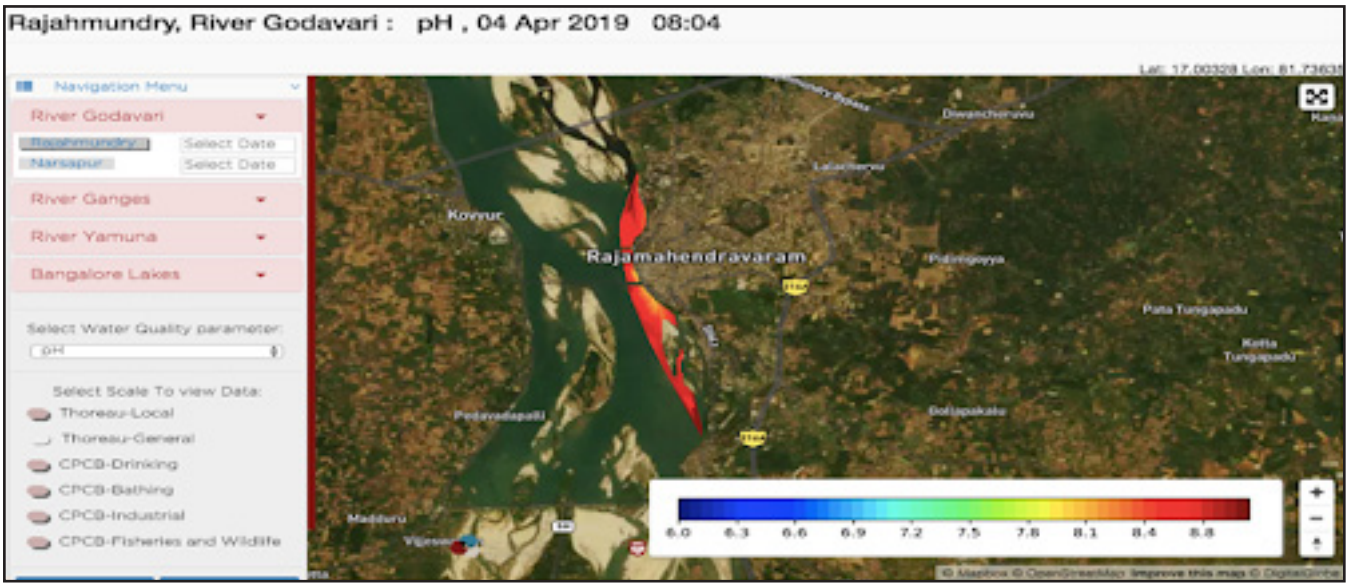



*Figure 8: Screenshot of a 2-dimensional, interactive representation of water quality data collected through the Water-to-Cloud system hosted on the open access thoreau platform with the option to choose water body, parameter and the type of scale given to the user*

These interactive maps allow the user to choose locations, dates, water quality parameters and the scale. They are also used to visualize lab-based point data across a map. Drains heatmap allows one to see lab and sensor data collected from different drains in a town, select/deselect the drains and select the timespan for the data. A similar heatmap approach is also used for drain monitoring as shown below. The average values of parameters are highlighted at the sampling points of each drain. A time slider can be used to select the time period between which values need to be averaged. This drains heatmap can be effective in monitoring the overall status of pollution in a town.

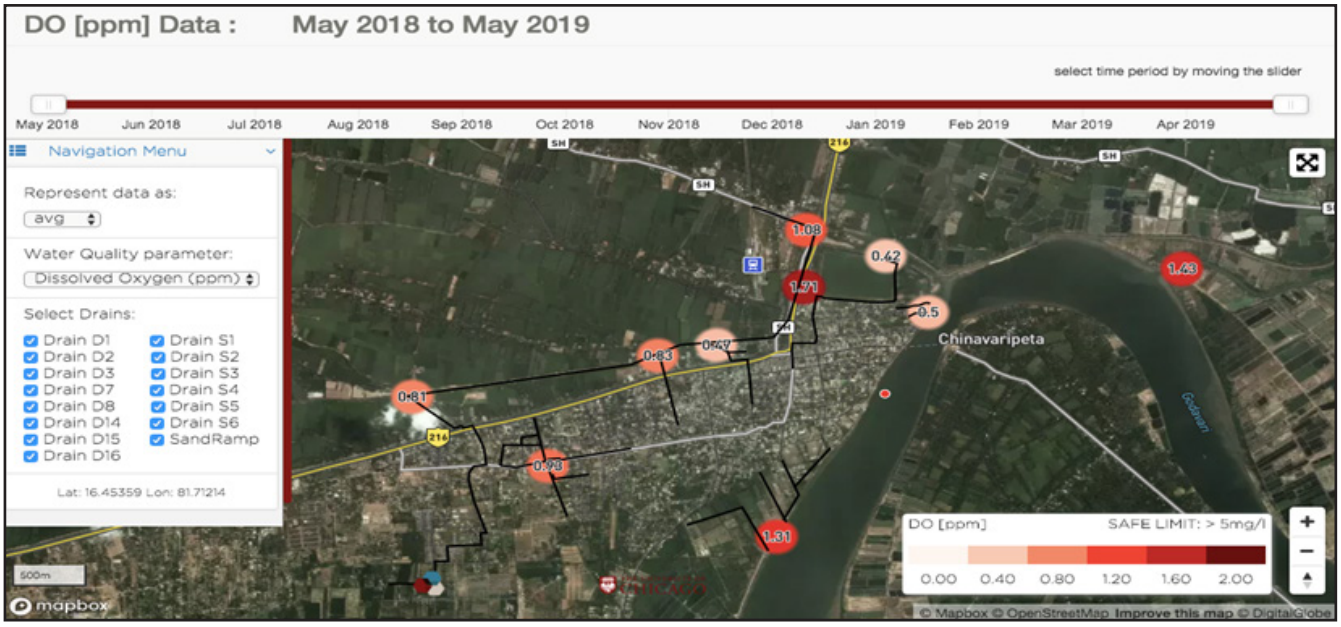



*Figure 9: Screenshot of an interactive heatmap showcasing water quality at various drain sampling points in Narspur with the options to select time range, parameters, drain and the type of analysis*

## 2.4.2 Laboratory Analysis

Sensor-based parameters are validated through laboratory measurements. Furthermore, no reliable sensors are available for a number of organic and inorganic parameters, including heavy metals. Hence, lab analysis provides additional layers of data to supplement the sensor data for any site. Grab samples are taken at the same location and at the same time as that of sensor readings. The lab results of this measurement are then compared with the corresponding sensor readings. This is to ensure that the calibrated sensors are working properly and the sensors values are within a reasonable range of the lab values. Samples are collected by following CPCB guidelines[13]

13 CPCB. (n.d.) Guide Manual: Water and Waste water Analysis. Retrieved from https://cpcb.nic.in/openpdffile php?id=UmVwb3J0RmlsZXMvMjA0XzE1MjQ2NTA4OTNfbWVkaWFwaG90bzEyODI3LnBkZg==

### 2.4.3 Statistical Analysis

While heatmaps are good to visualize variations in parameters over a given area on a certain day, statistical averages such as mean, median, mode and standard deviations may be used to understand the macro picture. In order to look at data from a particular site over different seasons or to compare water quality at two different sites, number of basic statistical tools were applied on the data gathered using sensors or lab sampling to provide powerful insights. Box plots, time series analysis and scatter plots have been used in later sections to bring out insights from the data. A description of these tools is given in appendix VII.

### 2.4.4 Machine Learning

Parameters such as BOD and COD are critical to understand organic pollution in water. However, it is challenging to monitor these parameters reliably through real-time, in-situ sensors. Currently, we rely on laboratory measurements which are both time and cost intensive. BOD test takes upto five days in laboratory. As a result, the number of data points available on these parameters are limited. With a lot more data available from in-situ (measurable on-site using sensor) parameters such as temperature, EC, pH, DO), etc., Water-to-Cloud team explored the possibility of predicting ex-situ (measurable in labs) parameters such as BOD and COD by drawing correlations with in-situ parameters. The idea was to use sensor fusion for measuring multiple parameters in real time and neural network models to estimate lab-based parameters from sensor-based parameters. This would reduce the reliance on costly, time consuming lab measurements. Various machine learning (ML) models were applied and the results are explained in a later section. Such models, if successful, would allow for a quick and efficient estimation of such crucial parameters to develop a deeper understanding of the pathogen levels in rivers.

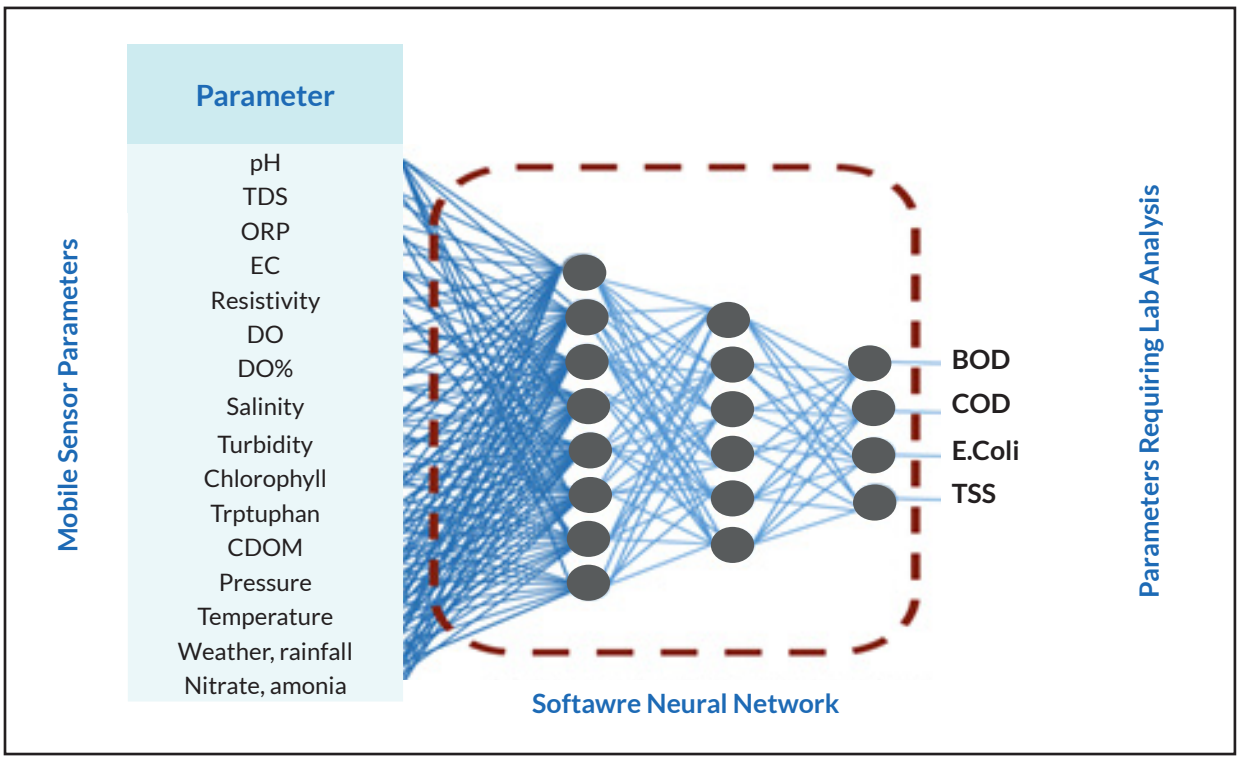



*Figure 10: Diagram depicting machine learning modelling to predict lab-measured parameters using sensor measured parameters by establishing correlation between them*

# 3 Godavari Monitoring Study

## 3.1 Project Context

Government of Andhra Pradesh (GoAP) launched Swachh Andhra Mission with a goal of achieving total sanitation in line with the vision of Swachha Bharat Mission (Urban). Sampoorna Swachhta Sankalp (S3) is an initiative of the Administrative Staff College of India (ASCI), Ernst & Young (E&Y) and the International Innovation Corps (IIC) to design, implement, and monitor improved decentralized sanitation systems in Andhra Pradesh to support this vision. Pritzker School of Molecular Engineering (PME) at University of Chicago was sub-contracted to ASCI through Bill and Melinda Gates Foundation to study the impact of waste water on River Godavari and monitor waste water quality of drains in riverfront towns of Narsapur, Palacole, and Kovvur under S3. These towns in the west Godavari district were suggested by the Municipal Administration and Urban Development Department (MA&UD) for transformation into model towns in terms of sanitation facility.

Water-to-Cloud system, developed by PME, would help understand how riverfront towns impacted water pollution, and also the impact of S3's septage management interventions across the three towns. Time varying sources of pollution such as malfunctioning sewage treatment plants could also be pinpointed through this technique. The project thus aimed at demonstrating that scalable water quality mapping systems can detect and predict water contamination.

## 3.2 Study Area

Godavari originates in Maharashtra and empties into Bay of Bengal via Andhra Pradesh after covering a length of 1465 km. It is India's second longest river after Ganga. Water-to-Cloud began its study in the Kovvur-Rajahmundry stretch and later scaled it to Narsapur stretch. The sites were selected in consultation with ASCI and local administration to monitor the impact of domestic, industrial or agricultural waste water entering from the riverfront towns into the river.

Rajahmundry and Kovvur are located on opposite banks of Godavari. Rajahmundry is hailed as the cultural capital of Andhra Pradesh and is revered as a religious place. Hundreds of pilgrims bathe at the ghats in Rajahmundry each year. It has a population of 341,831 as per 2011 census. Rajahmundry has a sewerage treatment plant with 30MLD capacity. However, certain reports estimate that the city produces 60 MLD of waste and only 10MLD gets treated at this plant. Ava drain carries waste water from the outlet of this treatment facility. In fact, most of the city's waste is carried by Ava drain and Nala channel[14].

14 Benjamin. (2016). Poor Drainage makes life miserable in Rajamahendravaram. Retrieved from thehansindia: https://www.thehansindia.com/posts/index/Andhra-Pradesh/2016-10-13/Poor-drainage-makes-life-miserable-in-Rajamahendravaram/258364

*Figure 11: Map of Rajahmundry in the East Godavari District of Andhra Pradesh depicting the boat ride route in River Godavari, waste water carrying drains, paper mill and sewage treatment plant*

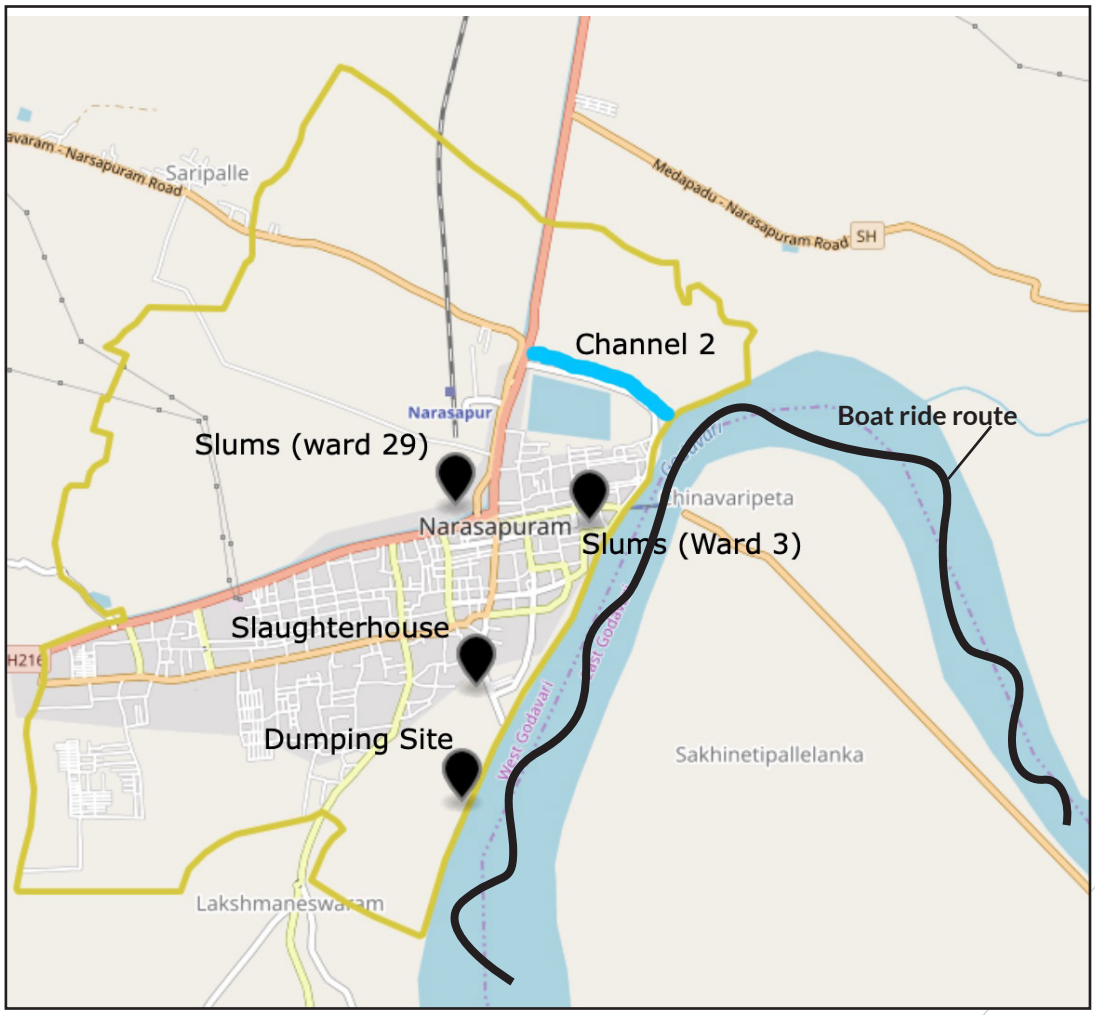



*Figure 12: Map of Narsapur in the West Godavari District of Andhra Pradesh depicting the boat ride route, slums, dumping site and slaughterhouse*

Narsapur is a relatively small town about 20 km before Godavari meets the Bay of Bengal. The waters here are highly affected by sea waters flowing back into the river. Since Narsapur does not have any major industries, most of the waste water is characterized by domestic waste. A solid waste dumping site is situated at the bank of Narsapur. A fecal sludge treatment plant (FSTP) has been set up by the town with support from ASCI in 2017. A slaughterhouse is found to discharge its waste water into a nearby drain directly. Water from a canal in the town, Pilla Kaluva is used for drinking by some of the village dwellers downstream. Unfortunately, some drains are connected directly to this canal in Narsapur rendering this water unsafe for consumption.

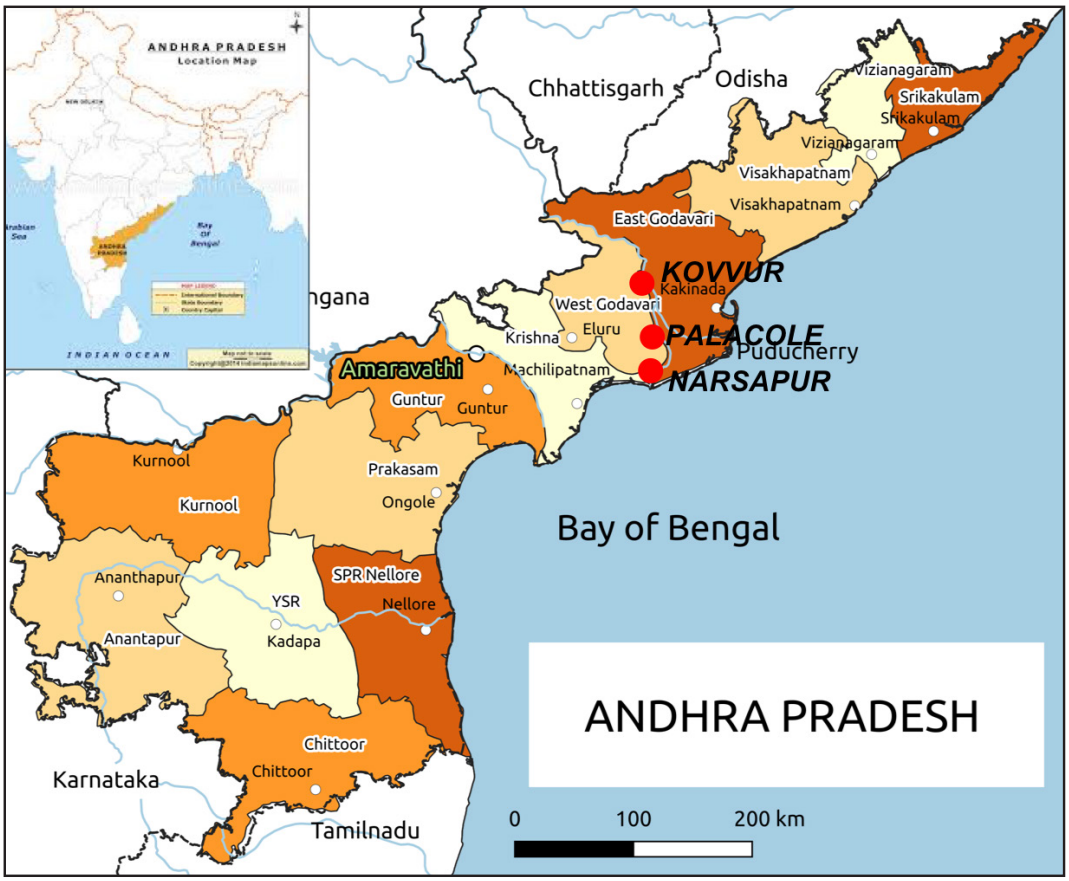



*Figure 13: Map showing three sites - Kovvur, Palacole and Narsapur identified for monitoring of wastewater drains*

A need was felt to go beyond monitoring river water quality to assess the waste water quality in different drains in Narsapur, Palacole and Kovvur. The general profile of the three towns is explained on the table below and the details of the drainage system in each town are described thereafter.

| Parameters | Narsapur | Palacole | Kovvur |
|---|---|---|---|
| Population (no.) | 58901 | 61284 | 39706 |
| Area (sq. km) | 11.32 | 4.7 | 24.56 |
| Density (persons per sq km) | 5203 | 13081 | 1617 |
| Decadal population growth (2001-11) (%) | 1 | 6 | 9 |
| Slum population as a percentage of total population | 60% | 40% | 50% |
| Sex ratio (Females per thousand of males) | 1041 | 1027 | 1084 |
| Literacy (%) | 80 | 79 | 73 |
| Individual water supply connections (%) | 44% | 46% | 30% |
| Availability of pucca drainage (%) | 92 | 80 | 52 |
| Road network (km) | 58 | 90.5 | 55 |

*Table 6: Town Profiles*

### 3.2.1 Narsapur drainage system

Narsapur was the first town selected for monitoring of the drain water quality. This monitoring took place between May 2018 and June 2019 on a weekly basis. The major drains in the city with minimum flow greater than 0.01MLD were selected for monitoring. Thirteen major drains with a cumulative waste water load of over 9 MLDs contributed by nearly 60,000 residents were identified. Cumulatively, they carry over 90 percent of waste water generated in Narsapur each day. To understand the drainage system in the city, it has been divided into three major sectors based on the collective outfall points of the drains.

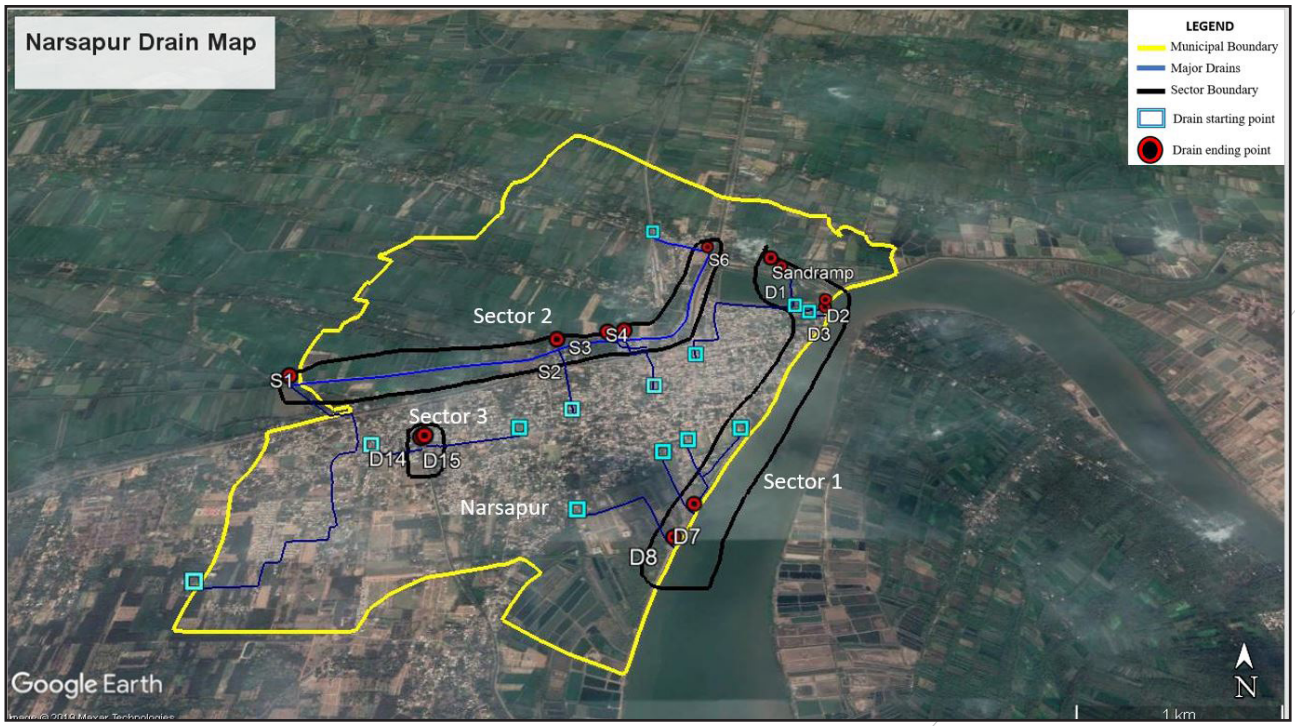



*Figure 14: Map of Narsapur depicting the drain pattern (in blue) and 13 sample collection points at the end of open drains (D1, D2, D3, D7, D8, D14 and D15), siphoned drains (S1, S2, S3, S4 and S6) and at Sand Ramp, where the waste water from sector 2 meets River Godavari*

### Sector 1: River Godavari

In sector 1, there are approximately 25,000 residents who discharge their waste water into River Godavari. Out of these 10 drains, five major drains i.e. D1, D2, D3, D7 and D8 with a total discharge around 4.74 MLD have been identified for monitoring. These drains discharge directly into River Godavari.

| Parameters | Sector 1 | Sector 2 | Sector 3 |
|---|---|---|---|
| Outfall canal/ river | Godavari | Pilla Kaluva | Muriki Kaluva |
| Total no. of drains (approx.) | 10 | 7 | 6 |
| No of drains identified for monitoring | 6 | 2 | 5 |
| Waste water generated (in MLDs) | 4.74 | 0.22 | 4.22 |
| Contributing population | 25748 (54%) | 3144 (6%) | 29108 (40%) |

***Table 7:*** *Details of sectors identified for Narsapur study*

### Sector 2: Pilla Kaluva

Pilla Kaulva is a freshwater canal that is a drinking water source for residents living downstream from Narsapur. However, a population of around 3000 living in five wards discharges waste water into this canal directly. There are around 15 small drain outlets which empty into Pilla Kaulva at various locations. The study is monitoring two major drains-D14 and D15-which contribute around 0.22 MLD of waste water, which might cause health issues to people who consume the canal's water for various purposes.

### Sector 3: Muriki Kaluva

Muriki Kaluva is a drain that flows into the town from north-west. About 4.22 MLD of waste water is discharged in to Muriki Kaluva by a population of approximately 29000 through the five major drain outlets-S1, S2, S3, S4 and S6 that have been identified for monitoring. It also carries waste water from other towns both from the western and northern parts of the west Godavari districts. All of the waste water in Muriki Kaluva is discharged into River Godavari via the canal at the sand ramp area. This canal meets Godavari in the upstream part of the town.

The details of all the selected drains in this study are provided in table below.

| Drain code | Sector | Lat. | Long. | Point where sub drain meets main drain |
|---|---|---|---|---|
| D1 | Sector 1 | 16.44586 | 81.70757 | drain outlet close to the burial ground |
| D2 | Sector 1 | 16.44274 | 81.71004 | drain outlet close to public toilets |
| D3 | Sector 1 | 16.44208 | 81.70981 | drain outlet nearer to the Durgamma temple |
| D7 | Sector 1 | 16.42684 | 81.70076 | drain outlet nearer to FSTP plant |
| D8 | Sector 1 | 16.42412 | 81.69976 | drain outlet nearer to the solid waste management site |
| D14 | Sector 2 | 16.43215 | 81.68577 | near pillakaluva-5th outfall point |
| D15 | Sector 2 | 16.43217 | 81.68576 | near pillakaluva-6th outfall point |
| S3 | Sector 3 | 16.43587 | 81.67757 | Kushana bed locations |
| S2 | Sector 3 | 16.43875 | 81.69264 | near Smt. Kattipalli Padma Suresh bridge |
| Sand Ramp | Sector 3 | 16.4388 | 81.69347 | Muriki Kaluva meets Godavari |
| S1 | Sector 3 | 16.43968 | 81.69722 | drain outlet from Kristen Peta |
| S6 | Sector 3 | 16.44488 | 81.7018 | drain outlet close to the railway station |
| S4 | Sector 3 | 16.44734 | 81.70389 | close to the sand ramp |

*Table 8: Details of selected drains*

### 3.2.2 Palacole drainage system

There are two major drains (MD 1 and MD 2) in Palacole, which carry the entire sewage of city and surrounding areas of the municipality. There are six small drains (sub drains 1 to 6) which carry waste water into these major drains. The major drains are often filled with waste such as plastic, leaves, clothes, paper etc., obstructing the flow. These drains are very long in length and the flow is limited due to their zigzag pattern. The major drains finally join the canal coming from upstream of the Palacole. This canal travels for one kilometer before merging with the River Godavari. In the Palacole drain map below, MD 1 Starting, MD 1 Middle and MD 1 Ending mark the starting, middle and end points of Major Drain 1 . MD 2 Starting, MD 2 Middle and MD 2 Ending mark the starting, middle and end points of Major Drain 2. These are the points where lab samples are collected and sensors are dipped for taking measurements.

*Figure 15: Map of Palacole depicting the drain pattern (in blue) and the six sample collection points along the two main drains*

### 3.2.3 Kovvur drainage system

Kovur is a small town with a population of less than 40,000. It has minor drains that finally merge with the irrigation canal and enter Godavari. Currently five sampling locations (D1,D2, D3, D4 and D5) have been identified for both sensor-based and lab-based monitoring. One location (D6) has been identified for only sensor-based monitoring, marked in yellow on the drain map below. Of these six sites, five sites have been identified for lab sampling.

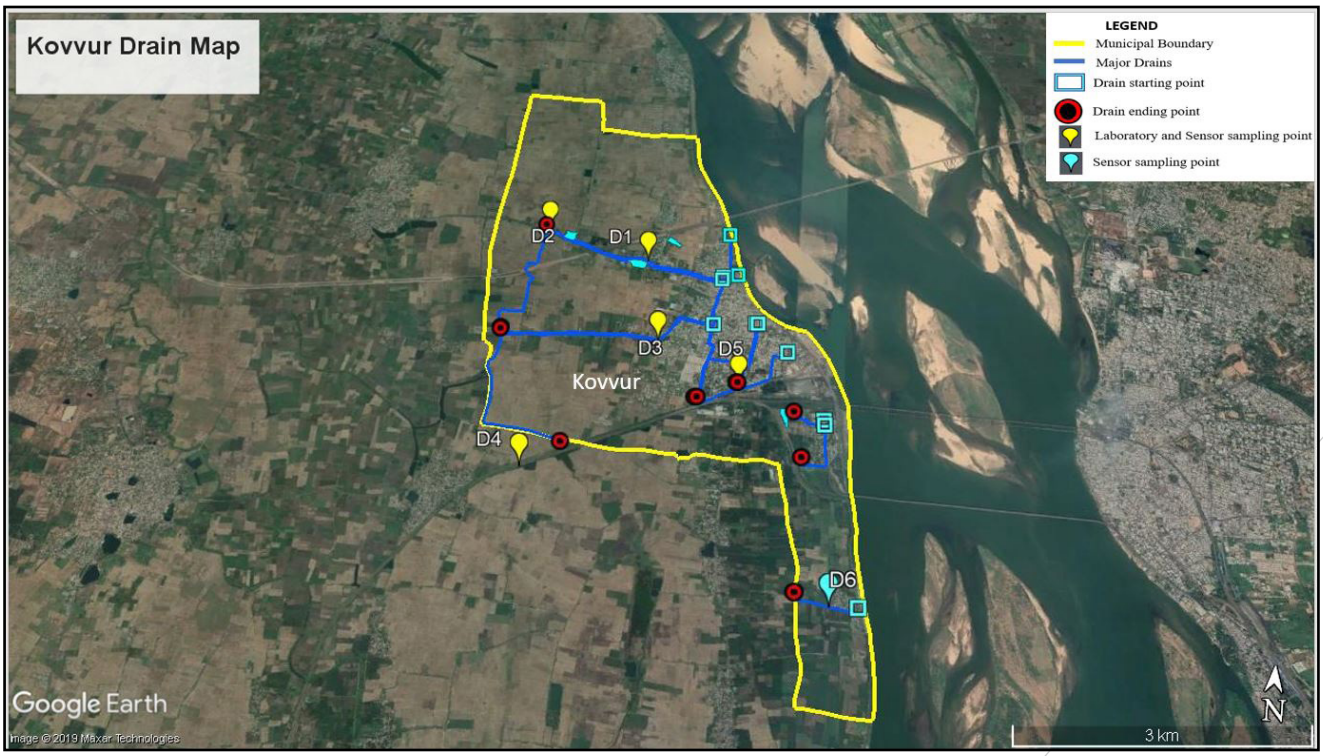



*Figure 16: Map of Kovvur depicting the drain pattern (in blue) and the six sample collection points*

## 3.3 Work Carried Out

### 3.3.1 Phase I: Scoping Study (October 2016 to March 2017)

Project planning

The project started with the aim to develop a cloud enabled water quality sensing network in the Godavari river to explore sensing techniques for rapid measurements that are indicative of anthropogenic contaminants in the water such as those related to human, agricultural and industrial waste. This would help understand how riverfront towns contributed to water pollution, and also understand the impact of BMGF funded septage management interventions led by ASCI across the three towns - Narsapur, Kovvur and Palacole.

*Figure 17: Picture showing Water-to-Cloud team conducting a boat ride along the Narsapur stretch of River Godavari*

### Sensor research

Various water quality monitoring parameters were identified for sensor-based monitoring in real time: pH, turbidity, dissolved oxygen, electrical conductivity, nitrate, ammonia, chlorophyll, tryptophan fluorescence, etc. A market research study was carried out to find off-the-shelf commercial sensors to measure these parameters. A multi-parameter sensor from Hanna Instruments, HI 9829, was identified to be the most suitable for selected physio-chemical parameters. The fluorescence or biochemical parameters were added later.

### Getting started

A field team was set up and requisite permissions were sought from local authorities in Kovvur to carry out the monitoring. Reconnaissance visits were made to Kovvur and Rajahmundry to map points where waste water from these towns entered Godavari. After studying entry points of drains, industrial sites such as International Paper APPM Ltd. in Rajahmundry, the discharge point of the drains leaving the sewage treatment plant, etc., a tentative route map was created for boat ride experiments.

## 3.3.2 Phase II: Piloting Water-to-Cloud System (April 2017 to September 2017)

### Pilot study

A team comprising a water chemist and a sensor engineer began working on the project full-time. Upon procurement of the sensors, lab calibration and on-field calibrations were carried out. A suitable boat and boat-person were identified. This technique was then deployed on Godavari in Rajahmundry. Initial troubleshooting helped fine tune the methodology in terms of how the sensors should be attached to the boat, fixing them to avoid incorrect values, holding them tight at a particular depth, etc.

### Sensor validation

To ensure the accuracy of sensor-based measurements, calibration protocols were followed (see appendix). Additionally, an independent, government-recognized laboratory was appointed to validate the results. Vimta Labs was chosen after evaluating a number of options Along with regular boat sorties, samples from the field were sent to the lab and all the parameter readings were verified to match with sensor readings. In addition, ex-situ parameters such as BOD, COD, total coliform, fecal coliform, etc. were measured by sending water samples to the lab.

### Replication to Narsapur

After a successful pilot at Rajahmundry-Kovvur stretch of River Godavari, operations were expanded to Narsapur to monitor the waste water leaving the town into Godavari. A similar boat ride technique was deployed to study spread of pollution. Domestic discharges and agricultural run offs were identified as main sources of waste water. There are no major industries in Narsapur. The water quality data was influenced by backwaters from the Bay of Bengal.

## 3.3.3 Phase III: Data Curation and On-ground Implementation (October 2017 to March 2018)

### Web platform for digital curation

While collecting credible data is important, disseminating it in easy-to-understand formats is crucial too. Decision makers may not understand technical data presented in tabular formats. However, giving a color code to each numerical value and creating data visualizations in the form of heatmaps by superimposing the geo-tagged water quality data makes interpretation easier. In order to democratize data access on pollution spread in the river, a full-time web developer was engaged to create a website (http://thoreau.uchicago.edu/water/home) to curate heat maps generated after each boat ride. This website presents interactive data visualizations allowing users the choice of site, date and parameter to analyze pollution spread in river through heatmaps.

### Drain monitoring in Narsapur

A need was felt to move from monitoring River Godavari to monitoring the drains that are responsible for the waste water entering the river. Narsapur was identified as one of BMGF's 10 model sanitation towns globally. It is located at the banks of River Godavari, 20 km before the river enters Bay of Bengal. Concentrated efforts were made to monitor waste water in drains and canals of Narsapur before the waste water reaches the river. ASCI and W2C started with monitoring drain water in Narsapur and prepared a detailed drain map. With the support of local authorities, major drains were identified based on the amount of waste load being carried in each drain. A baseline study was conducted to evaluate water quality of Narsapur's drains under the guidance of former Senior Advisor to Andhra Pradesh Pollution Control Board, who was hired as a consultant on this project. The results were discussed with experts at ASCI to plan for regular monitoring.

### Additional sensor procurement

While the Hanna HI-9829 sensor platform monitors physico-chemical parameters, a turners C3 fluorometer sensor platform has been procured to measure chlorophyll, CDOM fluorescence and tryptophan fluorescence. These provide an indication of the organic content in any water body. For example, tryptophan is an amino acid present in pathogens. Higher the tryptophan in water, higher the organic content, which would mean potentially higher BOD values. This is a relatively new approach but shows the promise of reducing BOD test time from a few days in lab to instantaneously in the field. In addition, during review meetings with various stakeholders, a need was felt to monitor the quantity along with the quality of waste water generated. A flow probe FP211 was thus purchased to measure the pollution load.

### 3.3.4 Phase IV: Data Analysis (April 2018 to September 2018)

**Using machine learning to establish correlations**

The team explored machine learning models to establish correlation between ex-situ parameters such as BOD and COD with in-situ parameters such as pH, EC, DO and turbidity. Various techniques and models were used to establish these correlations such as Simple Linear Regression, Multiple Linear Regression, Support vector machine (SVM), Adaboost and Artificial Neural Network (ANN). The results of using machine learning are discussed in later sections.

**Exploring real-time 24x7 monitoring options for physical/organic parameters**

Along with collecting sensor-based data from the River Godavari and regular monitoring of over 13 drains in Narsapur, the project has also been monitoring DO and turbidity at two critical sites all through the day. The idea was to monitor diurnal variations in parameters in peak and lean hours of the day. To comprehensively study the quality of waste water leaving Narsapur, we explored 24x7 real-time monitoring stationary sensors that can measure various physical, physico-chemical and organic parameters such as DO, turbidity, BOD and COD. Various national and international vendors were contacted to assess suitable options. DO and turbidity sensors were procured from a Japan-based company, Optex. BOD and COD are parameters that are measured in the laboratory with no real-time sensors available in the market yet. BOD takes upto five days to be measured in the lab. Surrogate parameters such as fluorescence and ultraviolet (UV) absorbance were explored. Multiple consultations were held with vendors, their customers and experts from ASCI following which a UK-based vendor was finalized to measure BOD in real time based on UV absorbance at 254 nm which captures the signature for organic content. An understanding of organic content at different locations on the drain could help in identifying proximity to insanitary toilets.

**Pollution load and data analysis**

Analysis of water quality data has been discussed in detail in Unit 5 and 6. Several experiments were conducted to understand the pollution load in terms of kilograms of COD and BOD contributed by each drain per day. Width and depth of each drain were measured. Using a flowmeter, the flow in each drain was measured at a 2-hour interval across the day to come up with the average flow rate for it. The drain dimensions and flow were used to calculate the total waste water leaving Narsapur, which is estimated to be ~9 MLD. The volume of waste water in each drain was multiplied with BOD and COD data to calculate the BOD and COD pollution load of each drain. The F.Coli levels helped pinpoint the hotspots of insanitary waste along the drains for ASCI to identify relevant interventions. The seasonal trends of various water quality parameters were identified. Water quality profile of each drain was analyzed and compared to draw insights on required interventions. The W2C team also provided the waste water quality data to support ASCI assess the need for setting up treatment plant in Narsapur.

### 3.3.5 Phase V: Scaling Up, 24x7 Monitoring (October 2018 to March 2019)

#### Operationalizing real-time stationary sensors

Real-time DO and Turbidity sensors from Optex, a Japanese company, have been installed at two critical drain locations in Narsapur to measure water quality every 5 minutes. The Optex sensors are tuned to display data in real time but do not transmit it to a cloud storage network. To make wireless transmission possible, the W2C team has worked with Advance Tech Pvt. Ltd. to develop a data logger to store values in a flash drive and transmit data to a cloud server using GSM technology. Two sensors have been installed in Narsapur-one on the drain, D1 passing through the Water Works department and the other in Muriki Kaluva, a major drain passing through the slum area. These locations were strategically chosen because together they carry nearly 40% of the town's waste to Godavari. This helped understand the pollution trend of drains throughout the day. An online dashboard was developed to visualize data generated from the stationary real-time sensors. The dashboard depicts temperature, turbidity and dissolved oxygen level in graphical and tabular formats. This can be accessed any time to check live status of water quality and understand daily, monthly or seasonal trends.

#### Scaling up to Palacole and Kovvur

Based on the success of drain water monitoring study in Narsapur, ASCI requested the W2C team to expand operations to Kovur and Palacole. Major drains were identified in both towns with the support of S3 team. The drain monitoring protocol was replicated at the new sites to collect regular data on drain water quality in these towns. Regular operations for drain monitoring have continued since then in all three towns.

#### Identification of Effluent Treatment Plant (ETP) site in Narsapur

Drain pollution data was analyzed to assess the need for an Effluent Treatment Plant (ETP) in Narsapur. Each drain was studied based on its pollution level and load. Drain D7, connected to the local slaughterhouse, was identified as a potential site based on high pollution load in the drain. The analysis was presented to ASCI to support their idea of setting up an ETP in concurrence with the local authorities.

*Figure 18: Picture showing waste water released near a slaughterhouse in Narsapur into a drain that finally meets River Godavari*

### 3.3.6 Phase VI: Wrapping up (April 2019 to September 2019)

#### Dashboard for drains

A drains dashboard was created for Narsapur to depict all the waste water quality parameters measured in the last couple of years. This is independent of the heatmaps from the boat sorties that are already curated on a regular basis. This dashboard enables comparison between different drains on various parameters to understand the pollution loading patterns in the town and help identify drains connected to insanitary toilets or poor sanitation facilities. A slider allows the user to move between various dates of data collection to observe temporal trends in waste water quality for each drain. The dashboard can be accessed at the Thoreau website.

#### Operationalizing UV-based BOD sensor

Real-time BOD and COD sensor has been procured from Photonic Measurements, a UK based company. It has been operationalized to measure real-time values of BOD, COD and TOC in rivers and drains. This sensor is dipped at identified locations on the drains to get real-time values. An orientation session was conducted by the sensor designer to train the team on calibration, handling and maintenance of the sensor. The sensor values were validated by independent laboratories. The readings were found to be highly unreliable based on limited data obtained so far and the sensors have now been sent back to the manufacturer for repair.

#### Wrapping up

For a smooth handover to ASCI, all data from various sensor and lab-based measurements in the river, drains and canals have been compiled to be handed over to ASCI. The data also remains available on the Thoreau web platform in raw form as well as in the form of heatmaps and summary statistics. The field team from S3 will be briefed about the operations and a knowledge transfer session can be held for the capacity building of local teams upon ASCI's request. The tech platform built can be made available to the town for future data visualization needs for water pollution data upon a written request made to W2C by the town authorities.

# 4 Research Findings

The research findings in this section are based on the following data points collected between April 2017 and June 2019.

| | Narsapur | Palacole | Kovur |
|---|---|---|---|
| # Drains monitored | 13 | 6 | 6 |
| # Lab samples (drains) | 163 | 54 | 44 |
| # Sensor data points (drains) | 626 | 110 | 159 |
| Area Surveyed in Sq. Km (River) | 3.2 | N/A | 11 |
| # Boat rides (River) | 68 | N/A | 93 |
| # Lab samples (River) | 107 | N/A | 112 |
| # Sensor data points (River) | 66039 | N/A | 130998 |

*Table 9: Summary of data collected for Water-to-Cloud*

## 4.1 Pollution Hotspots Identification

### 4.1.1 Major Drains in a Town

A spike in EC values was observed at the meeting point of Ava drain with River Godavari in Rajahmundry. Sewage water from this drain is seen to be released into the river just ahead of the agriculture canal at Dowleswaram (see figure 19). In Narsapur, most of the town's domestic waste reaches Godavari through channel 2 marked on the map. Channel 2 is connected to Muriki Kaluva drain which carries 40 percent of Narsapur's waste water. CDOM heatmap shows an abrupt spike around this area due to the waste water mixing with river waters (see figure 20). From the heatmaps shown below, the transition from green to red zones around a major drain (Ava drain in Rajahmundry and Channel 2 in Narsapur) occurs in less than 50 m. This shows discharge of waste by these drains and its impact on Godavari.

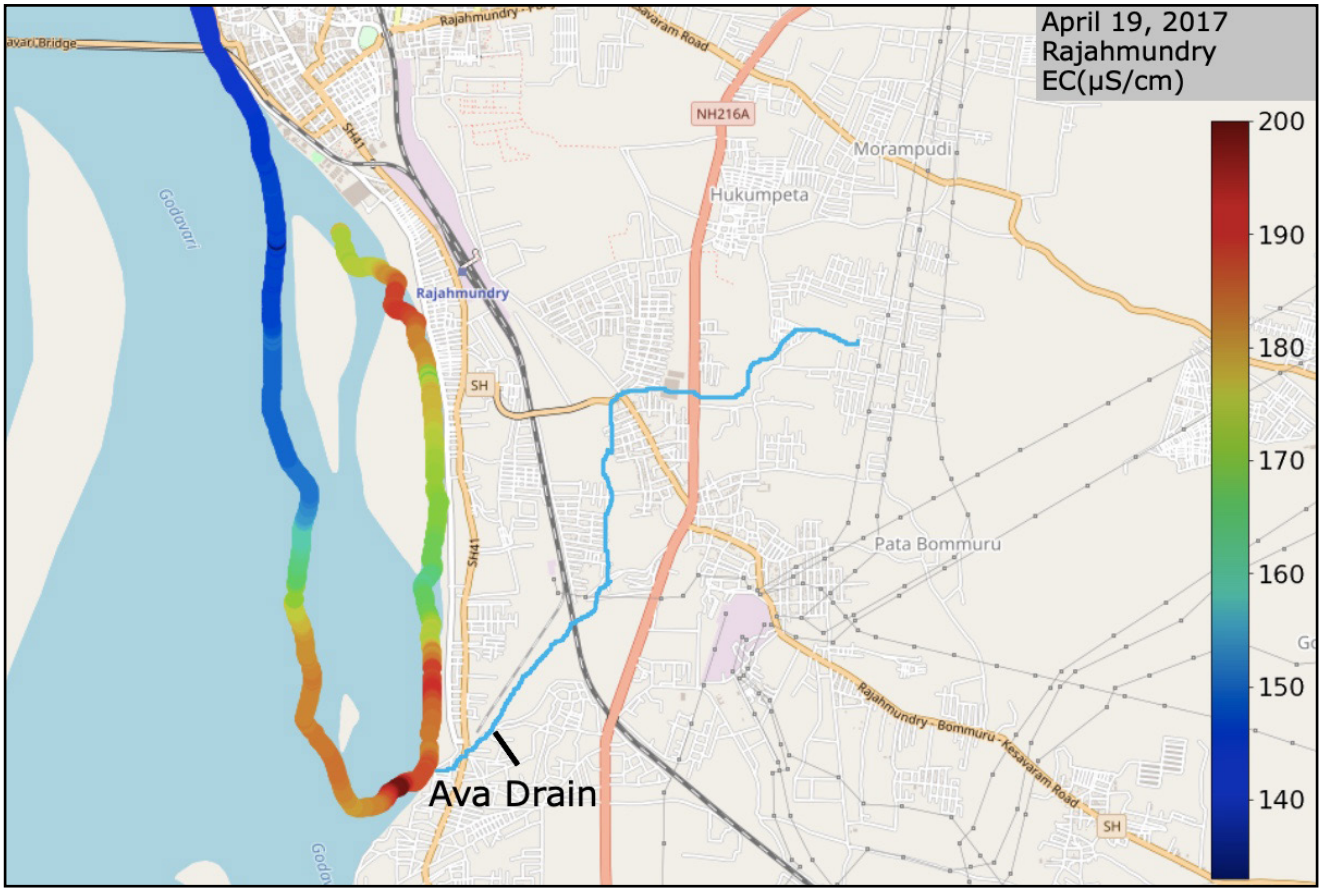



*Figure 19: Heatmap tracing the boat trajectory along Rajahmundry stretch of River Godavari to depict an increase in electrical conductivity (µS/cm) around the area where Ava drain, Rajahmundry's biggest waste water drain, meets River Godavari*

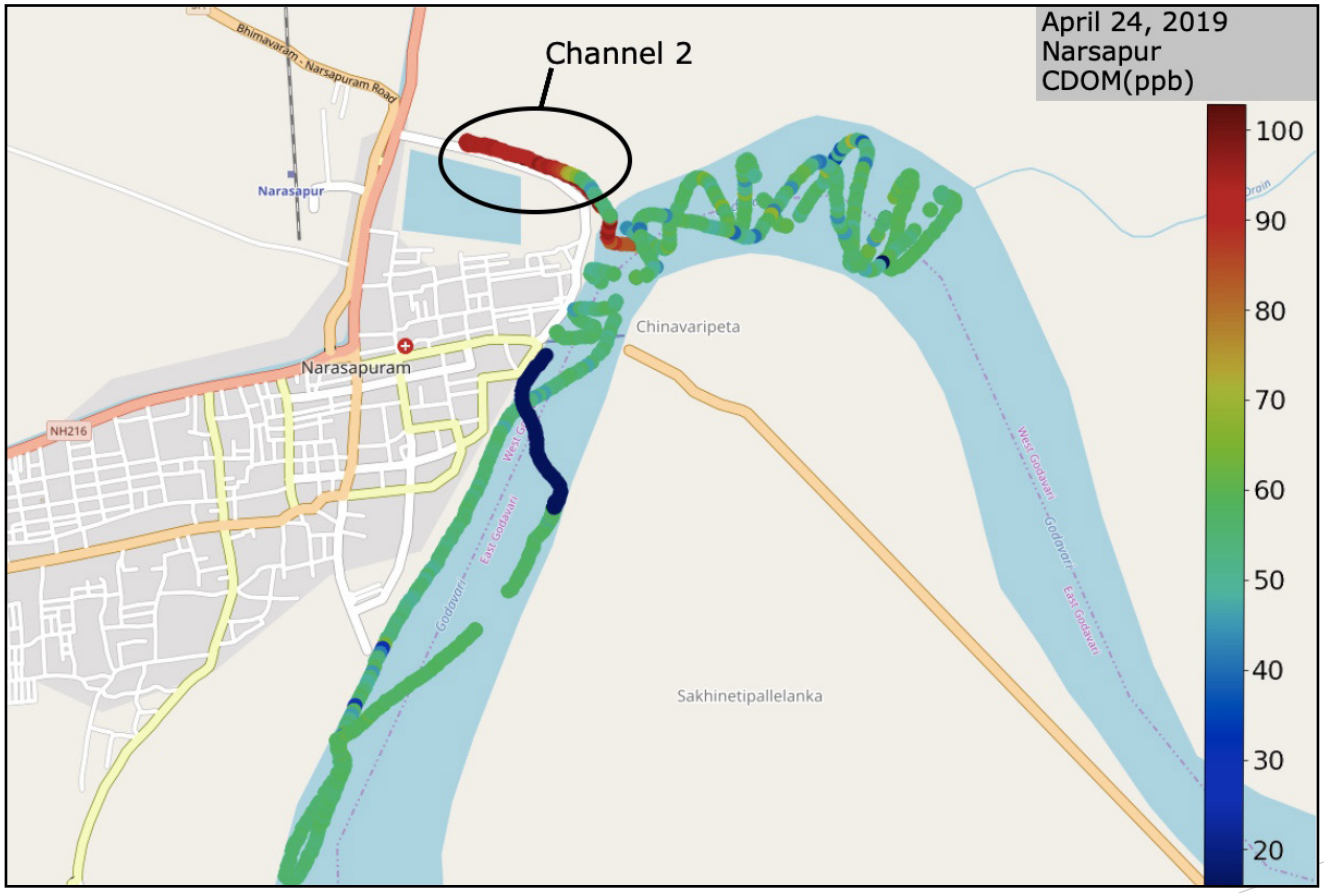



*Figure 20: Heatmap tracing the boat trajectory along Narsapur stretch of River Godavari to depict a steep increase in CDOM (ppb) as the waste water from channel 2 meets River Godavari*

### 4.1.2 Industrial Discharge

Sudden rise in pH values near International APPM Ltd – a paper mill in Rajahmundry, is indicative of contaminants possibly being dumped from the paper industry. EC plots showed a similar spike.

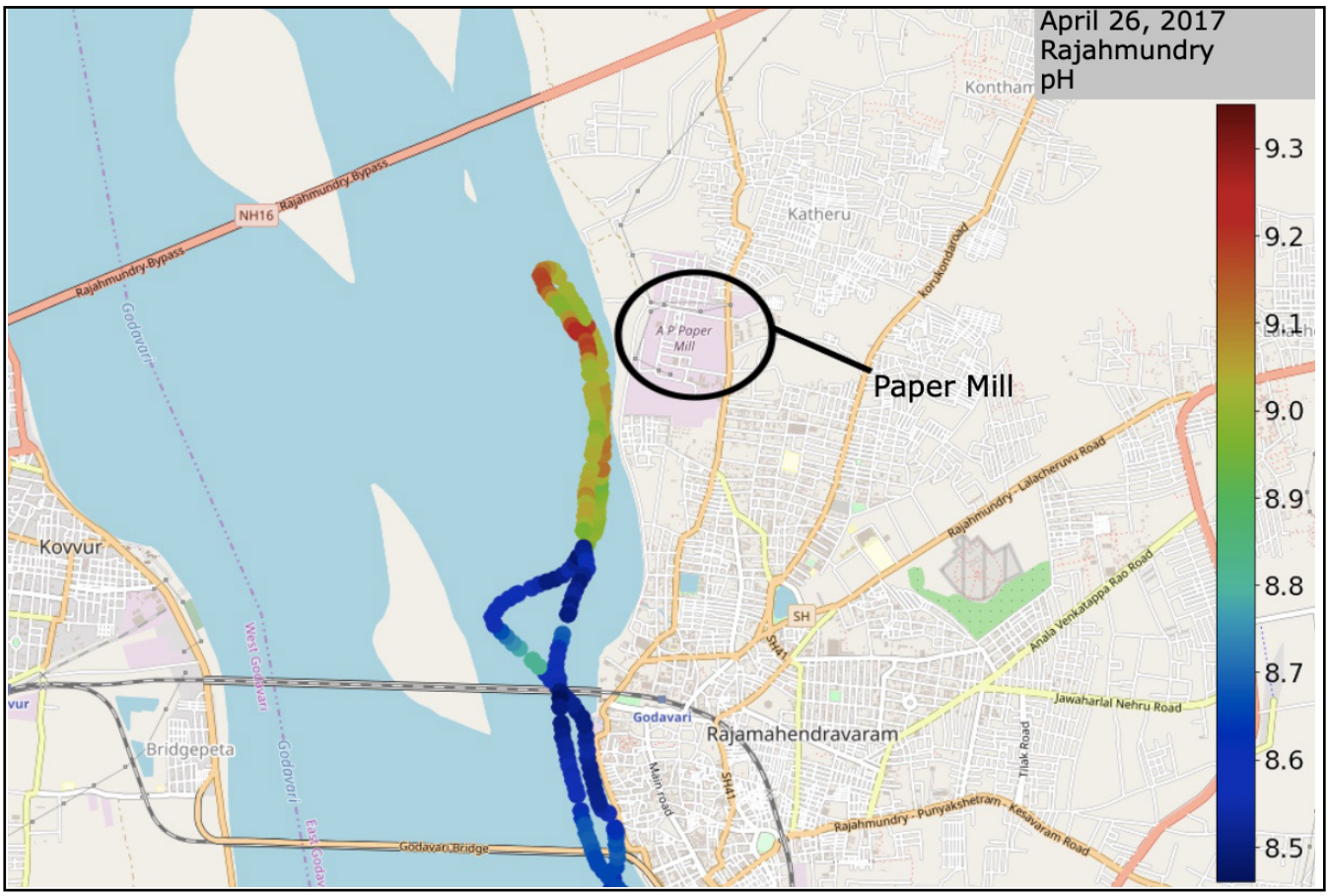



*Figure 21: Heatmap tracing the boat trajectory along Rajahmundry stretch of River Godavari shows a sudden increase in pH near the paper mill*

### 4.1.3 Leaching Effects

The solid waste dumping site at Narsapur is on the river bank. A huge heap of waste is seen at this site. It is known to accumulate waste over the last few years. Leaching is the phenomenon by which harmful chemicals from landfills formed from dumping of waste dissolve and migrate into natural resources like river, groundwater, etc[15]

A spike in EC observed at the solid waste dumping site in Narsapur indicates an occurrence of leaching which impacts the river water quality. Such chemical contamination has long-term harmful effects. Proper designing of the landfill site will prevent this.

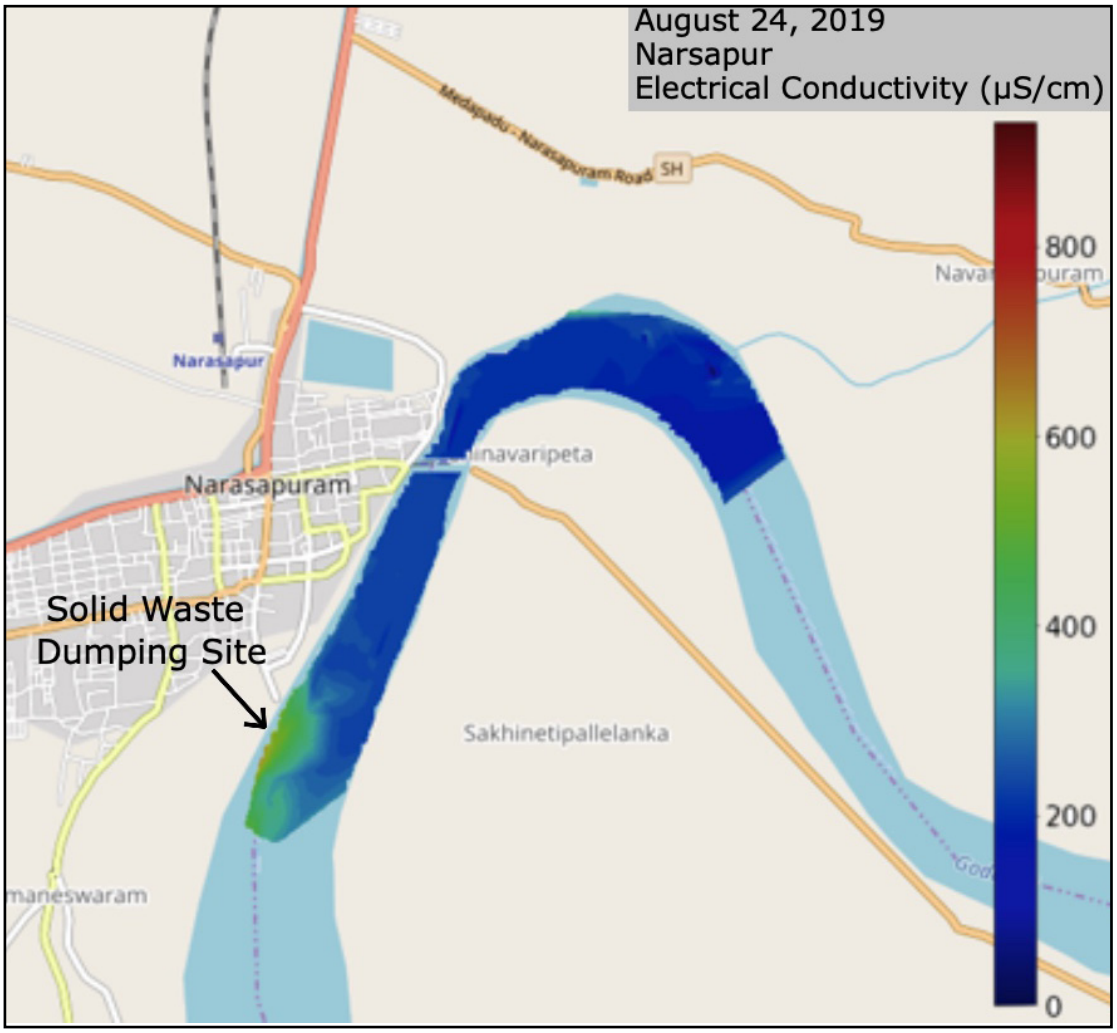



*Figure 22: Heatmap capturing the leaching effect next to a solid waste dumping site along Narsapur stretch of River Godavari depicted through a steep increase in electrical conductivity (µS/cm)*

15 IIT-Guwahati. (n.d.). Retrieved from http://www.iitg.ac.in/coeiitg/lecture22.pdf

## 4.2 Direct and Indirect Usage of Water

### 4.2.1 Suitable for Bathing

The table below represents classification of water quality based on the water quality index as specified by CPCB. Class A water is safe for drinking while Class B is suitable for bathing.

| WQI | Quality classification | Class by CPCB |
|---|---|---|
| 65-100 | Good to excellent | A |
| 50-65 | Medium to good | B |
| 30-50 | Bad | C |
| <=30 | Bad to very bad | D, E |

*Table 10: Water Quality Classification*

An analysis of Water Quality Index (WQI) in Godavari at Rajahmundry-Kovvur stretch using the sensor data gathered from April 2017 to April 2019 from four locations with varying characteristics was carried out. 103 sensor-based water quality measurements are used to create WQI. pH, EC, DO and turbidity were used. The following equation was employed to calculate WQI as suggested by Wunderlin[16].

$$WQI_{sens}=(C_{pH}+C_{EC}+C_{DO}+C_{turbidity})/4$$

Here $C_{pH}$, $C_{EC}$, $C_{DO}$ and $C_{turbidity}$ represent the normalization factor for each of these parameters. Normalization factor is value to convert each parameter to 0 - 100% scale depending on its analytical value, where 100 represents the highest value. The table below shows the normalizing factors for the respective parameters.

The C values are detailed in the table below.

| Parameter | Relative weight (pi) | Normalization factor (C) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 100 | 90 | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0 |
| Conductivity (µS/cm) | | <750 | <1000 | <1250 | <1500 | <2000 | <2500 | <3000 | <5000 | <8000 | ≤12000 | >12000 |
| DO | 4 | ≥7.5 | >7 | >6.5 | >6 | >5 | >4 | >3.5 | >3 | >2 | ≥1.0 | <1.0 |
| pH | 1 | 7 | 7-8 | 8-8.5 | 8.5-9 | 6.5-7 | 6-6.5 | 5.5-6 | 4-5 | 3-4 | 2-3 | 1-2 |
| Turbidity | 2 | <5 | <10 | <15 | <20 | <25 | <30 | <40 | <60 | <80 | ≤100 | ≥100 |

*Table 11: Normalized factor (C) values*

The results show that most of the times water quality in Godavari at Rajahmundry-Kovvur stretch was found to be of medium to good quality, except in one instance. This implies suitability for Class B-bathing water standards. The following figure depicts the graphical representation of the results.

16 Wunderlin, S. F. (2000). Use of water quality indices to verify the impact of Coardoba city (Argentina) On SUQUõ A. Wat. Res. Vol. 34, No. 11, 2915 - 2926.

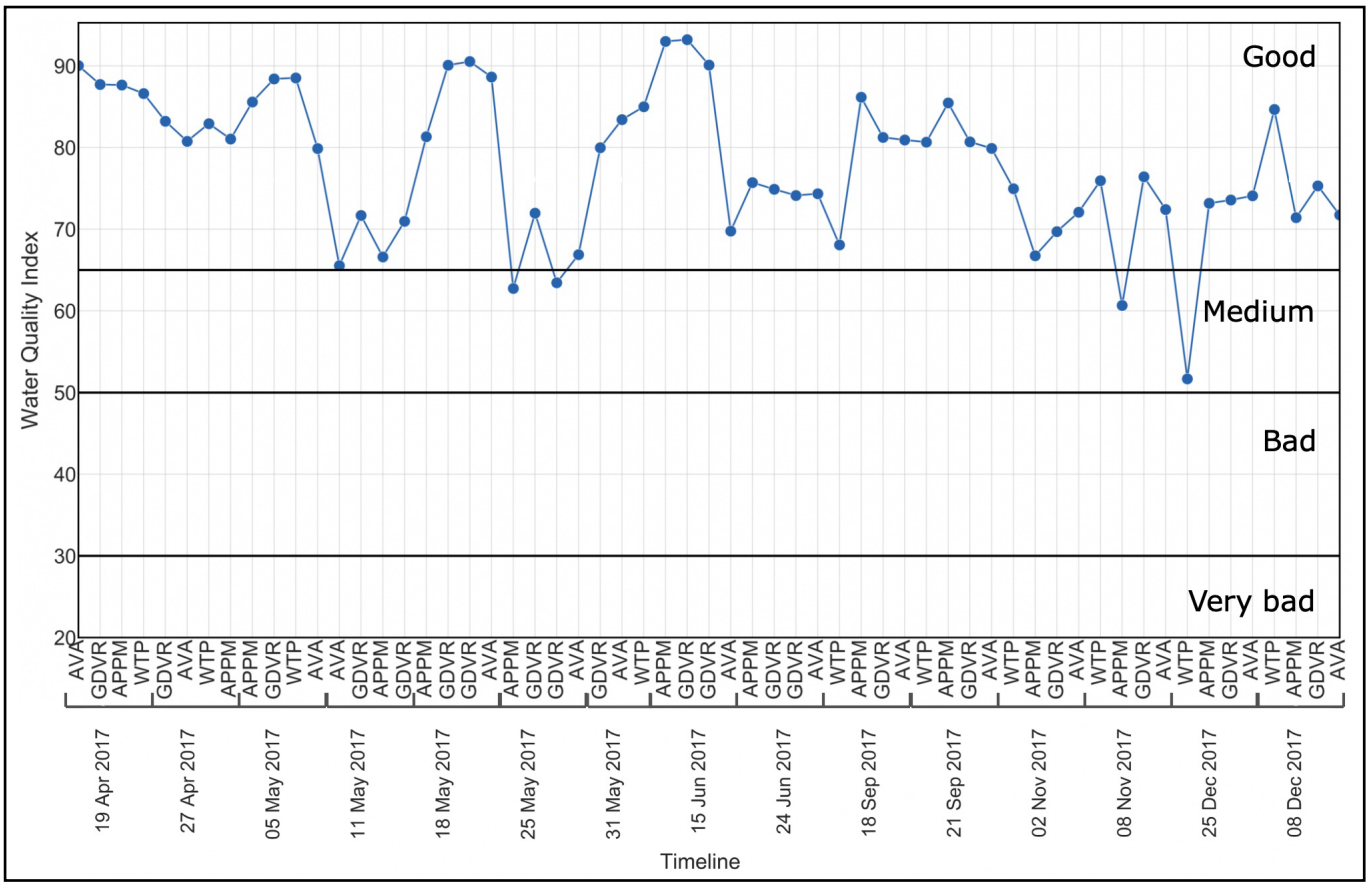



*Figure 23: Water quality index plot of Rajahmundry stretch of River Godavari calculated using sensor parameter (EC, DO, pH, turbidity) values measured at four points with distinct water quality charachteristics: (i) meeting point of Ava waste water drain and river (AVA), (ii) middle of the river (GDVR), (iii) Near paper mill (APPM) and (iv) near Water Treatment Plant (WTP), noted on X-axis, between April 19, 2017 and December 8, 2017*

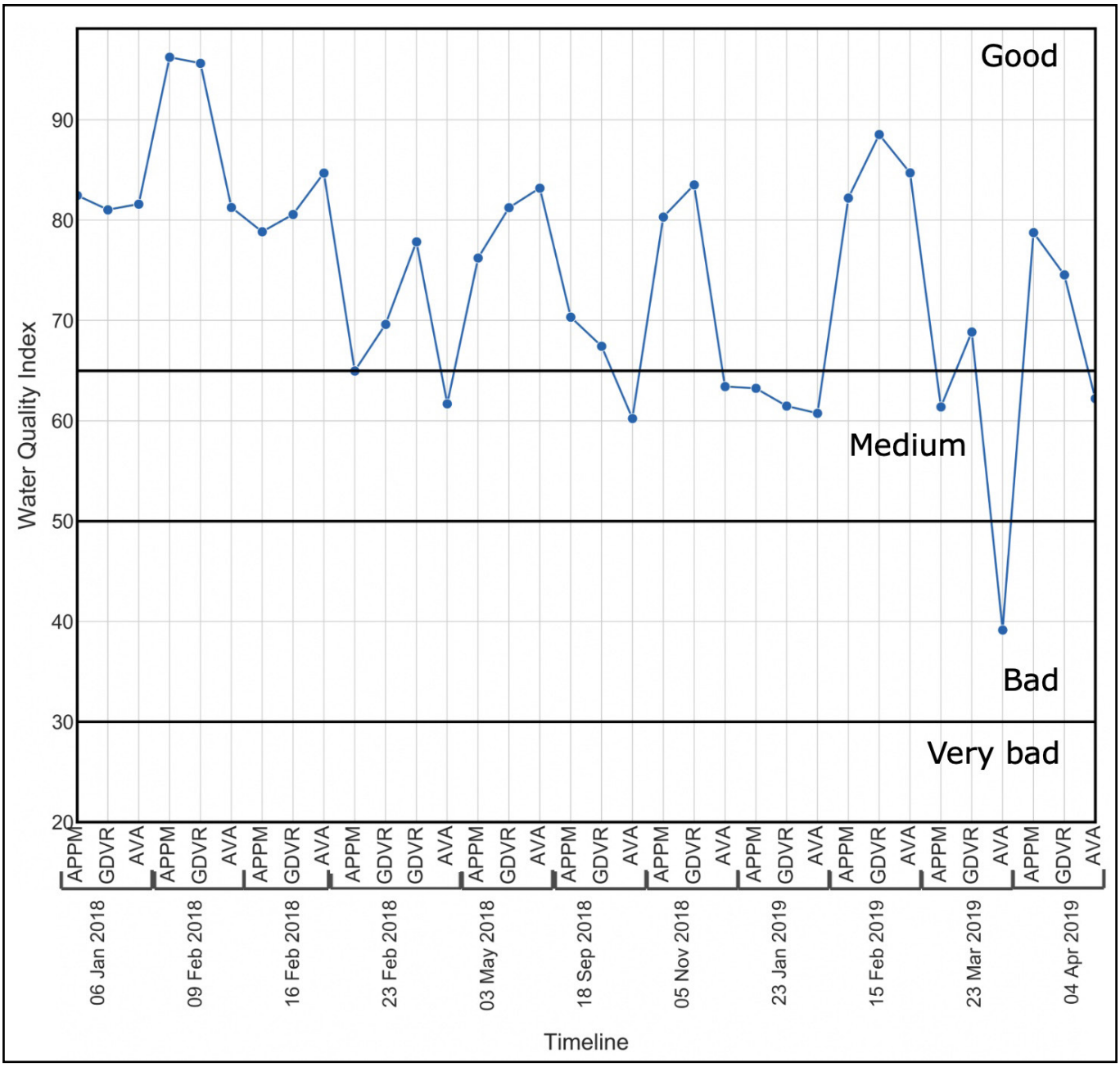



*Figure 24: Water quality index plot of Rajahmundry stretch of River Godavari calculated using sensor parameter (EC, DO, pH, turbidity) values measured at four points with distinct water quality charachteristics: (i) meeting point of Ava waste water drain and river (AVA), (ii) middle of the river (GDVR), and (iii) Near paper mill (APPM) noted on X-axis, between January 6, 2018 and April 4, 2019*

### 4.2.2 Unsafe for Drinking Sometimes

As seen from WQI graph, Godavari waters may not always match drinking water standards. Furthermore, when heavy metal analysis was carried out using an ICP-MS (Inductively Coupled Plasma Mass Spectrometry) instrument in laboratory, several parameters were found to be above the permissible limits in some months of the year. For Narsapur, the concentration of aluminium (Al), chromium (Cr), manganese (Mn), iron (Fe), nickel (Ni), copper (Cu), zinc (Zn), selenium (Se) and lead (Pb) is found to be above the permissible limit as per the Bureau of Indian Standards (BIS 2014). For Rajahmundry, the concentration of Al, Fe, Mn, Ni and Pb was found above the BIS limit. The data for heavy metals exceeding limits is available in appendix VIII. The collected samples indicate that Mn concentrations were higher than the permissible limits in one out of three cases observed in Narsapur. Fe values were above the BIS limits for about 20 percent of the samples in Rajahmundry and for about 69.5 percent of the samples in Narsapur. Similar alarming values were found for Cu and Pb.

High concentration of Mn in drinking water may cause neurodegenerative disorder and sideroblastic anaemia. Fe is an essential element for human body but excess of Fe i.e. > 300ppb may cause hemochromatosis. At later stage, it may increase the risk of arthritis, cancer, liver problems, diabetes and heart failure. Similarly, Cu is also an essential element for metabolic process, immune system and formation of red blood cells in human body. However, high concentration of Cu in water may cause cancer and neurological disorder and its deficiency might cause anaemia or leukopenia. High concentration of Pb in drinking water might cause convulsions, major neurological damage and organ failure in young children.

### 4.2.3 Toxicants Entering Food Chain

High concentration of some heavy metals may affect aquatic life and accumulate in them to enter the food chain. This might lead to another dimension of human health risk. River water is often used for irrigation activities and concentration of ions and heavy metals acts as source of nutrient for the crops. The concentration of heavy metals exceeding the permissible limits might also affect the growth and reduce fertility of the soil. In this study the Godavari river water quality is analyzed using the irrigation water quality provided by Central Water Commission (CWC) 2018 and the concentration of Cu and Zn was found above the standards at downstream in Narsapur. However, all the heavy metals were within the limits at Rajahmundry. High concentration of Cu in irrigation water may restrict the growth of the plants. High concentration of Zn is also toxic to crops at widely varying pH condition. The Zn toxicity was reduced with the increase in pH values.

## 4.3 Evidence based WASH Interventions

### 4.3.1 Locating Insanitary Toilets & Slum Zones

For town authorities to set up targeted interventions such as improvement in sanitation infrastructure, they would benefit from knowing which parts of the town may have a concentration of insanitary toilets. Often, slum dwellers lack proper sanitation facilities and have no access to a toilet or do not use it. Data from drain monitoring correlated with insanitary toilets or slum zones in Narsapur, thus making this an effective method to identify areas which require immediate attention from authorities. The following table shows CPCB limits for discharge into water bodies.

| Parameter | CPCB limits for discharge into water bodies[17] |
|---|---|
| pH | 6.5-9.0 |
| BOD (mg/l) | Not more than 10 |
| COD (mg/l) | Not more than 50 |
| TSS (mg/l) | Not more than 20 |
| NH 4-N (mg/l) | Not more than 5 |
| N-total (mg/l) | Not more than 10 |
| Fecal Coliform (MPN/100ml) | Less than 100 |

*Table 12: CPCB limits for discharge into water bodies*

Fecal coliform (FC) in the waste water mainly results from human or animal excreta. However, the waste from paper and pulp mills, plant materials and agricultural runoff can also increase fecal coliform concentration in waste water. High FC values observed in drains S3, S4, D3, D7 and D14 indicate high load of organic contamination. Refer to Narsapur Drainage System in Unit 3 to see the location of drains. From local authorities, it is known that ward 3 and ward 29 have the bigger slums in Narsapur and also a concentration of insanitary toilets. Drain S4 passes through ward 29 while drain D3 is close to ward 3.

17 CPCB. (2015). CPCB Envis Newsletter on Control of Pollution. Retrieved from http://www.cpcbenvis.nic.in/envis_newsletter/ENVIS%20Newsletter%20Jan%20-%20Apr%202015.pdf

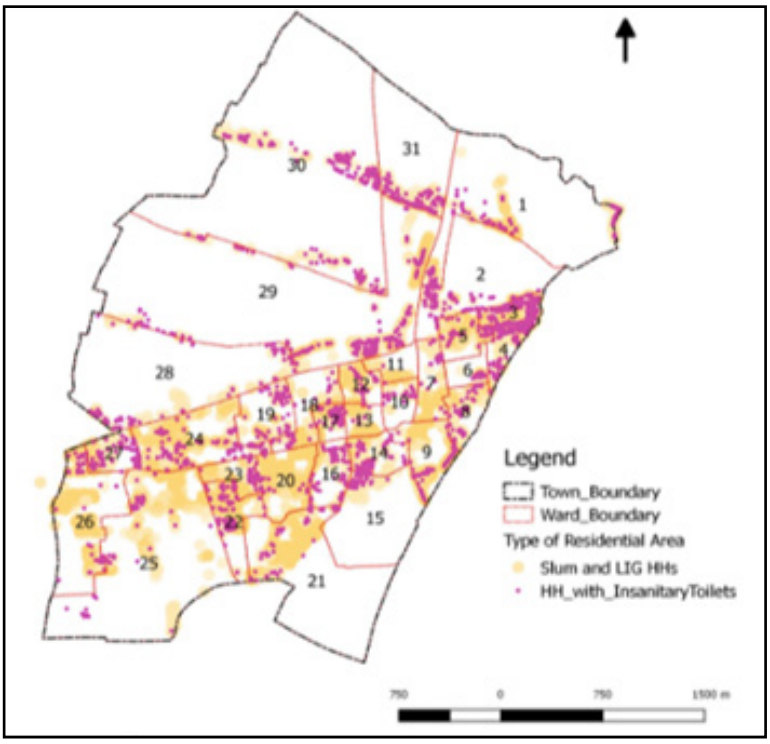



*Figure 25:* Heatmap showing spatial distribution of households with toilets (in green) and without toilets (in red) in Narsapur

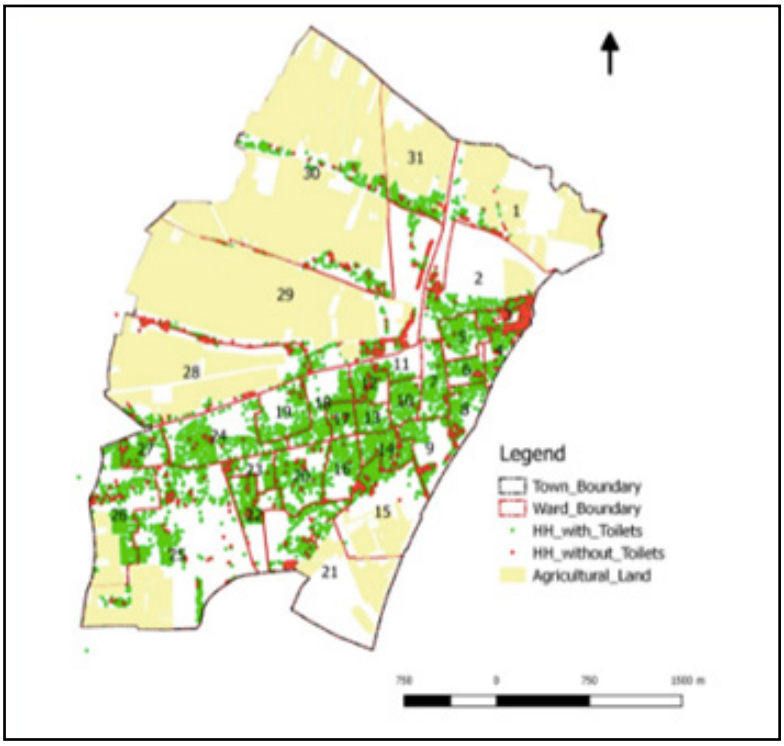



***Figure 26:*** *Heatmap showing spatial distribution of slums (in light mustard) and insanitary toilets (in pink) in Narsapur*

High median FC values in drain D3, S3 and S4 could be because of slums which discharge sewage directly into the drains. In drain D7, the high fecal coliform concentration is due to the waste from the slaughterhouse that enters the drains D14 and D15, which in turn enter Pilla Kaluva. Pilla Kaluva also shows contamination which may affect health of villagers downstream since they use the canal water for drinking. This assessment indicates the need for town authorities to put in place necessary interventions to reduce such contamination.

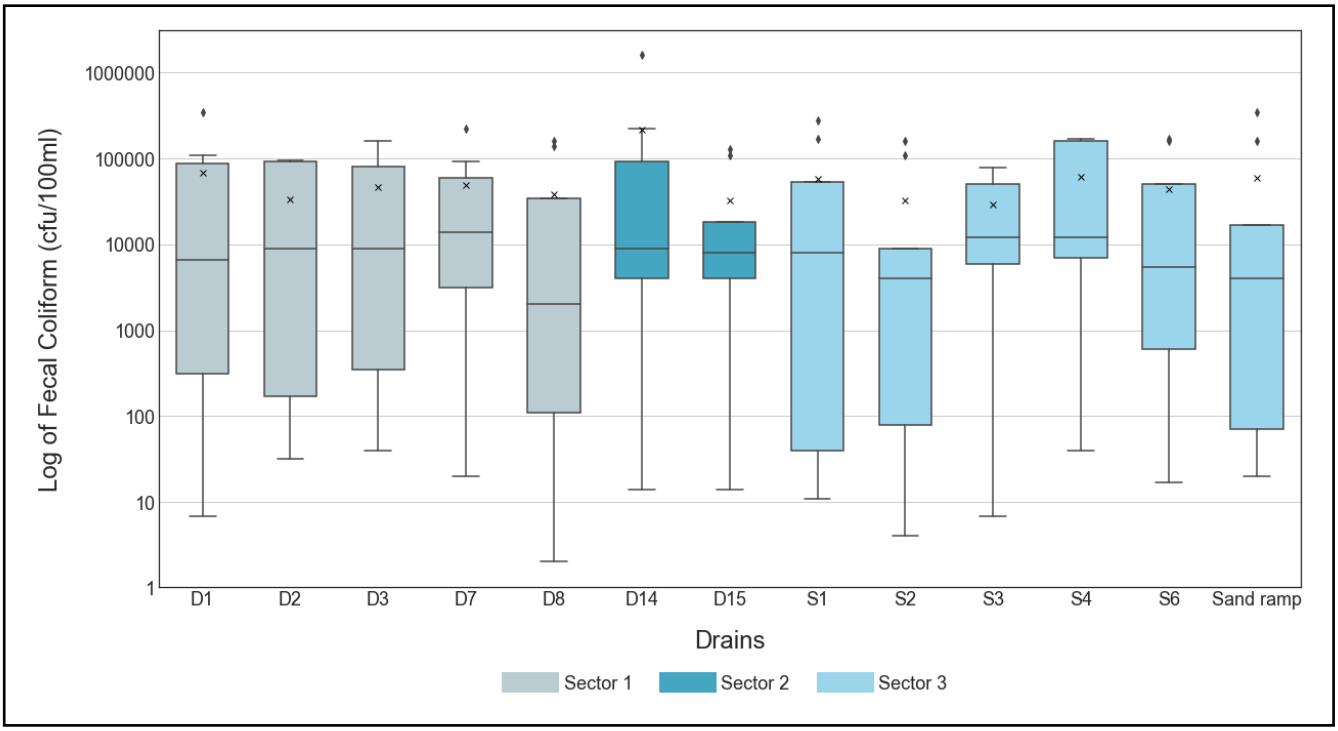



*Figure 27: Box plots showing logarithm of fecal coliform concentation (CFU/100ml) at the 13 sampling points (detailed in figure 14) on major drains in Narsapur*

Similarly in Palacole and Kovvur, the FC numbers are high, indicating human excreta or animal waste mixing directly into these storm water drains. In Kovvur, drain D1 has much higher median FC than other drains indicating the presence of a potential insanitary toilet zone nearby.

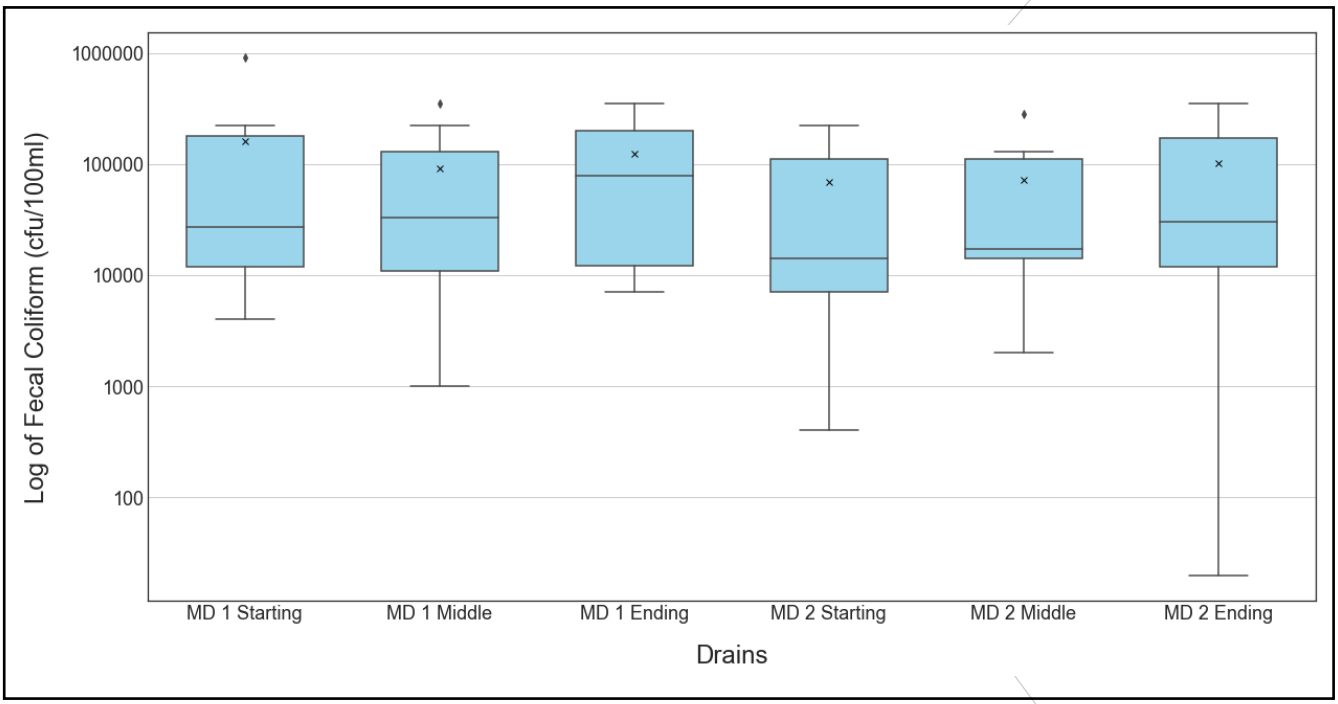



*Figure 28: Box plots showing logarithm of fecal coliform concentation (CFU/100ml) at the starting, middle and ending of two major drains (MD1 and MD2 as described in figure 15) of Palacole*

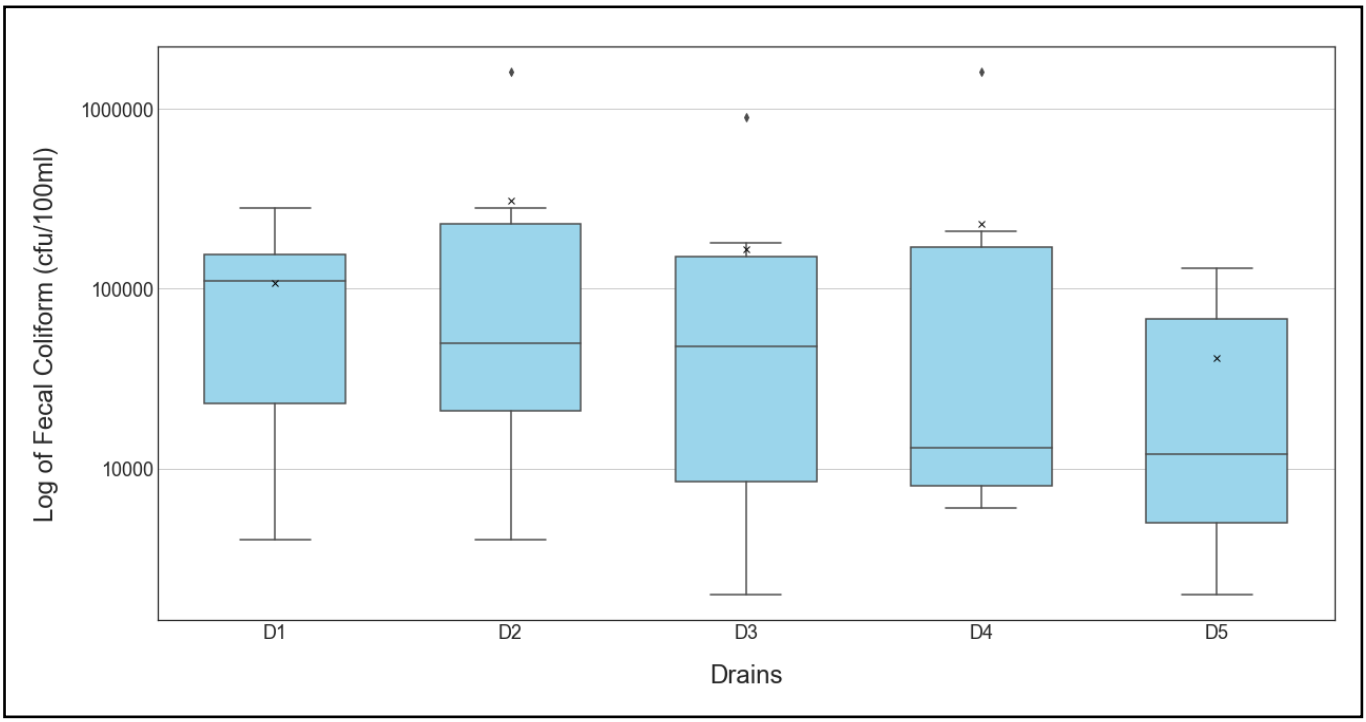



*Figure 29: Box plots showing logarithm of fecal coliform concentration (CFU/100ml) at the five sampling points (D1, D2, D3, D4 and D5 as described in figure 16) on major drains of Kovvur*

### 4.3.2 Case for Treatment Plants

The waste coming from the cities may contain a large amount of organic matter such as food, vegetables, human and animal excreta, paper fibers, plant material, and humus. This high amount of organic matter increases the biological and chemical oxygen demand of water, making its quality unsuitable for human and aquatic life.

The study's findings indicate a need to establish sewage treatment plants in all three towns based on the high BOD and COD concentrations in the drains. A decentralized approach can be adopted since the total waste water generated in these towns is small. The following plots summarize BOD and COD levels in each drain for all three towns.

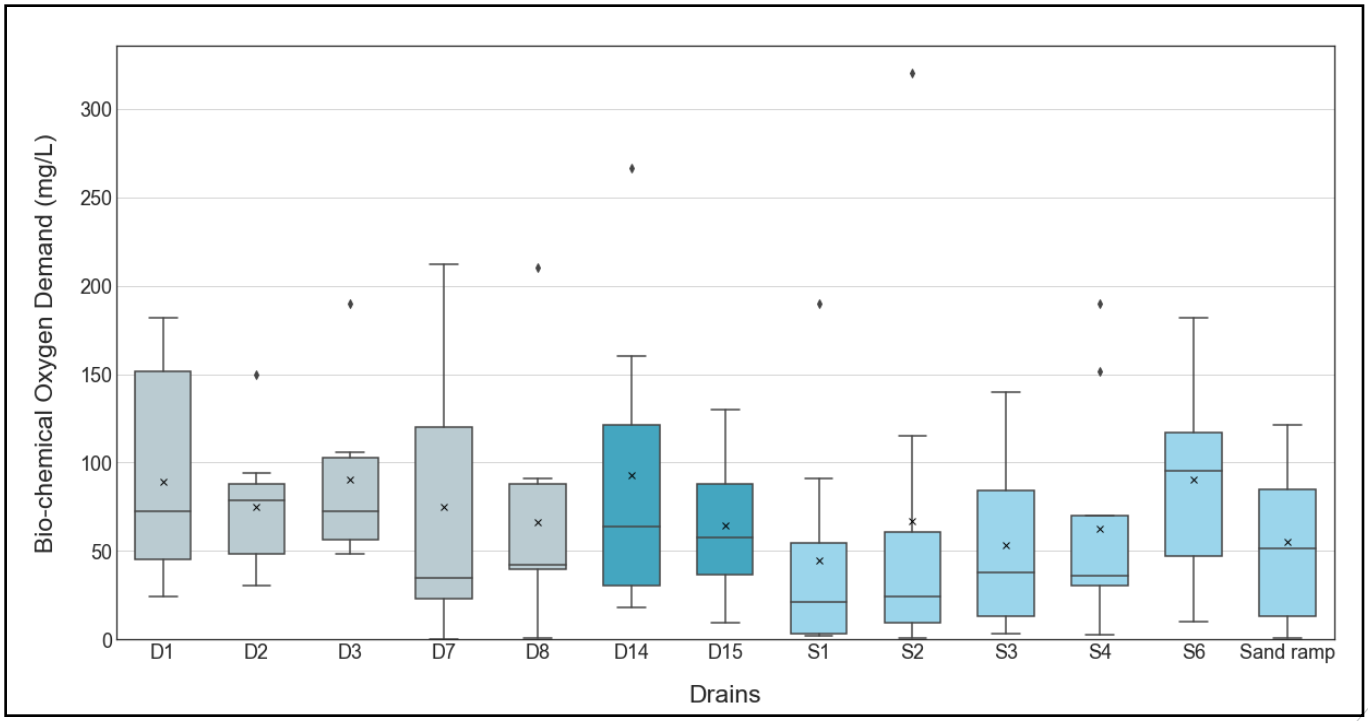



*Figure 30: Box plots showing BOD values (mg/l) at the 13 sampling points (D1, D2,... D15, S1, S2... and S6 as described in figure 14) on major drains of Narsapur*

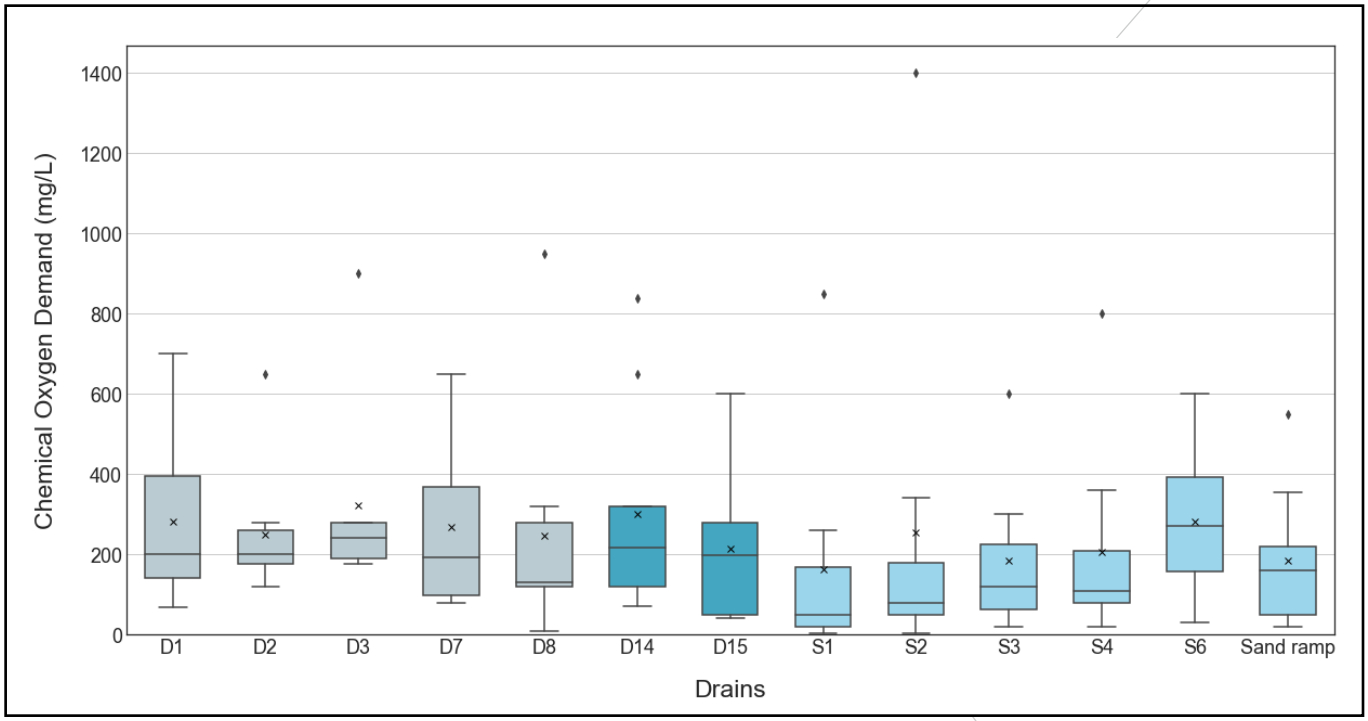



*Figure 31: Box plots showing COD values (mg/l) at the 13 sampling points (D1, D2,... D15, S1, S2...and S6 as described in figure 14) on major drains of Narsapur*

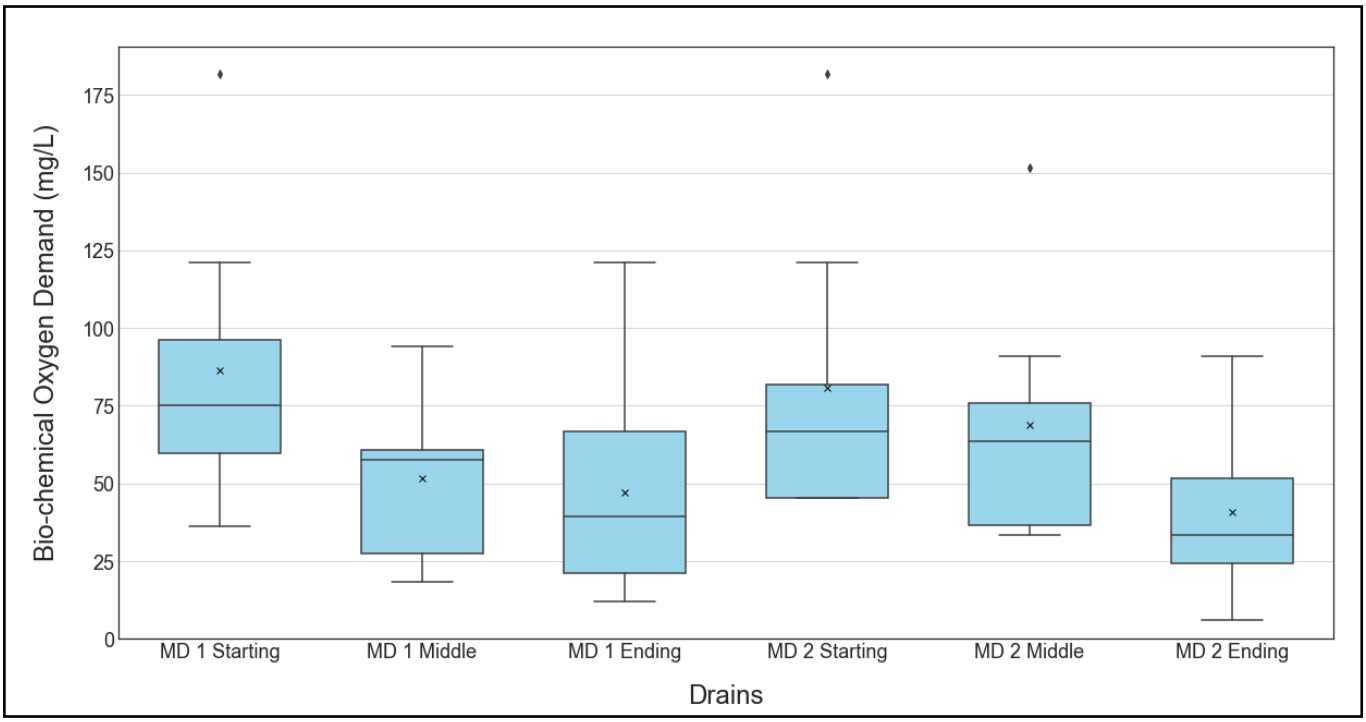



*Figure 32: Box plots showing BOD (mg/l) values at the starting, middle and ending of two main drains (MD1 and MD2 as described in figure 15) of Palacole*

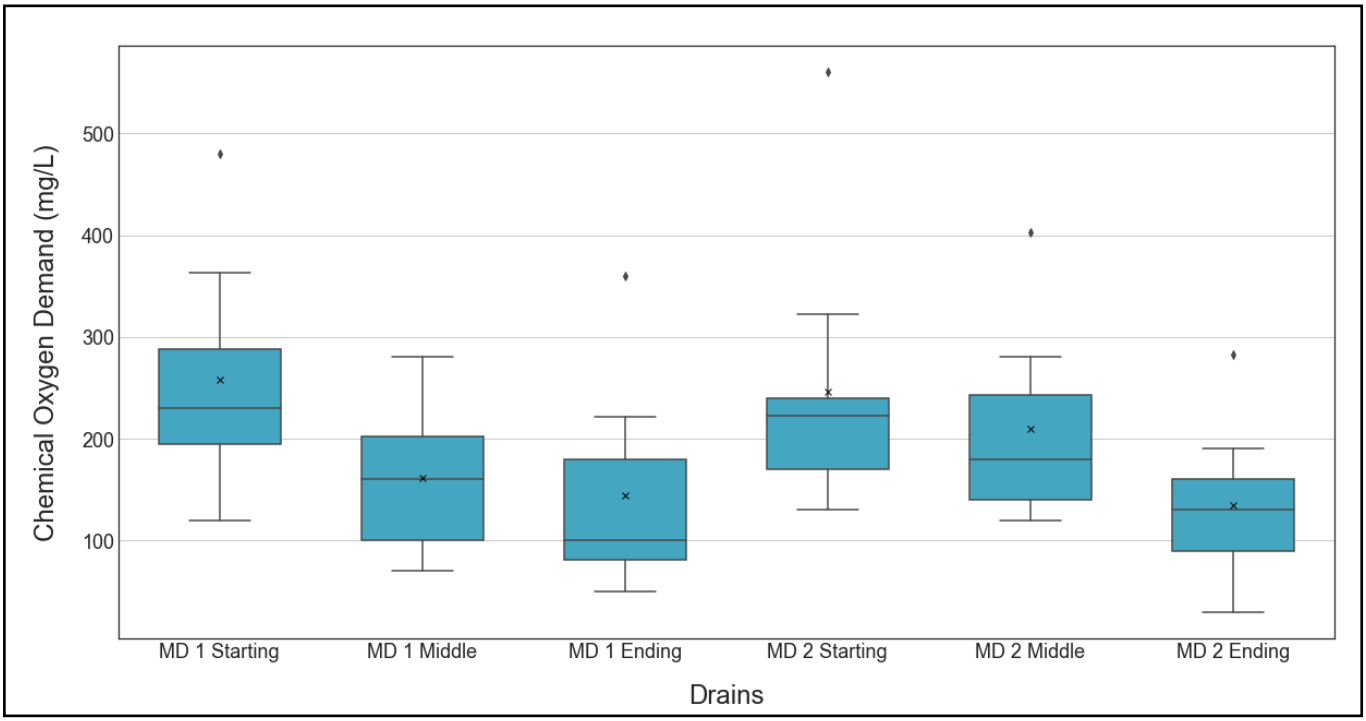



*Figure 33: Box plots showing COD (mg/l) values at the starting, middle and ending of two main drains (MD1 and MD2 as described in figure 15) of Palacole*

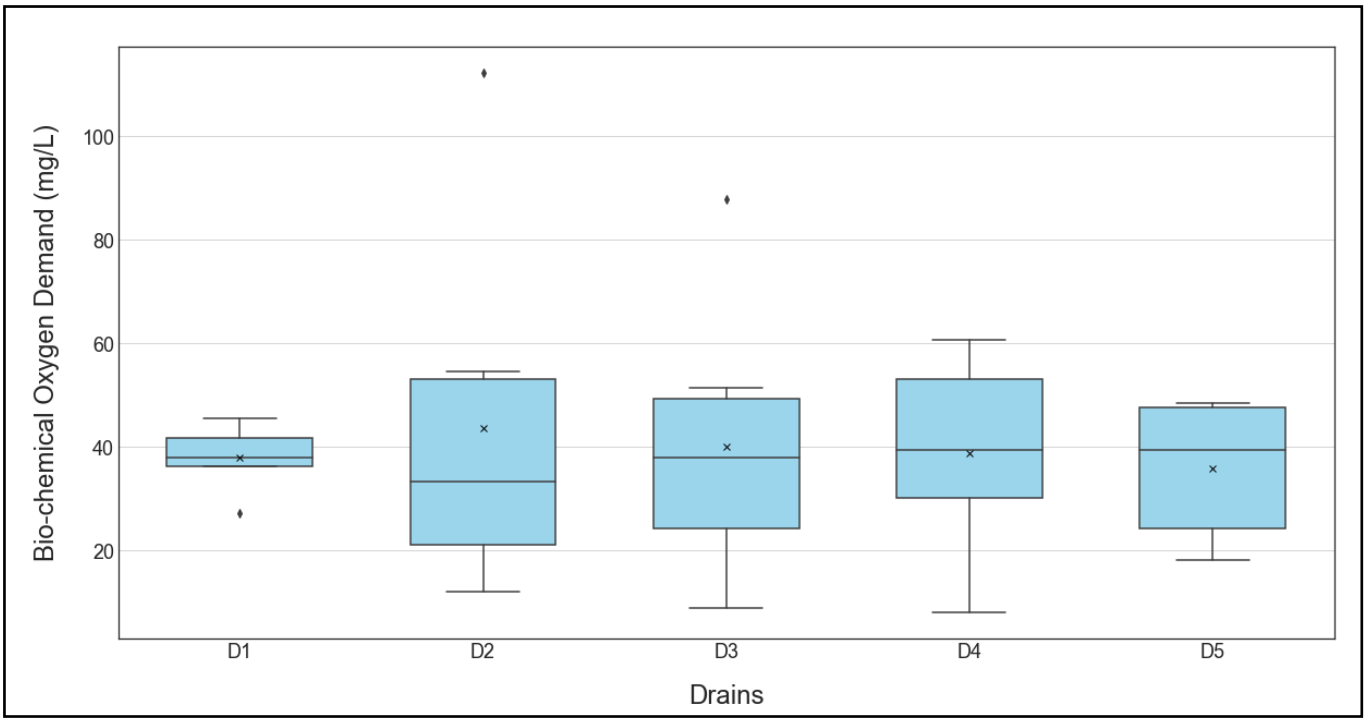



*Figure 34: Box plots showing BOD (mg/l) values at the five sampling points (D1, D2, D3, D4 and D5 as described in figure 16) on major drains of Kovvur*

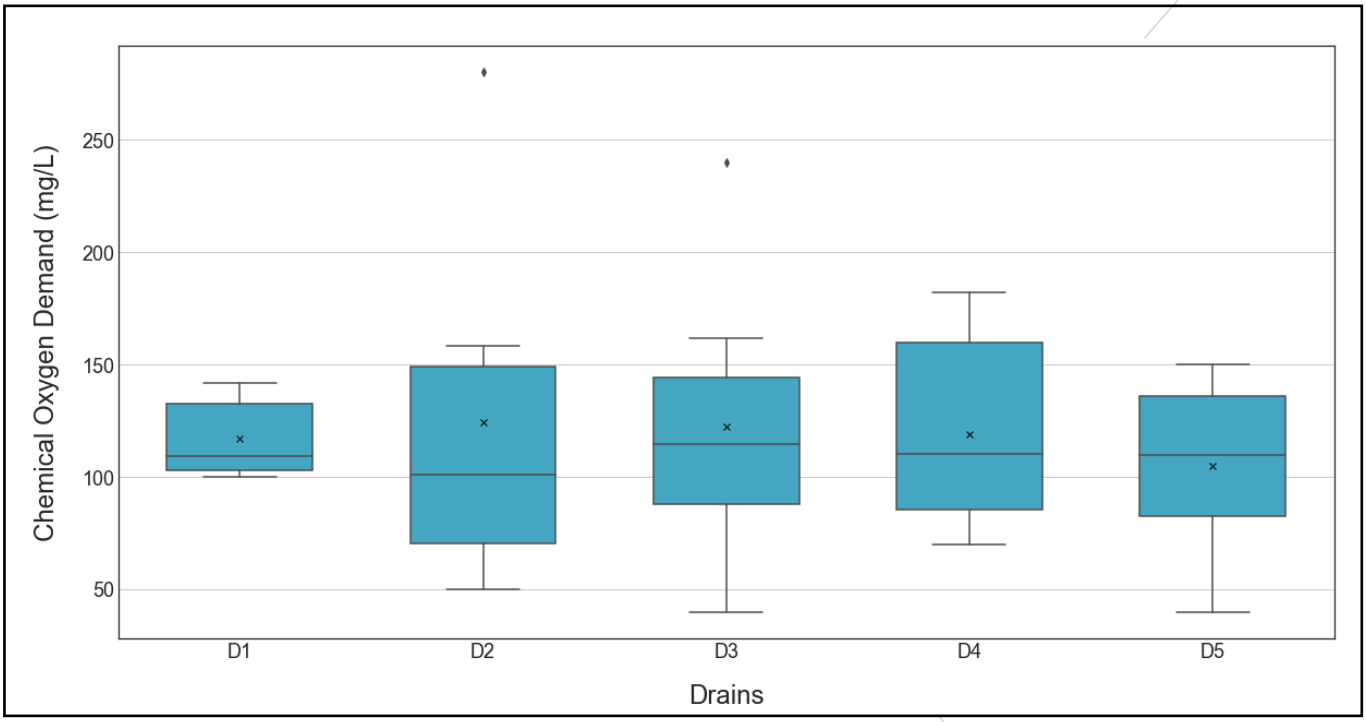



*Figure 35: Box plots showing COD (mg/l) values at the five sampling points (D1, D2, D3, D4 and D5 as described in figure 16) on major drains of Kovvur*

## Pollution load

To further assess if a treatment plant was required, total pollution load was calculated for each drain using channel flow data and BOD or COD concentrations. To illustrate, the following table estimates total BOD and COD load in kg per day for each drain in Narsapur on June 11, 2018 and July 18, 2018.

| Drain Name | June 11, 2018 | | July 18, 2018 | |
|---|---|---|---|---|
| | BOD (Kg/day) | COD (Kg/day) | BOD (Kg/day) | COD (Kg/day) |
| D1 | 238.384 | 981.579 | 37.160 | 96.223 |
| D2 | 125.477 | 543.733 | 51.195 | 147.644 |
| D3 | 33.386 | 158.142 | 10.332 | 31.008 |
| D7 | 393.947 | 1506.269 | 92.230 | 340.788 |
| D8[18] | | | | |
| D14 | 11.085 | 45.033 | 4.233 | 11.570 |
| D15 | 20.080 | 92.679 | 15.786 | 48.459 |
| S1 | 163.020 | 729.300 | 9.095 | 42.059 |
| S2[19] | | | | |
| S3 | 19.803 | 84.870 | 2.192 | 9.706 |
| S4 | 26.876 | 2527.016 | 5.120 | 309.686 |
| S6 | 473.816 | 34.681 | 109.925 | 5.100 |
| Sand ramp[20] | | | | |
| Total without D8, S2 and Sand Ramp | 1512.809 | 6703.302 | 338.008 | 1042.243 |

*Table 13: BOD and COD load per drain*

Based on the data, two findings emerge:

- Pollution load variation across drains and months is high.
- D7 (the drain near slaughterhouse) has high pollution load indicating a need for an effluent treatment plant at the slaughterhouse in addition to STP for the town.

18 Due to backwater effect, the WW generated values for this drain may not be representative. Hence, not included here in calculations.
19 Flow data for S2 not available since discharge point was not reachable.
20 Sand ramp is where a number of drains merge so this is excluded in the summation calculations for accuracy.

## BOD/COD ratio

To validate the need for STP, waste water characteristics were analyzed by calculating the ratio of BOD and COD. Low BOD/COD ratio i.e. <0.2 indicates non-biodegradable waste. However the high ratio i.e. >0.4 indicates that the waste is completely degradable. The values between 0.2-0.4 indicate that the waste is slowly degradable. Based on these values, the action required is listed below.

| BOD/COD Ratio | Type of Waste | Action required |
|---|---|---|
| <0.2 | Non-biodegradable waste | Needs urgent treatment |
| Between 0.2 and 0.4 | Slowly biodegradable waste | Waste treatment can help in quick degradation of waste |
| >0.4 | Completely biodegradable waste | Does not need treatment |

*Table 14: Waste type basis BOD/COD ratio*

In sector 1, BOD/COD ratio of most of the water samples was found to be between 0.2 and 0.4 indicating that the waste is slowly degradable and its treatment might help in fast degradation. However, in March 2019, at drain D7 and in September 2018 at D8, the BOD/COD ratio was found to be <0.2. This decline in the ratio suggests that at times the proportion of non-biodegradable waste is high in the drains and needs urgent treatment.

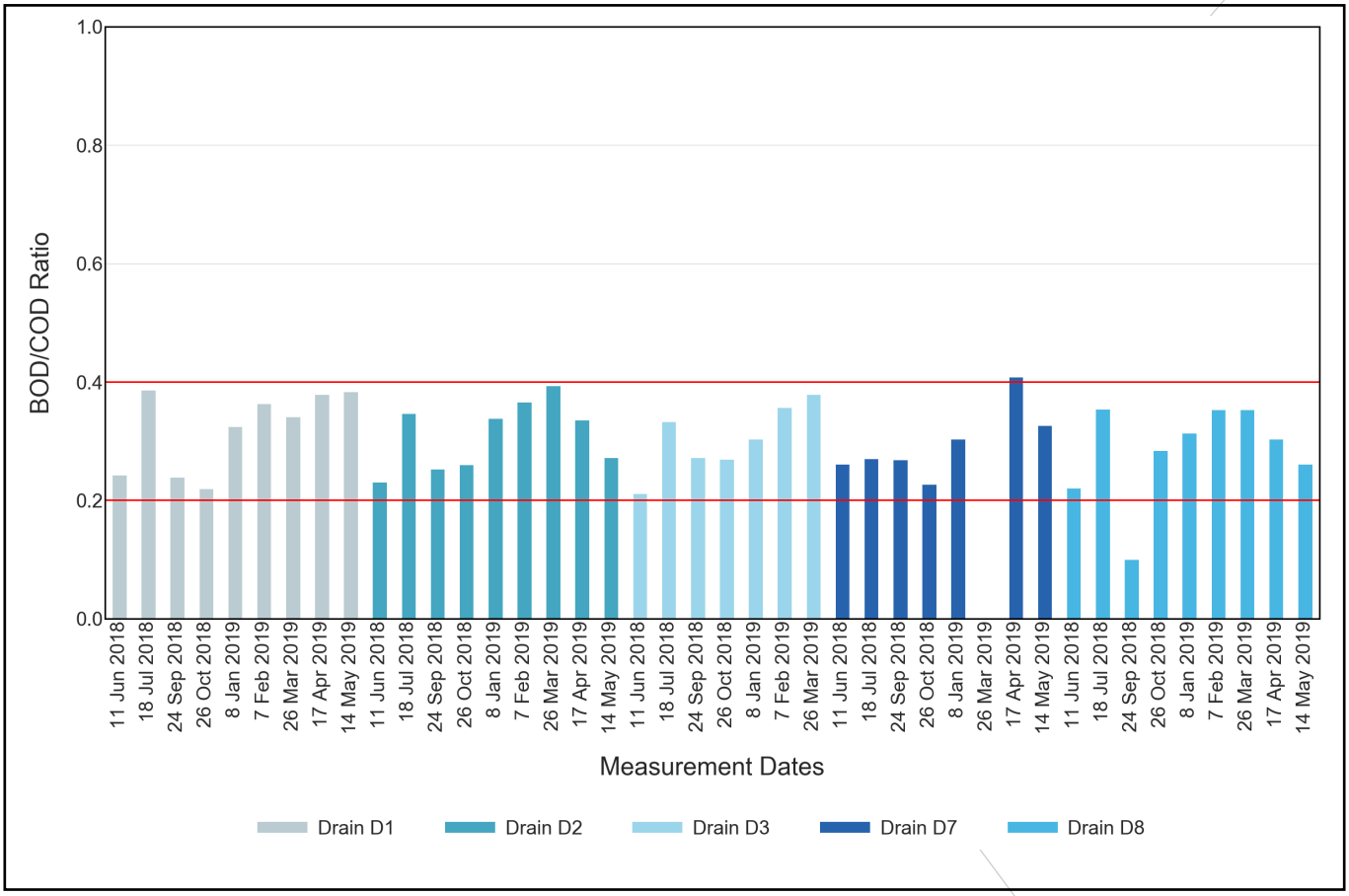



*Figure 36: Bar chart showing BOD/COD ratio for drains in sector 1 (D1, D2, D3, D7 and D8 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur*

All the samples collected from sector 2 have values ranging from 0.2 to 0.4 except one value in drain 15 that is >0.8. This indicates the dominance of slow degradable waste in the drain.

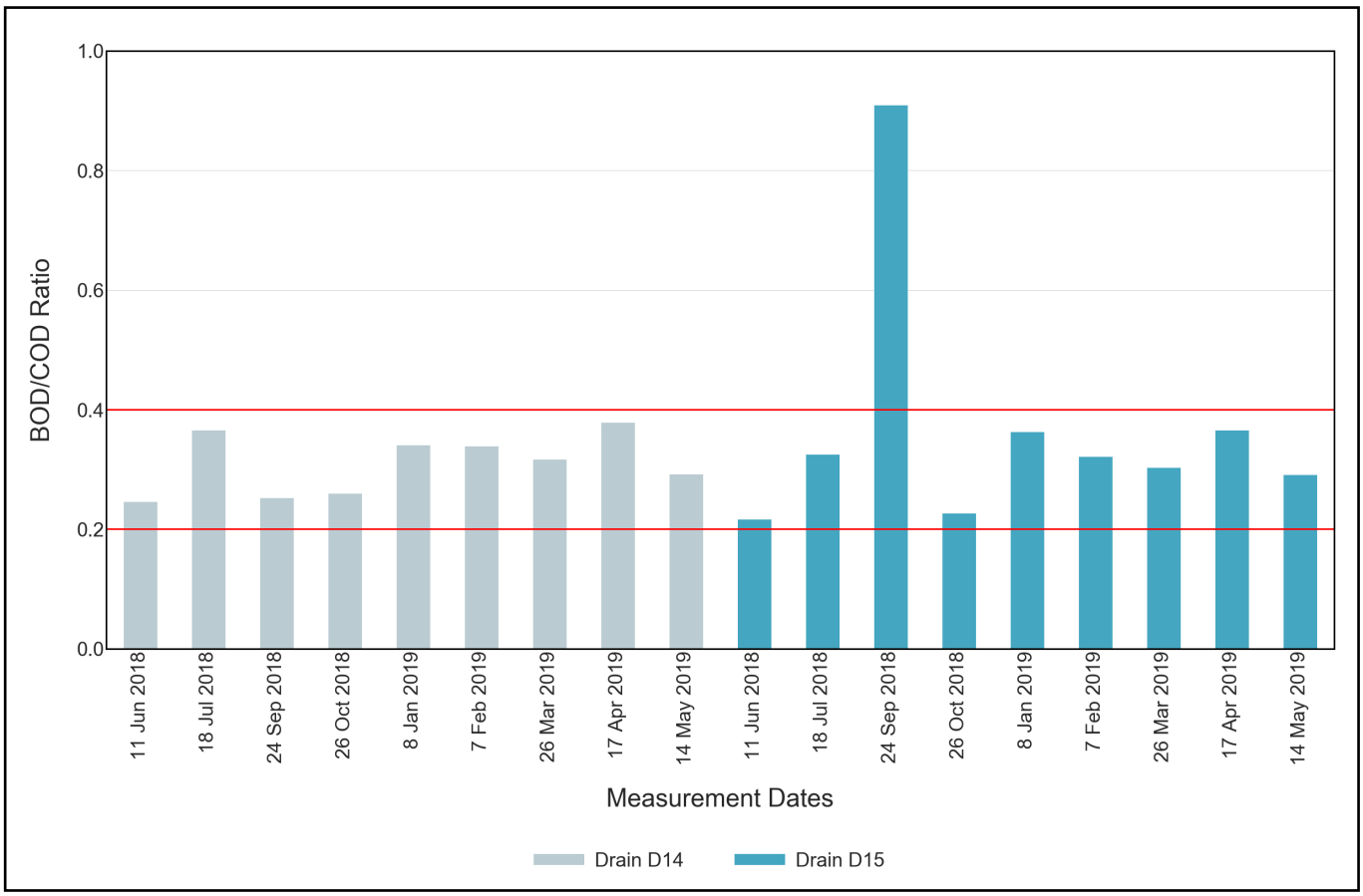



*Figure 37: Bar chart showing BOD/COD ratio for drains in sector 2 (D14 and D15 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur*

Compared to sector 1 and sector 2, the dominance of non-biodegradable waste is found to be relatively high in sector 3. It is evident from the figure that one sample in S1, two in S2, one in S3 and two in S4 have BOD/COD values < 0.2. However, two samples in S1, one in S4 and two in S6 have ratios exceeding 0.4, while the ratios of rest of the samples lie between 0.2-0.4. This indicates mixed waste characteristics in the drains and requirement for waste water treatment before disposal. Sector 3 has some slum settlements which lack proper sanitation facilities. As a result, a high proportion of non-biodegradable waste might be affecting this sector.

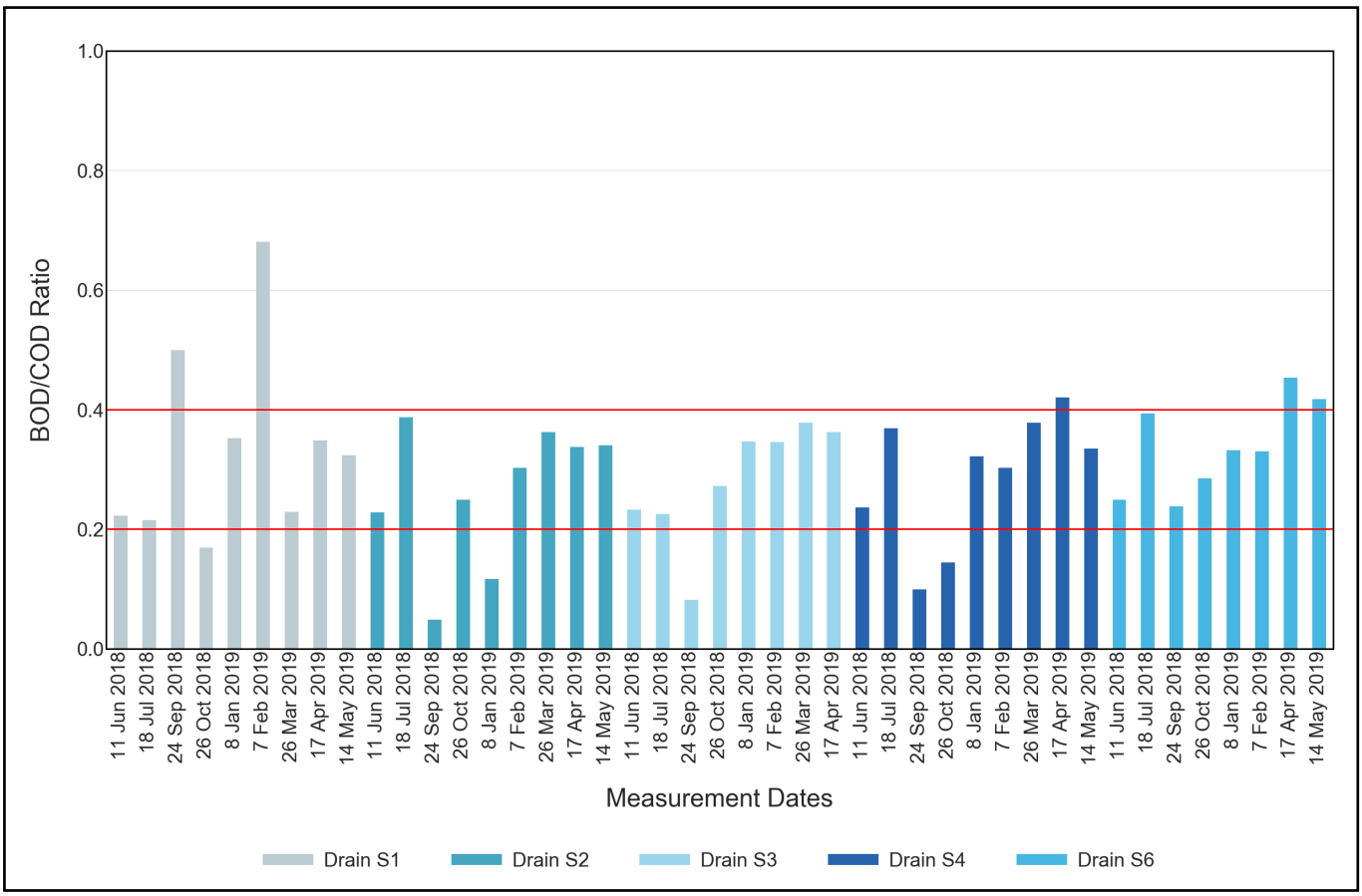



*Figure 38: Bar chart showing BOD/COD ratio for drains in sector 3 (S1, S2, S3, S4 and S6 as described in figure 14) on different days between June 11, 2018 and May 14, 2019 (noted on X-axis) in Narsapur*

Most of Narsapur's waste characteristics indicate that BOD/COD is between 0.2 and 0.4 which implies that a treatment facility may not be urgent. However, for long-term safe sanitation planning, especially in the context of Narsapur being designed as a sanitation model town, a sewage treatment plant may be installed.

### 4.3.3 Regular Cleaning of Drains

While setting up a treatment plant in Narsapur may take some time, waste water from households is discharged daily. Some drains were seen to be clogged with solid waste or with stagnant waters with algae and sludge in them. DO & turbidity sensors were installed by the project in two areas, one around Muriki Kaluva and another in drain D1 passing through the water treatment plant with the aim to understand if continuous monitoring can assess the need for regular cleaning of drains.

#### Real-time DO

Dissolved oxygen concentration is regulated by the balance between atmospheric re-aeration, photosynthesis and respiration. The change in seasonal and diurnal DO concentration is often observed in the drains or water bodies. Generally, the DO concentration increases till afternoon due to increased rate of photosynthesis by aquatic plants. A fall in DO is observed in the night due to its consumption for respiration. However, in waste water drains, this trend depends on the mixing of waste water in the drain, level of organic load, BOD and COD. The graph below depicts the DO data (in percentage) between December 28, 2018 and January 10, 2019 from real-time sensors installed on the Muriki Kaluva drain.

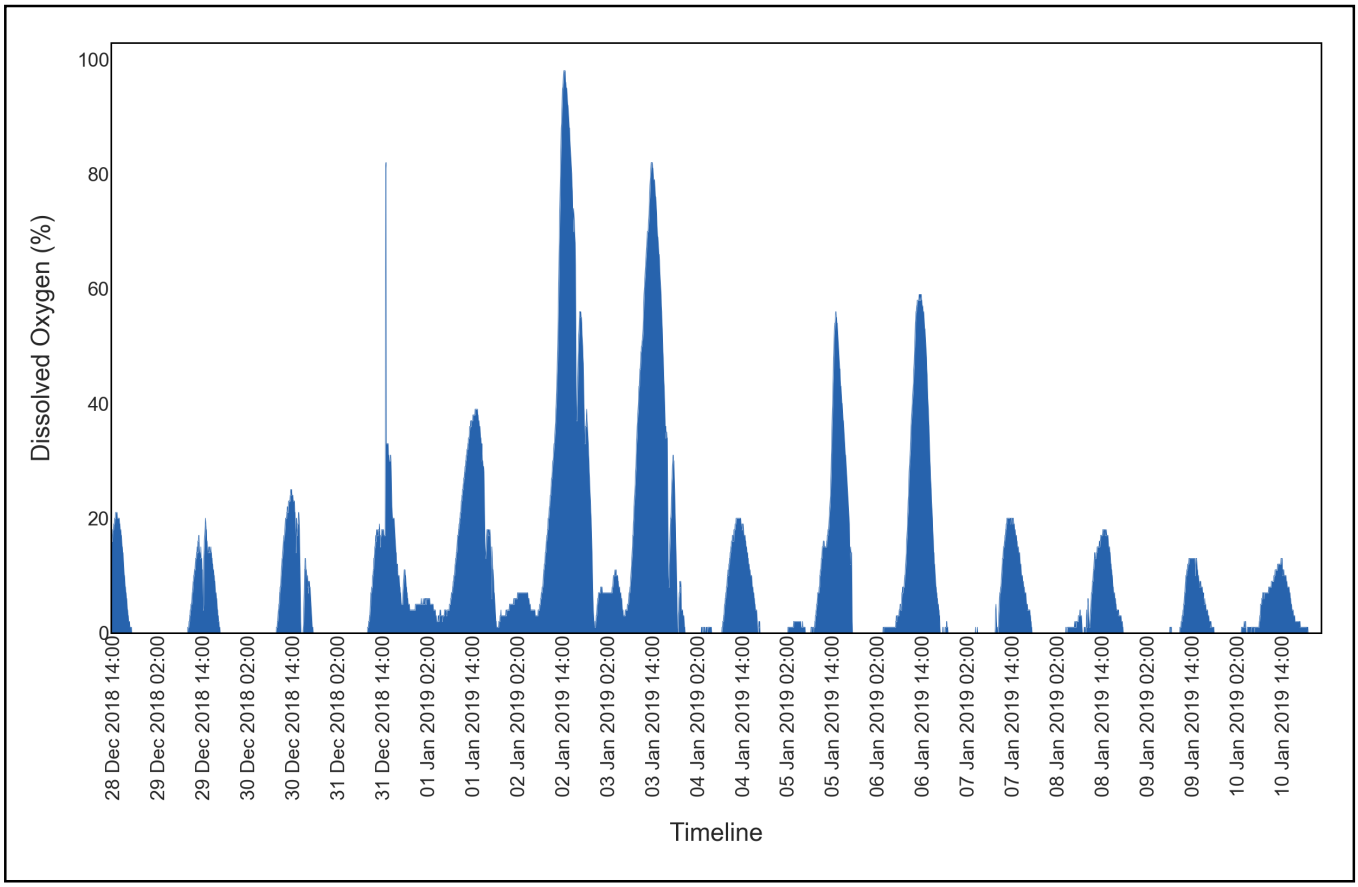
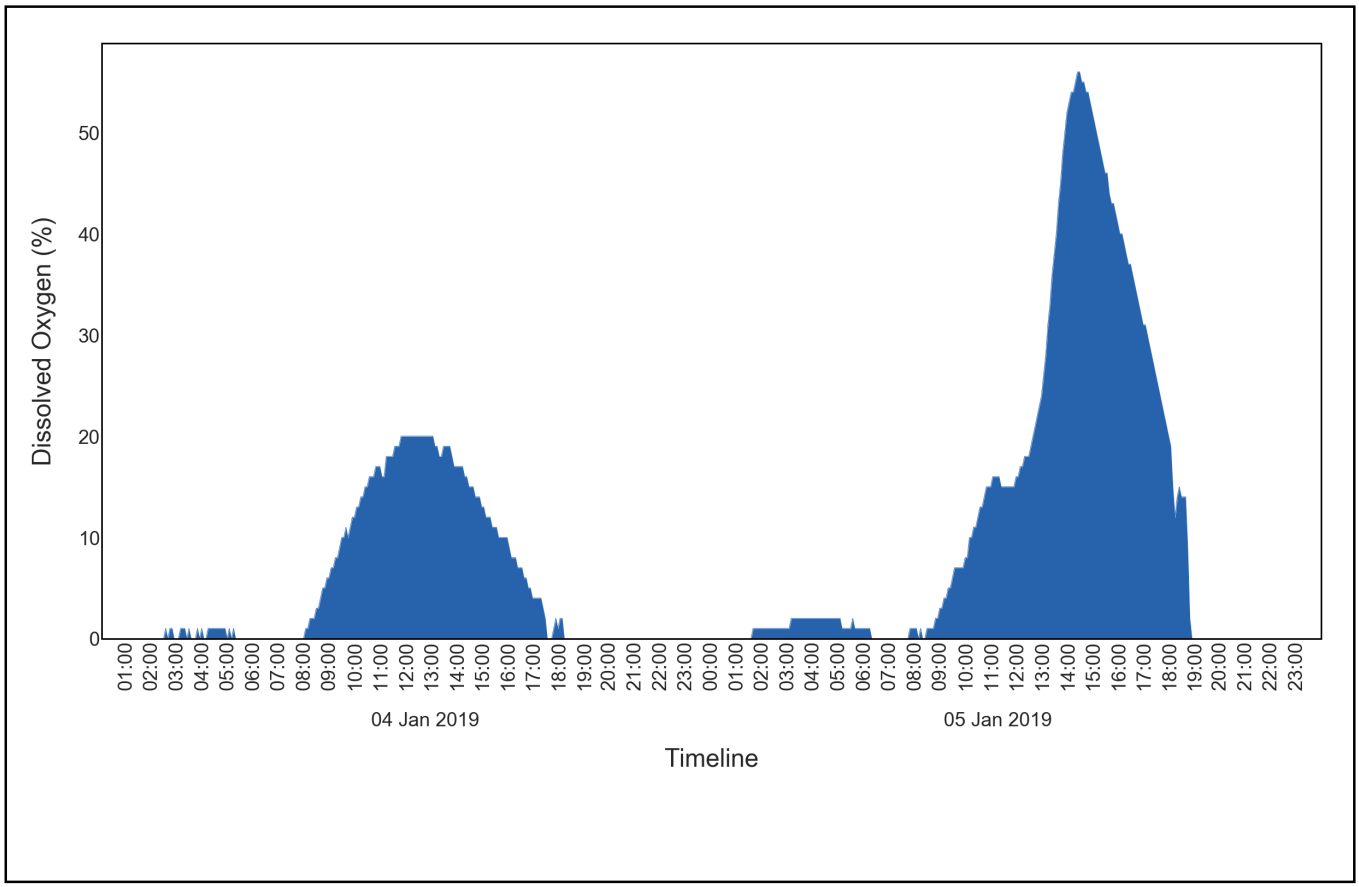



*Figure 39: imeplot of diurnal variation in dissolved oxygen concentration (mg/L) in Muriki Kaluva waste water canal across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval*

*The graph below depicts a zoom-in version of the DO data (in percentage) on January 4 and 5, 2019 from real-time sensors installed on the Muriki Kaluva drain*




*Timeplot of diurnal variation of dissolved oxygen concentration at Muriki Kauva waste water canal across different times between (noted on X-axis) on January 4 and 5, 2019, recorded at a 5-minute interval*

The DO concentration of the drain at Muruki Kaluva peaks during afternoon. It starts increasing during 8 am and declines post afternoon. One probable reason is that water at Muriki Kaluva is mostly in the stationary state with minimum flow which restricts the mixing of atmospheric oxygen in the drain during lean discharge time. Algal growth has been observed in Murki Kalva drain. Due to photosynthesis by these algae and other phytoplankton in the drains, the DO concentration increases in the morning. However, the concentration decreases post afternoon because of the reduction in intensity of sunlight that limits the release of oxygen by algae or phytoplanktons. No such variation in DO was observed in drain D1 at the water treatment plant. This is because the water has a higher flow in that drain and no algae or phytoplankton growth was observed. The clear pattern of rise and fall in DO values in a day at Muruki Kaluva indicates high algae growth in it and require regular cleaning and continuous flow. Stagnant drain waters could also be breeding grounds for mosquitoes causing health hazards.

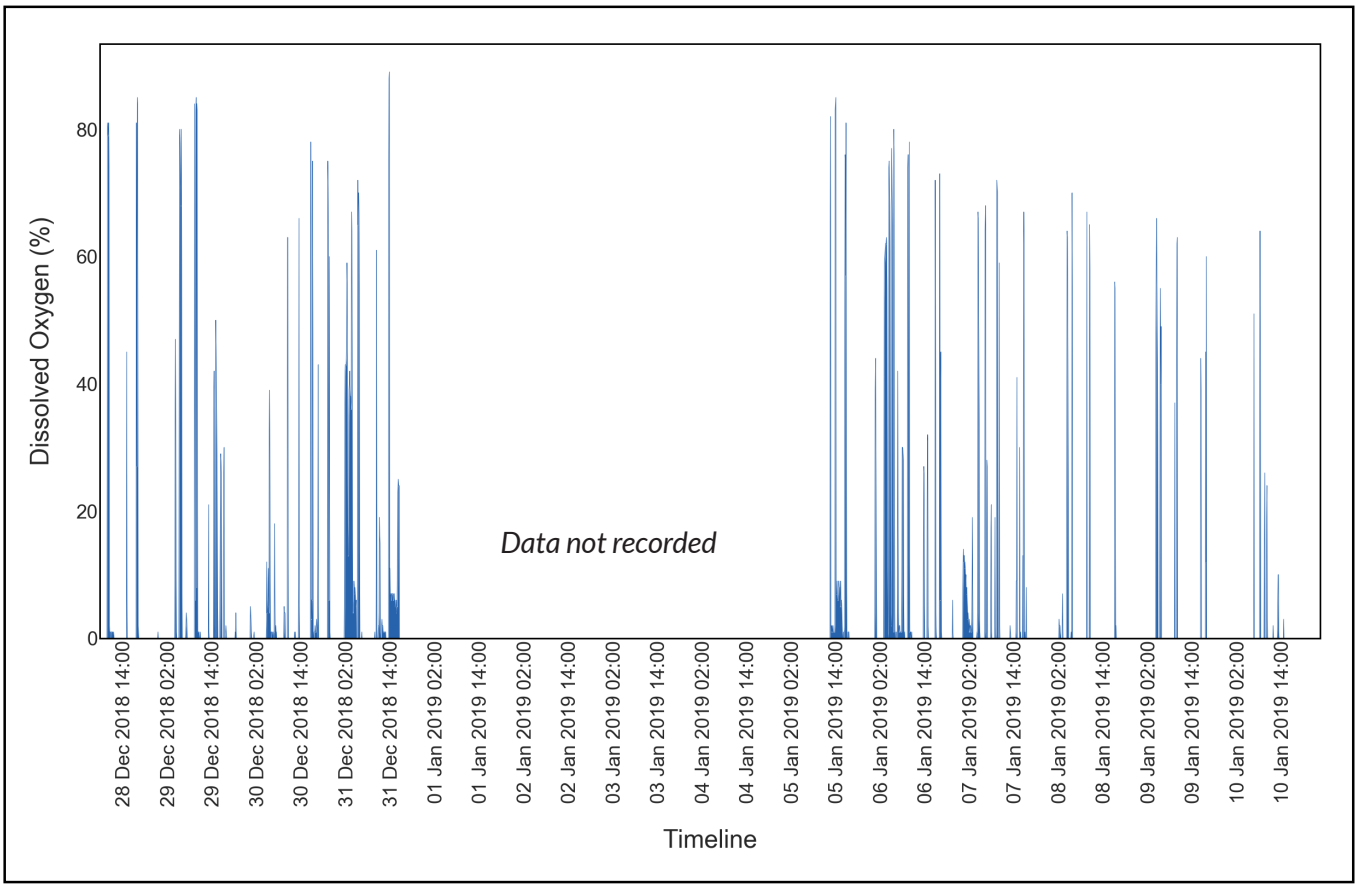



***Figure 40:*** *Timeplot of diurnal variation in dissolved oxygen concentration (mg/L) in drain D1 (as described in figure 14) passing through Narsapur Waterworks Department across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval (data not recorded between January 1, 2019 and January 5, 2019)*

## Real-time Turbidity

From the two real-time turbidity sensors installed at Muriki Kaluva and water treatment plant, no clear patterns tend to emerge between December 28, 2018 and January 10, 2019 as depicted in the graphs below.

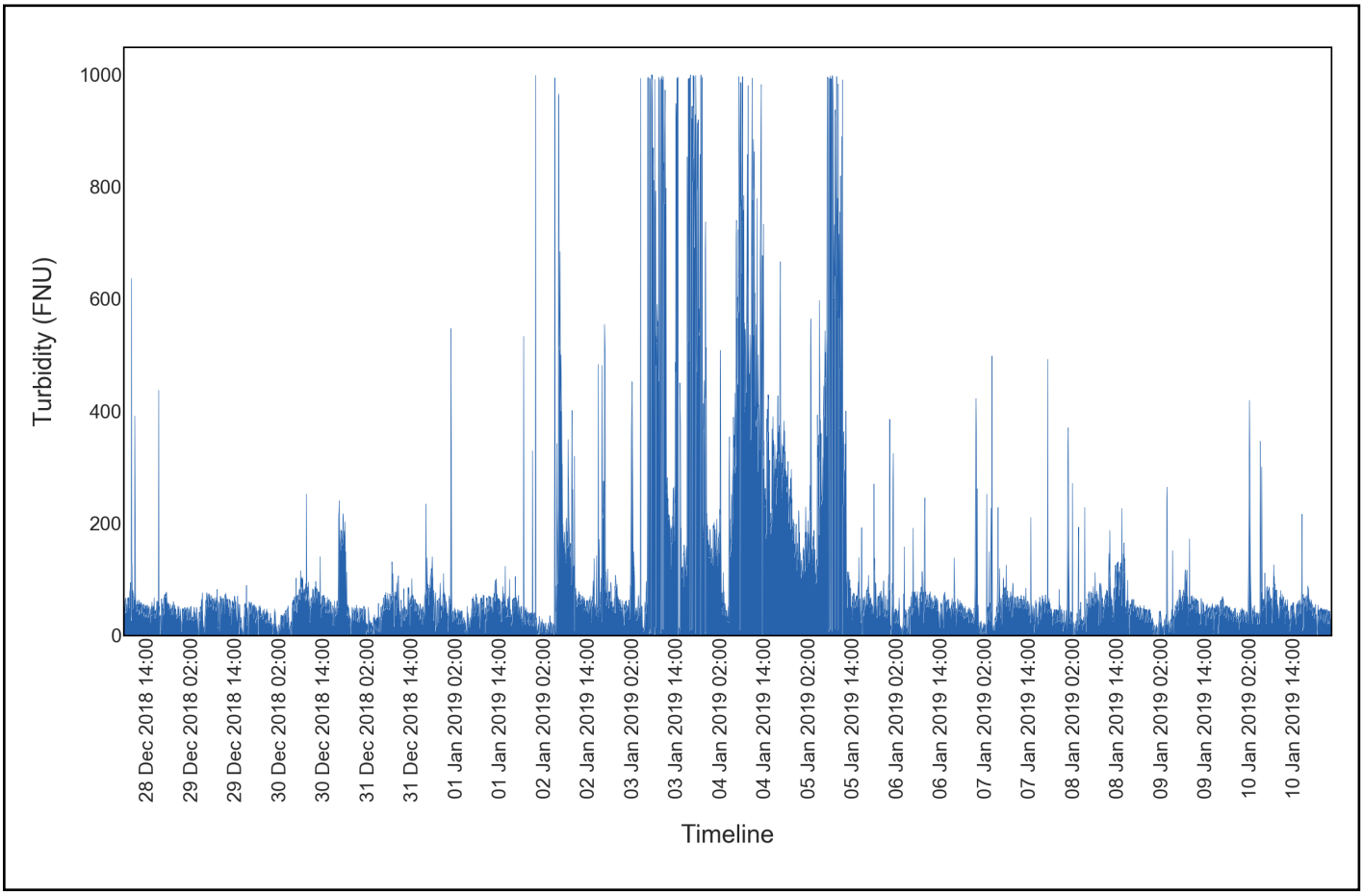



***Figure 41:*** *Timeplot of diurnal variation in turbidity concentration (FNU) in drain D1 (as described in figure 14) passing through Narsapur Waterworks Department across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval*

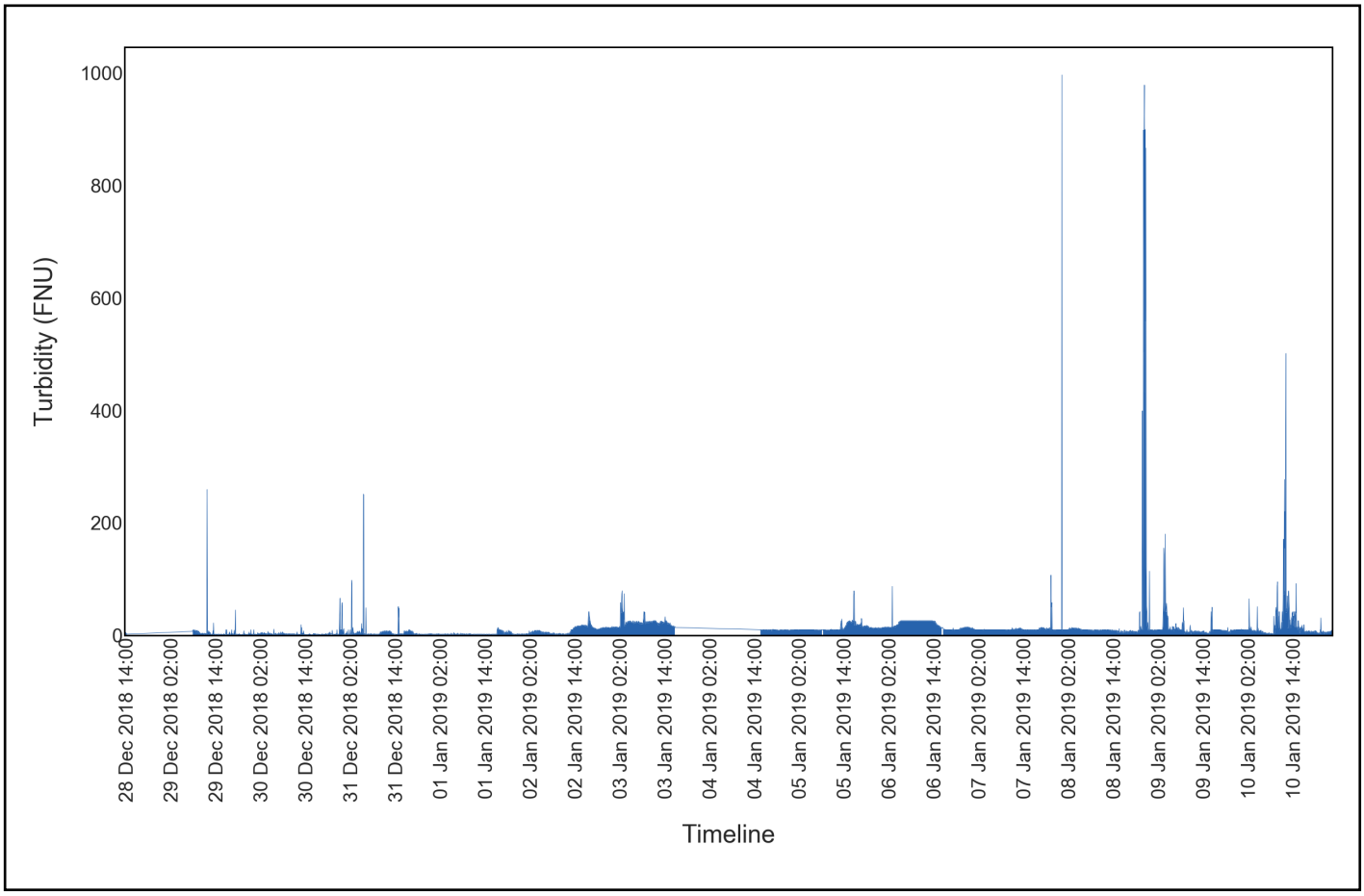



***Figure 42:*** *Timeplot of diurnal variation in turbidity concentration (FNU) in Muriki Kaluva waste water canal across different days between December 28, 2019 and January 10, 2019 (noted on X-axis), recorded at a 5-minute interval*

## 4.4 Other Findings

### 4.4.1 Nature's Attempt at Revival

Seasonal variations in multiple parameters show trends where pollution levels in monsoons are reduced as compared to summers. This is mainly due to higher quantum of freshwater in the river due to rains or opening of barrage gates releasing water from upstream. While the pollution load in the river from these towns remains similar across seasons, the dilution due to increased fresh water improves the water quality. This is nature's way of reviving the river basin ecosystem annually.

**Electrical Conductivity reduces in monsoons in Narsapur but remains the same in the Rajahmundry-Kovvur stretch of River Godavari**

Most days of the year, the conductivity values are extraordinarily high in Godavari in Narsapur because of sea water influence except during some days in the monsoon. Low EC values in monsoons indicate the opening of fresh water barrage during August/September and high runoff from the adjoining catchment areas. Pin-pointing sources of pollution is more effective in the monsoon season when overall EC values are low (e.g. leachates from solid waste collection site leaking into Godavari). See picture titled monsoon 2017.

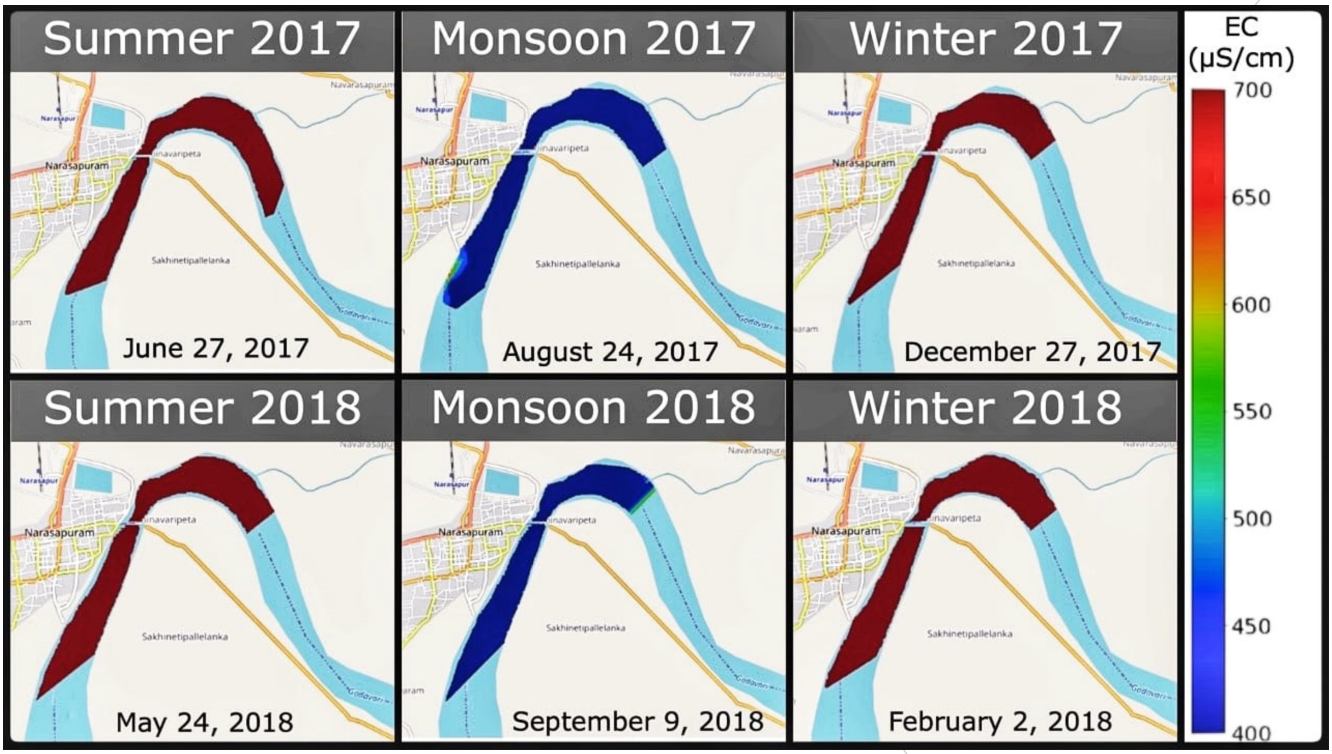



*Figure 43: Heatmaps showing seasonal variation of electrical conductivity (µS/cm) in Narsapur stretch of River Godavari during different seasons in 2017 and 2018*

### Chlorophyll level rise in summers

Chlorophyll provides green color to plants. High chlorophyll in river or surface water bodies indicates high algae. These algae serve as the primary producer and play an important role in the food chain. However, excess concentration can cause algal bloom and eutrophication. High chlorophyll level also indicates excess of nutrient in the aquatic system. Generally, chlorophyll concentration increases during the summer because of relatively less flow rate, high intensity of sunlight and availability of nutrients. Chlorophyll levels were observed to rise in the summers at Narsapur stretch of River Godavari.

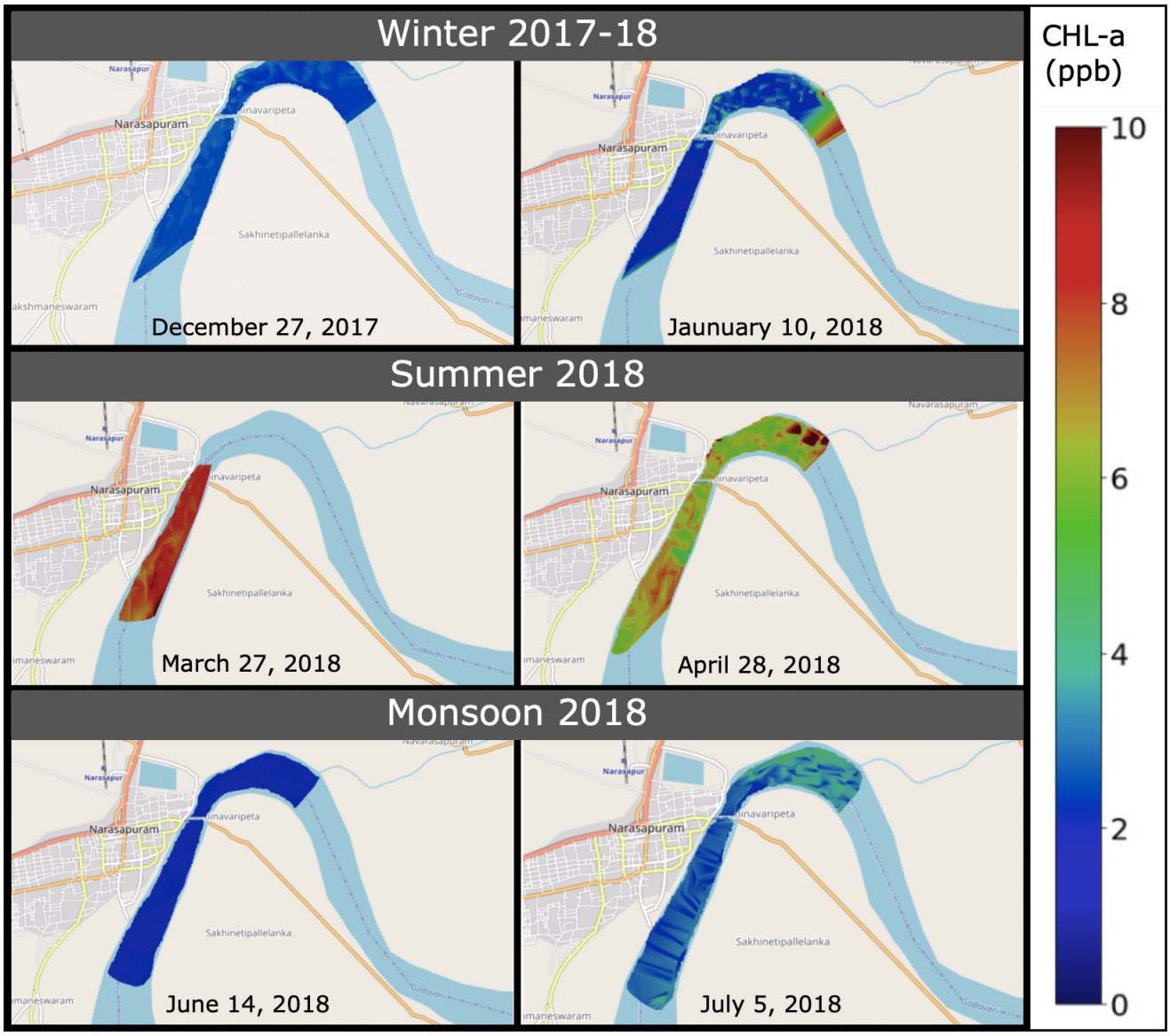



*Figure 44: Heatmaps showing seasonal variation of chlorophyll (mg/L) in Narsapur stretch of River Godavari during different seasons*

### 4.4.2 Fish Kills-Inverse Relation between EC & DO

Based on the sensors reading at Narsapur and Kovvur - Rajahmundry stretch of River Godavari, is negatively correlated with the dissolved oxygen concentration. It is evident from the figure that the water with high EC has relatively low DO and when we plot all measured values of DO against EC it shows a negative trend line. High concentration of EC in water indicates high concentration of dissolved salt/ions. However, DO concentration in water is affected by several factors including atmospheric mixing/ pressure, temperature, salinity and growth, and photosynthesis of phytoplankton in the water bodies. The solubility of gases in water decreases by the addition of salts, resulting in a decrease in DO. This relation has a seasonal effect due to monsoons but increase in EC due to domestic discharge or salts from industrial waste can lead to a decrease in DO. Aquatic life needs DO > 5mg/l for survival. Fish kill events have been recorded when DO levels fall below 3mg/l.

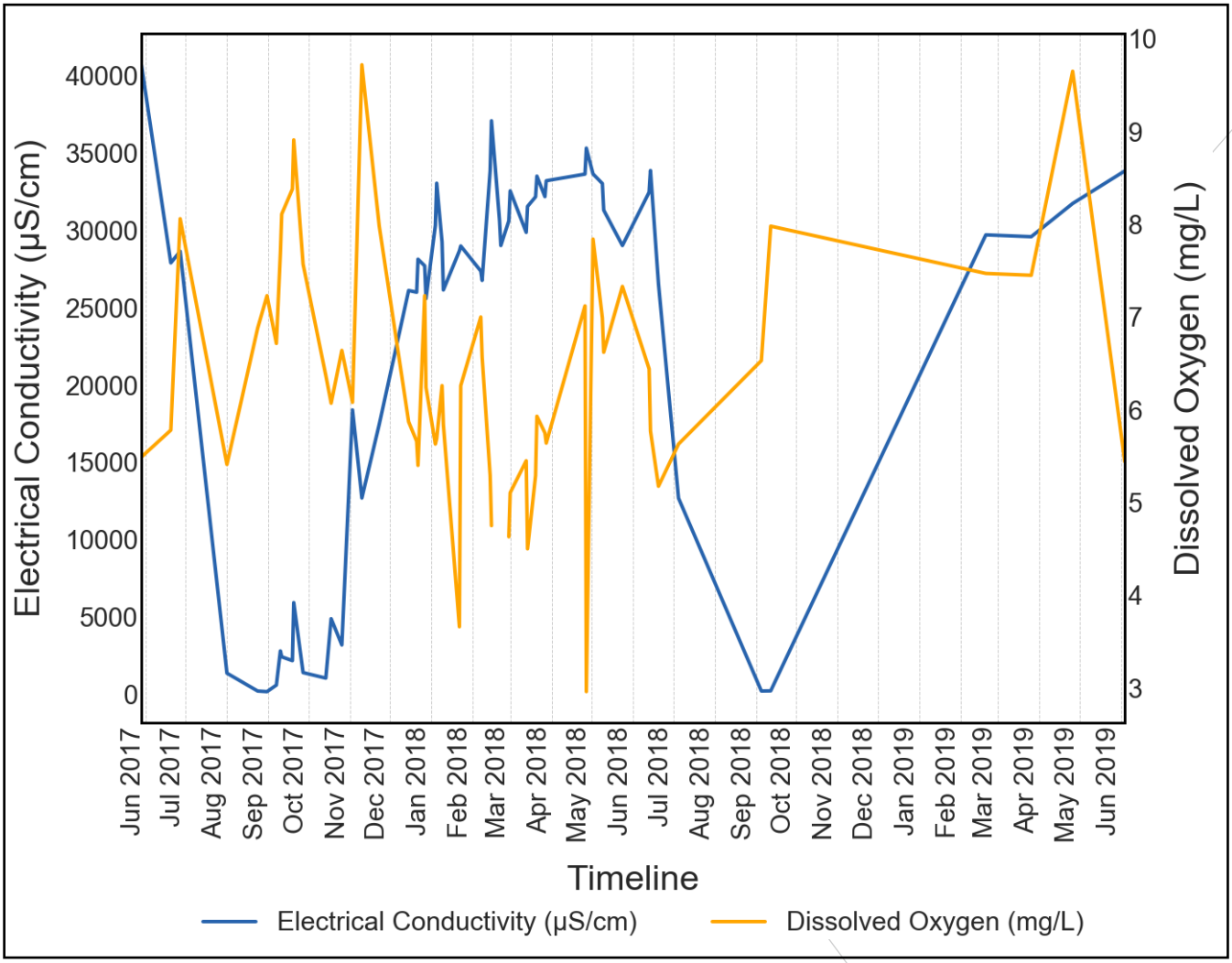



*Figure 45: Timeplots showing variation in dissolved oxygen concentration (mg/L) and electrical conductivity (µS/cm) between July 2017 and June 2019 (noted on X-axis) in Narsapur stretch of River Godavari*

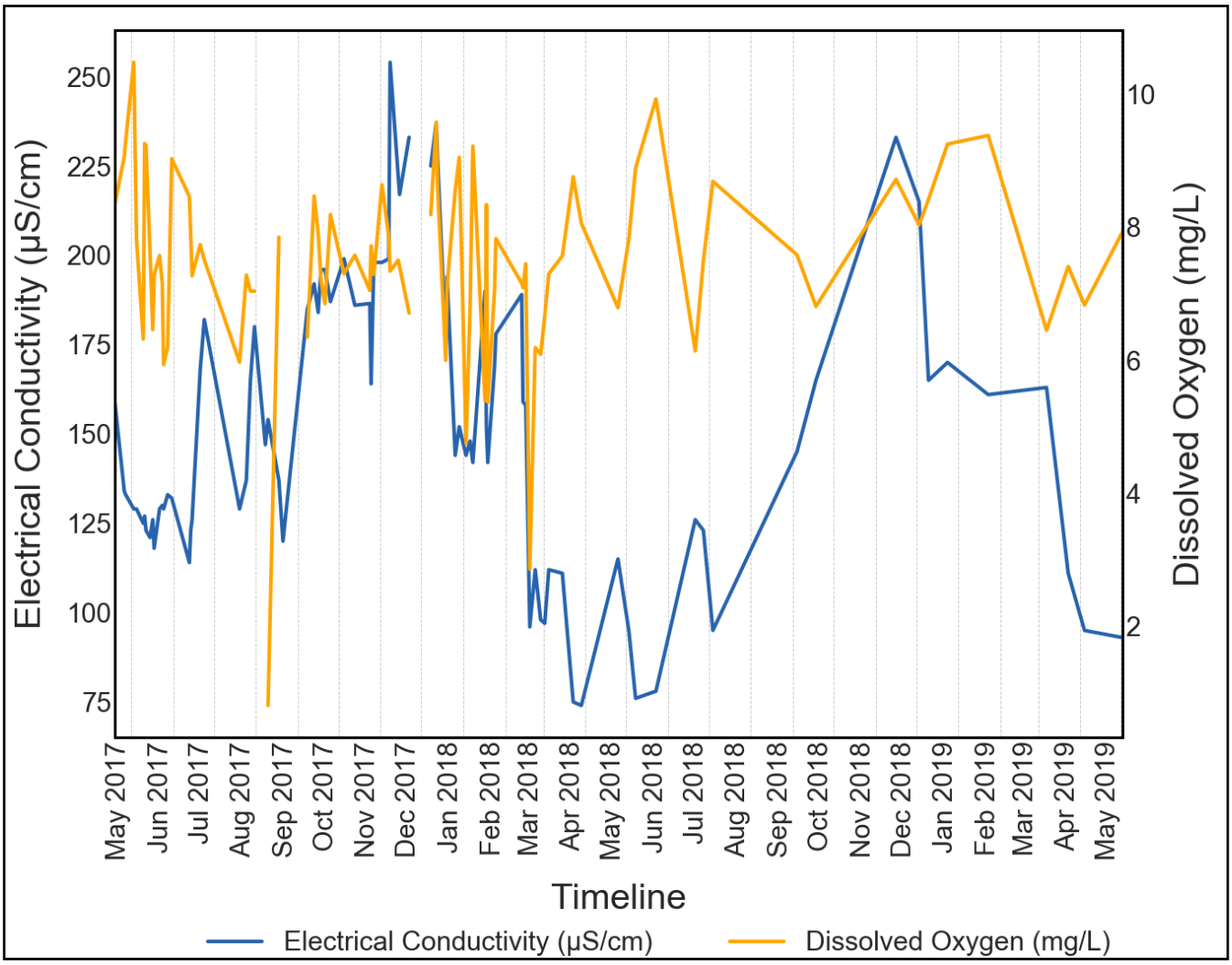



*Figure 46: Timeplots showing variation in dissolved oxygen concentration (mg/L) and electrical conductivity (µS/cm) between May 2017 and May 2019 (noted on X-axis) in Rajahmundry stretch of River Godavari*

### 4.4.3 Water Quality Prediction Models

In this section, we attempt to predict values of ex-situ parameters – COD and BOD, using in-situ parameters which are measured using sensors. The following figure shows the classification of machine learning techniques. Our prediction problem lies around regression.

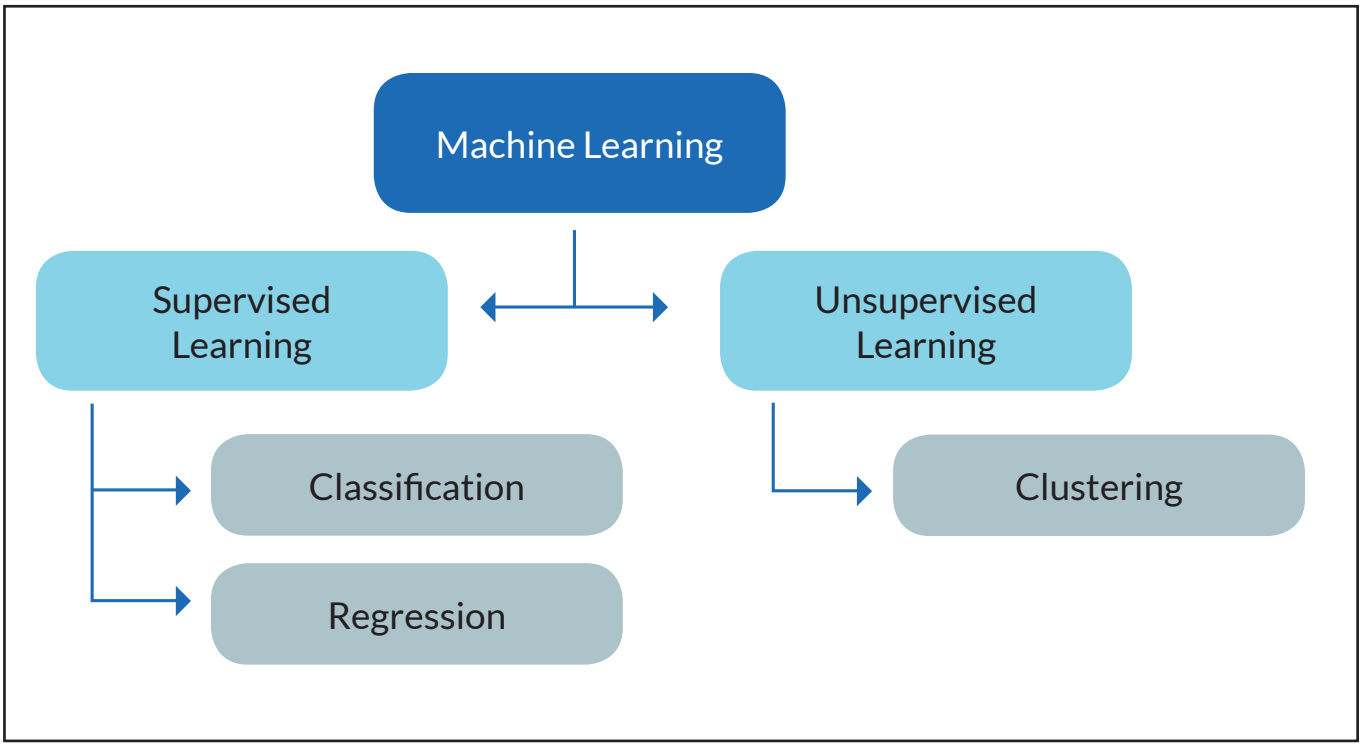



*Figure 47: Diagram depicting various machine learning approaches*

The collected data was subjected to different regression machine learning techniques such as artificial neural network-multiple layer perceptron (ANN-MLP), Support vector mechanism and AdaBoost. Parameters such as temperature, electrical conductivity, pH, turbidity, DO, Chlorophyll-a, CDOM, Tryptophan and nitrate are used in different combinations to predict COD and BOD. While nitrate can be measured using a sensor as well as in lab, the values used in models here were lab-based values since only limited sensor-based nitrate values were available with us. The respective machine learning techniques were applied to the data collected across various rivers in India from April 19, 2017 to September 24, 2018. The table below shows the descriptive statistics of the data.

| Descriptive Statistics | | | | | |
|---|---|---|---|---|---|
| | N | Minimum | Maximum | Mean | St.Dev |
| Temperature (°C) | 259 | 19.12 | 34.8 | 29.28 | 2.79 |
| Conductivity (µS/cm) | 267 | 85 | 48700 | 5813.94 | 10684 |
| pH | 260 | 6.62 | 9.44 | 7.86 | 0.55 |
| Turbidity (FTU) | 269 | 0.1 | 606 | 41.67 | 70.55 |
| DO (mg/L) | 261 | 0 | 13.6 | 4.90 | 3.12 |
| Chlorophyll-a (µg/L) | 127 | 0.16 | 77.14 | 6.19 | 13.01 |
| CDOM (µg/L) | 128 | 39.65 | 321.9 | 117.25 | 50.93 |
| TRPN (µg/L) | 128 | 6.4 | 78.43 | 22.04 | 16.19 |
| Nitrate (mg/L) | 260 | 0.8 | 39.8 | 5.41 | 5.73 |
| COD (mg/L) | 261 | 0 | 1950 | 120.98 | 226.97 |
| BOD (mg/L) | 260 | 1 | 380 | 33.81 | 56.80 |

***Table 15:*** *Descriptive statistics of water quality data collected across various rivers in India from April 19, 2017 to September 24, 2018*

## Simple linear regression

Regression technique allows not only to investigate the relationship between variables but also to model the relationship between variables. This enables one to make predictions about what one variable will do based on the scores of some other variables. The table below depicts the correlation coefficients between COD, BOD and other parameters. There is a weak correlation observed between COD, BOD and conductivity, nitrate, CDOM and tryptophan; negative correlations were observed between COD, pH and DO. Similar trend was observed for BOD. Though these relations are weak, they show statistical significance (since significance (2 tailed) is <0.05).

| | | Temp (°C) | EC (µS/cm) | pH | | DO (mg/L) | Nitrate (mg/L) | Chl-a (µg/L) | CDOM (µg/L) | |
|---|---|---|---|---|---|---|---|---|---|---|
| BOD | Pearson Correlation | 0.1348 | 0.3004 | -0.2598 | 0.0145 | -0.2430 | 0.2462 | -0.0717 | 0.2324 | 0.2878 |
| | Sig. (2-tailed) | 0.0328 | 0.0000 | 0.0000 | 0.8179 | 0.0001 | 0.0001 | 0.4290 | 0.0091 | 0.0011 |
| | N | 251 | 252 | 245 | 253 | 245 | 257 | 124 | 125 | 125 |
| COD | Pearson Correlation | 0.0610 | 0.3036 | -0.2945 | 0.0111 | -0.1879 | 0.1582 | -0.0421 | 0.2170 | 0.3052 |
| | Sig. (2-tailed) | 0.3346 | 0.0000 | 0.0000 | 0.8604 | 0.0031 | 0.0109 | 0.6409 | 0.0147 | 0.0005 |
| | N | 252 | 253 | 246 | 254 | 246 | 258 | 125 | 126 | 126 |

*Table 16: Correlation between COD, BOD and other parameters*

## Multiple linear regression

Simple linear regression enables prediction of the value of one variable based on the value of a single predictor variable whereas multiple regression allows you to use multiple predictors. To predict the COD and BOD, four different combinations of input parameters are applied. The input parameter combinations are as follows:

a. Model-01 = Temperature, pH, EC, DO and turbidity.
b. Model-02 = Temperature, pH, EC, DO, nitrate and turbidity.
c. Model-03 = Temperature, pH, EC, DO, turbidity, chlorophyll-a, CDOM and tryptophan.
d. Model-04 = Temperature, pH, EC, DO, turbidity, nitrate, chlorophyll-a, CDOM and tryptophan.

Table below depicts the results of multiple regression where R2 is the proportion of variance explained by an independent variable in a regression model, RMSE is the root mean square error representing the difference between predicted and actual observed values and p is the level of significance. The 4th combination of parameters is able to explain only 34% and 37% of the data for BOD and COD respectively. All the models show p-value is <0.05 indicating that these models are statistically significant. However, they cannot be used to explain the variability of the targeted parameters fully.

| Bio-chemical oxygen demand | | | | | |
|---|---|---|---|---|---|
| | N | R | $R^2$ | f-value | p-value |
| Model-01 | 235 | 0.454 | 0.206 | 11.92 | 0.000 |
| Model-02 | 235 | 0.5 | 0.25 | 12.7v | 0.000 |
| Model-03 | 111 | 0.583 | 0.34 | 6.619 | 0.000 |
| Model-04 | 111 | 0.586 | 0.34 | 5.92 | 0.000 |
| **Chemical oxygen demand** | | | | | |
| Model-01 | 235 | 0.449 | 0.201 | 11.646 | 0.000 |
| Model-02 | 235 | 0.473 | 0.224 | 11.053 | 0.000 |
| Model-03 | 111 | 0.6 | 0.365 | 7.474 | 0.000 |
| Model-04 | 111 | 0.61 | 0.37 | 6.609 | 0.000 |

*Table 17: R, $R^2$, f and p-values resulted from multiple regression*

## Support Vector Machine

As linear models are not able to predict the targeted parameters, the data was treated with non-linear techniques such as artificial neural network, support vector regression and AdaBoost. Support vector machine (SVM) is based on a technique called the kernel trick to transform the data, and based on these transformations it finds an optimal boundary between the possible outputs. Simply put, it does some extremely complex data transformations, then figures out how to separate the data based on the labels or outputs as defined. It is a discriminative classifier formally defined by a separating hyperplane.

**a.** Basic concept: SVMs were developed by Cortes & Vapnik (1995) for binary classification. Their approach may be roughly sketched as follows:

**b.** Class separation: This part looks for the optimal separating hyperplane between the two classes by maximizing the margin between the classes. Closest points are the points lying on the boundaries and are called support vectors, and the middle of the margin is our optimal separating hyperplane.

**c.** Overlapping classes: Data points on the "wrong" side of the discriminant margin are weighted down to reduce their influence.

**d.** Nonlinearity: When we cannot find a linear separator, data points are projected into a higher-dimensional space where they effectively become linearly separable.

**e.** Problem solution: The whole task can be formulated as a quadratic optimization problem which can be solved by known techniques.
A program able to perform all these tasks is called a Support Vector Machine.

The figure below shows the scatter diagrams between measured COD, BOD and predicted COD and BOD with the SVM technique. Table below shows the calculated values of $R^2$, and RMSE and p.

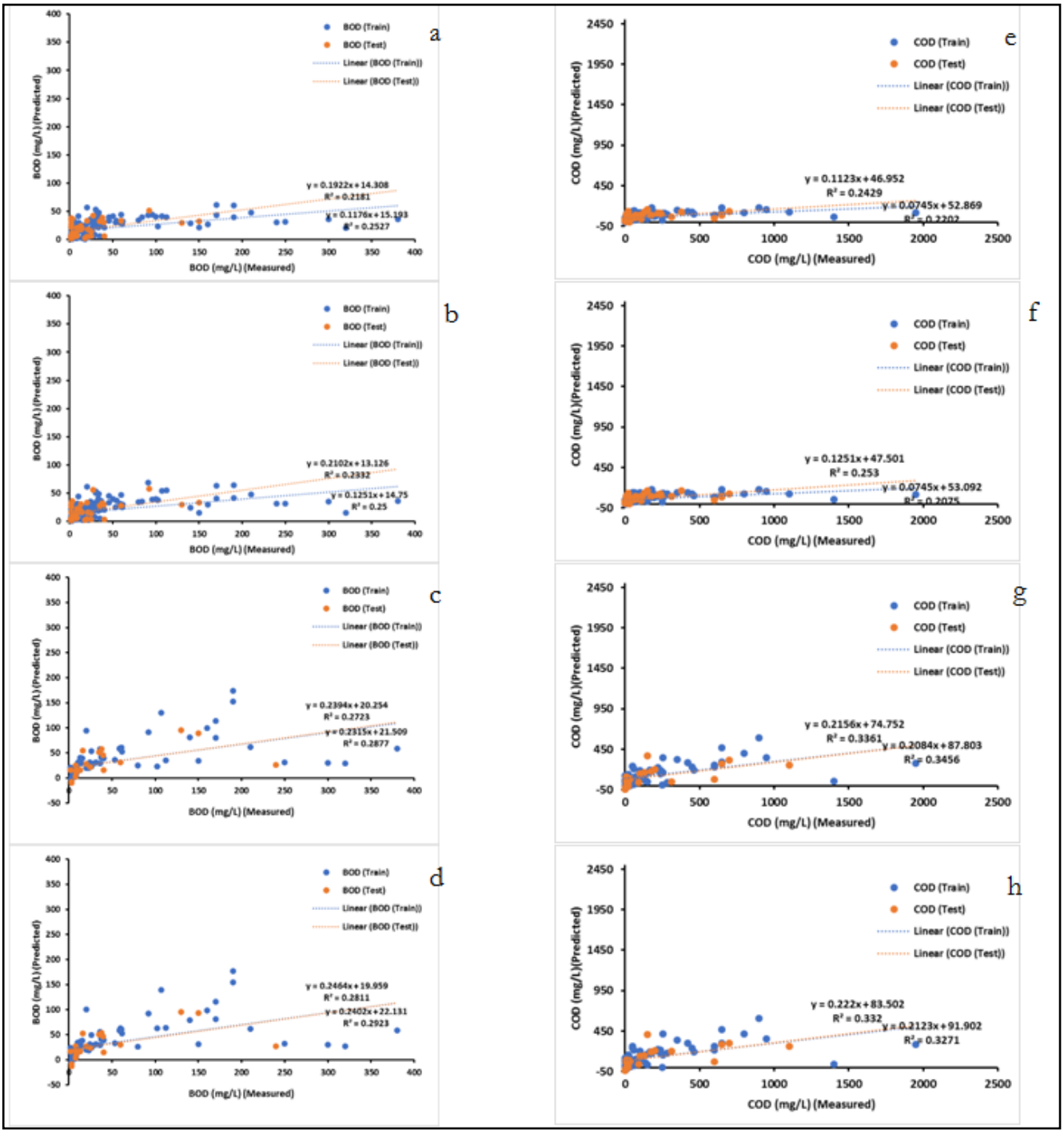



*Figure 48: a, b, c and d are the scatter plots between BOD measured and predicted BOD using SVM method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using SVM method with the four sets of input parameters*

| | | | BOD | | | | COD | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | R | $R^2$ | RMSE | p-value | R | $R^2$ | RMSE | p-value |
| Model 01 | Train | 160 | 0.503 | 0.253 | 12.8 | 0.000 | 0.469 | 0.220 | 35.6 | 0.000 |
| | Test | 41 | 0.467 | 0.218 | 12.1 | 0.002 | 0.493 | 0.243 | 35.5 | 0.001 |
| Model 02 | Train | 160 | 0.500 | 0.250 | 13.7 | 0.000 | 0.456 | 0.208 | 37.0 | 0.000 |
| | Test | 41 | 0.483 | 0.233 | 12.7 | 0.001 | 0.503 | 0.253 | 38.5 | 0.001 |
| Model 03 | Train | 73 | 0.536 | 0.288 | 30.4 | 0.000 | 0.588 | 0.346 | 96.5 | 0.000 |
| | Test | 19 | 0.522 | 0.272 | 25.2 | 0.022 | 0.580 | 0.336 | 96.5 | 0.009 |
| Model 04 | Train | 73 | 0.541 | 0.292 | 31.2 | 0.000 | 0.572 | 0.327 | 102.5 | 0.000 |
| | Test | 19 | 0.530 | 0.281 | 25.4 | 0.020 | 0.576 | 0.332 | 100.3 | 0.010 |

*Table 18: R, $R^2$, f and p-values resulted from SVM*

## AdaBoost

Boosting is a general ensemble method that creates a strong classifier from several weak classifiers. This is done by building a model from the training data, then creating a second model that attempts to correct the errors from the first model. Models are added until the training set is predicted perfectly, or a maximum number of models are added. AdaBoost was the first successful algorithm developed for binary classification. Modern boosting methods are built on AdaBoost, most notably stochastic gradient boosting machines. In the figure, a,b,c and d are the scatter plot between BOD measured and predicted using the AdaBoost technique with the four sets of input parameters. Similarly, e, f, g and h are the scatter plots between COD measured and predicted through the AdaBoost technique. Table below shows the $R^2$, and f and p values calculated. The calculated values in the training dataset can explain 97% and 95% data variability for BOD and COD respectively. But the AdaBoost machine is not able to explain the test datasets as evidenced by the 0.455 and 0.137 $R^2$ for BOD and COD respectively.

| | | | BOD | | | | COD | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | R | $R^2$ | RMSE | p-value | R | $R^2$ | RMSE | p-value |
| Model - 01 | Train | 160 | 0.937 | 0.877 | 19.9 | 0.000 | 0.965 | 0.932 | 60.8 | 0.000 |
| | Test | 41 | 0.644 | 0.415 | 19.8 | 0.002 | 0.377 | 0.142 | 67.7 | 0.010 |
| Model - 01 | Train | 160 | 0.940 | 0.880 | 19.9 | 0.000 | 0.962 | 0.927 | 63.7 | 0.000 |
| | Test | 41 | 0.500 | 0.250 | 18.2 | 0.000 | 0.553 | 0.3 | 45.5 | 0.000 |
| Model - 01 | Train | 73 | 0.980 | 0.950 | 16.0 | 0.000 | 0.979 | 0.958 | 63.9 | 0.000 |
| | Test | 19 | 0.470 | 0.221 | 44.5 | 0.042 | 0.673 | 0.453 | 107.0 | 0.001 |
| Model - 01 | Train | 73 | 0.985 | 0.971 | 13.6 | 0.000 | 0.980 | 0.952 | 64.4 | 0.000 |
| | Test | 19 | 0.675 | 0.455 | 39.3 | 0.001 | 0.370 | 0.137 | 318.0 | 0.118 |

*Table 19: R, $R^2$, f and p-values resulted from SVM*

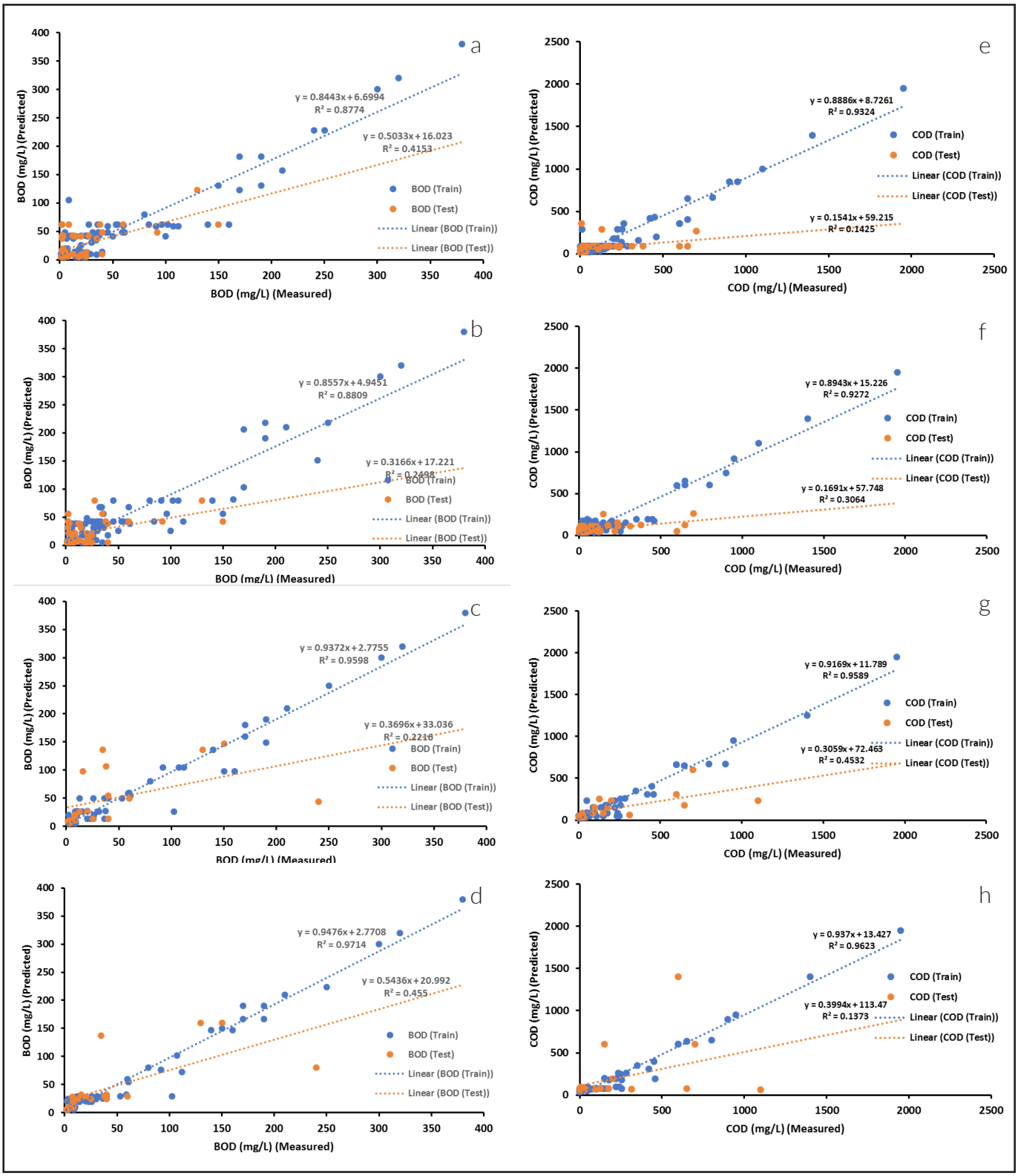



*Figure 49: a,b,c and d are the scatter plots between BOD measured and predicted BOD using AdaBoost method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using AdaBoost method with the four sets of input parameters*

## Artificial Neural Network (ANN)

ANN model, inspired by the brain structure of a human, was used to further predict COD. The Levenberg-Marquardt algorithm was used to train the network. Figure shows the typical architecture of ANN. For the present study we used one hidden layer with fifty neurons.

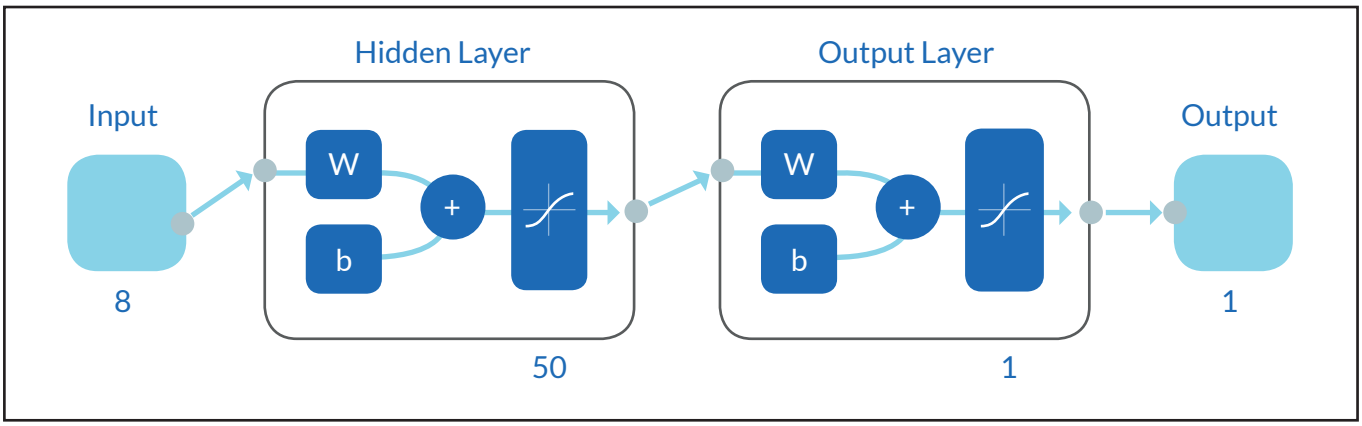



*Figure 50: Diagram depicting the architecture of Artificial Neural Network*

| | | N | R | $R^2$ | RMSE | p-value | R | $R^2$ | RMSE | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| Model - 01 | Train | 160 | 0.794 | 0.631 | 30.1 | 0.000 | 0.757 | 0.574 | 121.9 | 0.000 |
| | Test | 41 | 0.406 | 0.165 | 41.5 | 0.008 | 0.647 | 0.419 | 117.4 | 0.000 |
| Model - 02 | Train | 160 | 0.799 | 0.639 | 30.8 | 0.000 | 0.854 | 0.728 | 109.9 | 0.000 |
| | Test | 41 | 0.369 | 0.136 | 32.7 | 0.017 | 0.491 | 0.2 | 173.7 | 0.001 |
| Model - 03 | Train | 73 | 0.937 | 0.877 | 27.4 | 0.000 | 0.980 | 0.960 | 65.1 | 0.000 |
| | Test | 19 | 0.447 | 0.200 | 49.1 | 0.050 | 0.266 | 0.071 | 322.6 | 0.271 |
| Model - 04 | Train | 73 | 0.994 | 0.989 | 8.7 | 0.000 | 0.991 | 0.983 | 42.2 | 0.000 |
| | Test | 19 | 0.837 | 0.700 | 46.7 | 0.000 | 0.617 | 0.381 | 118.6 | 0.008 |

*Table 20: R, $R^2$, f and p-values resulted from ANN*

In the figure below, a, b, c and d are the scatter plot between BOD measured and predicted BOD using ANN - MLP technique with the four sets of input parameters respectively. Figure- e, f, g and h are the scatter plots between COD measured and predicted COD using ANN - MLP technique method with the four sets of input parameters respectively. The results from ANN-MLP for BOD are able to explain both train and test data sets with the 4th model 99% and 70% data variability. Though this technique is able to explain the data variability for BOD, the RMSE values are on the higher side. Thus, it is not a reliable prediction model. Whereas for the COD, ANN-MLP fails to explain the test datasets variability. The results of predictive analysis using the machine learning techniques are in good progress. For the optimization of these techniques, more data is needed.

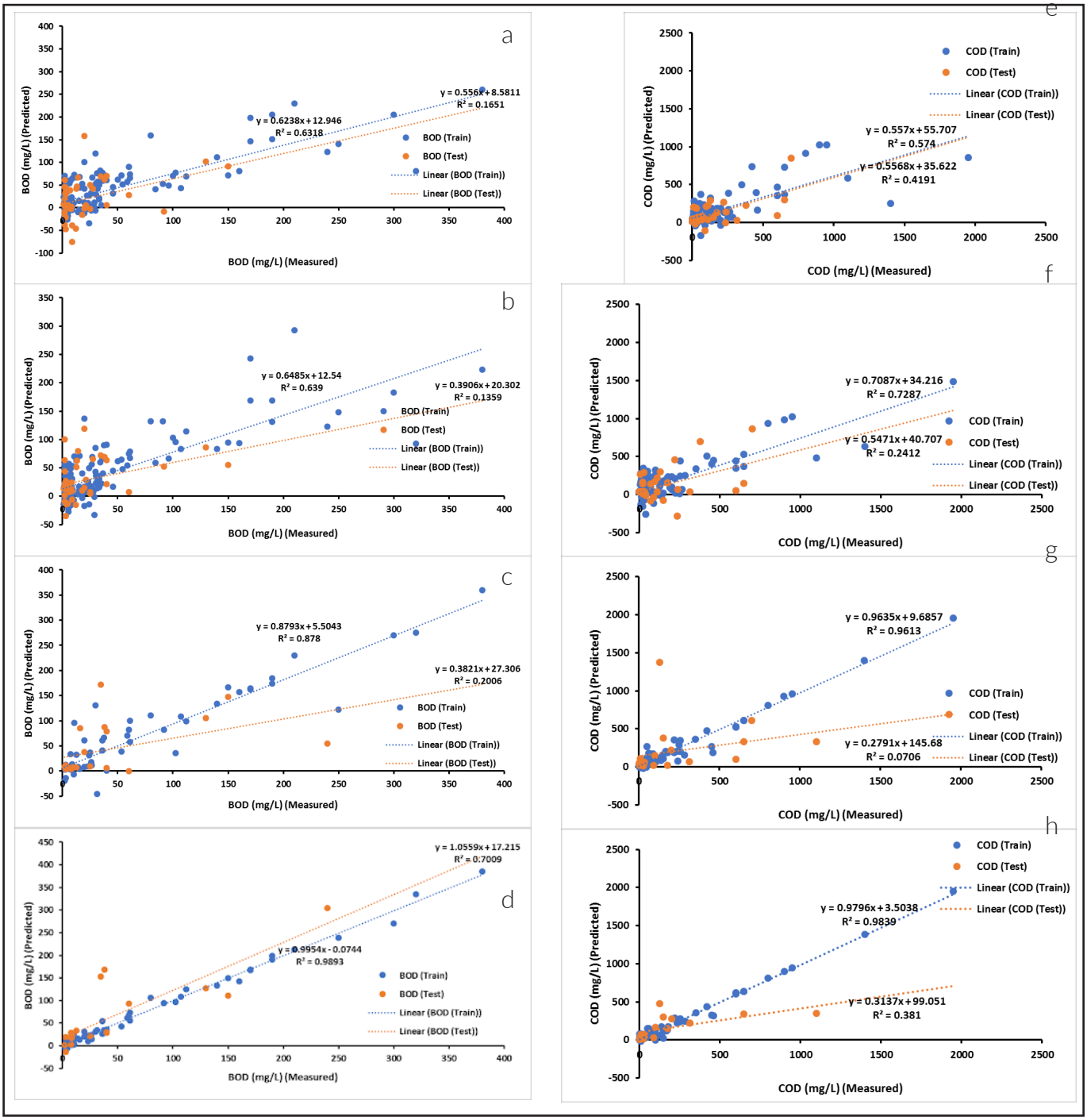



*Figure 51: a,b,c and d are the scatter plots between BOD measured and predicted BOD using ANN-MLP method with the four sets of input parameters; e,f,g and h are the scatter plots between COD measured and predicted COD using ANN-MLP method with the four sets of input parameters*

# 5 Challenges

While non-stationary sensor-based monitoring can provide quick and sufficient data, this approach has its own limitations. The following challenges need to be addressed to ensure adoption of such high resolution mapping techniques for water quality assessments at scale.

## 5.1 Sensor Technology

There are no available commercial sensors for crucial organic parameters such as BOD, COD, total coliform, etc. which may provide accurate values in real time. These are crucial indicators for safe drinking and bathing water and measuring them in labs is time-intensive. Proxy parameters are monitored using sensors such as the Turners C3 and Photonic measurements UV254! but they are expensive and their readings may be affected by turbidity. Thus, sensing technology may need a few more years before this approach can become truly cost-effective.

## 5.2 Calibration and Maintenance

Proper sensor operations rely on precisely calibrated sensors. Calibration is a tedious exercise. Inaccuracies in data may be noticed if the sensors or their accessories (panels, connecting cables, batteries, etc.) are not maintained in good condition. This not only leads to a faulty understanding of water quality by decision makers, but also wastage of resources and efforts spent on conducting the monitoring. Sensor probes for certain parameters such as nitrate and pH have to be replaced from time to time. Long lead times for such probes make it imperative to have backup and plan in advance to not stall operations in between. Repairs and replacement may often not be possible locally. Sending it to another location (usually back to the manufacturer in a different country) may impede the experiments for weeks. All of this could make it a cumbersome process.

## 5.3 Trained Technical Operators

Diligent field staff who are careful in handling sensor equipment must be employed for smooth operations. They must be trained in sensor calibration, maintenance and usage. They must attach sensors properly to the rod before fixing it to the boat. Although sensors are automated, it is advisable for field operators to monitor its readings through the hand held meters or evaluate readings on the laptop to check if all probes are functional and the data is as expected based on observed water conditions on that day. They should be capable of troubleshooting any issue by themselves or through an expert to mitigate any sensor malfunctioning or damage. Availability of such human resources is limited in India today and this could be a bottleneck for scaling.

## 5.4 Logistics

Boat-based measurements are dependent on availability and navigability of boats. Bad weather due to heavy rains or waters with algal bloom may restrict movement of boats. In fact, a boat may not even be available in some areas. Moving to remote-controlled (RC) float boats to carry sensors or unmanned automatic vehicles (UAV) may reduce reliance on these external factors but such vehicles may not be portable which is the USP of this technique.

## 5.5 Ground Situation Analysis

Actionability of data may be limited if water quality data is analyzed in isolation. Metadata on amount of sunlight, rainfall, increased anthropogenic activities on a given day due to a special occasion, unusually active or inactive drains, opening or closing of barrage gates, etc. requires notes from field observations. This information enriches the data collected for better interpretation. However, such data may not be accessible to the public through various government portals or shared internally between departments for ease of analysis.

## 5.6 Acceptable Monitoring Protocol

The non-stationary automated sensor-based approach is fairly recent. While CPCB has a robust protocol in place for manual sampling and laboratory testing, no government protocol exists for non-stationary sensor-based monitoring in a standardized manner.

Adoption of a generic monitoring protocol irrespective of the technique will make it easier to crowd source data from different agencies using diverse methods of monitoring. Our on-ground learnings indicate that such a protocol should account for the following from the perspective of sensor-based monitoring:

- **Site Classification:** Classify sites based on a baseline study of pollution levels and known sources in any area.
  *Class 1:* Critical sites suspected to have multiple pollution hotspots due to continuous waste water inflow from domestic or industrial drain inlets.
  *Class 2:* Sites which are relatively less affected by point sources of pollution.
  *Class 3:* Need based assessment sites.
- **Frequency:** Class 1 sites should be prioritized for once a week mobile sensor-based monitoring. This can assist the authorities in charting out a spatial map of pollution hotspots in the water body. This can be supplemented with lab tests. To understand the trend of incoming waste water pollution and take regulatory action to curb pollution where necessary, they need to be monitored at random times of the day. Class 2 sites should be monitored at least twice a month in the morning and evening. Seasonal variation and any sudden variations are important to be noted at these sites. Class 3 sites may have to be monitored multiple times based on the water management issue to be addressed.

- **Source Apportionment:** Upon identifying waste water discharge points, it is important to monitor pollution trends in drains to identify the exact source of any waste. This would provide evidence to authorities to initiate regulatory action. This information can be used to apply the principles of Extended Producer Responsibility (EPR) to such entities for them to be accountable for their own waste water.

- **Decentralization:** While it is important to set a minimum frequency of monitoring for each site by CPCB, the decision to add new sites and increase the frequency of testing should be with the state pollution control boards to let them decide if a particular stretch may need to be monitored with higher resolution for a few weeks due to specific events or increased pollution temporarily. This flexibility ensures that critical information is collected while minimizing effort and expenditure by resource mobilization.

- **Interoperability:** A validation process should be put in place for comparison and correlation between sensor data and lab data collected at the same point and time with acceptable limits defined by the government. Each sensor manufacturer may need to procure a validation certificate once it is tested to meet these standards.

- **Local Conditions:** Sites near coastal areas may have waters highly affected by tides in the adjoining seas. For such rivers, precaution should be taken to ensure that monitoring happens during low tides to eliminate any backwater effect. Excessive growth of water hyacinths or algal bloom may restrict movement of boat and such areas should be avoided.

- **Metadata:** Information on rainfall, solar irradiance, GPS location, exact time of data collection, data on industrial, agricultural and domestic sources of pollution and the gaps between waste water generated and treated, etc. should be noted and available with the water quality data to increase trustworthiness of such data.

- **Navigation:** It should be ensured that depth of water at the monitoring sites is at least 3 feet to allow navigability of boat and to avoid damage to sensors due to contact with the river bed. Sensor probes should be fixed at a depth of at least 1 feet below the surface of water for accurate measurements. For surface measurements, a depth of 1-3 feet is recommended.

- **Capacity Building:** Field staff should be trained once every six months to ensure retention of knowledge on data collection and testing practices to be followed.

# 6 The Way Forward

Having collected, analyzed and studied water quality data from rivers and drains over the last three years, the study tested the feasibility of an innovative technique for monitoring in the Indian context through this grant. The possibility of getting newer information about pollution in an area and its potential cause was also studied. The next logical step was to customize such analysis for the needs of various stakeholders for them to act upon it.

Through an additional grant from the Tata Centre for Development at UChicago, the study began working towards simplifying reporting of technical data through a macro level understanding of any water body. The Water Quality Index creation was a step in this direction making it easy for laypersons to relate to it. With increased public awareness and civil society interest, the hope is to mainstream water quality issues. Working with IITs and other reputed academic partners, we intend to create a common vocabulary for universal acceptance of data and analysis.

The initial results from the machine learning models applied to water quality data are encouraging. With additional input data from our monitoring exercise as well as from other sources, these models will be further refined with the goal of making data available at scale. As of August 2019, we have collected 325,927 data points from 252 boat sorties at 11 sites. With more data and improved ML models, the water sector will be revolutionized.

Since early 2018, the W2C approach was discussed with the Central Pollution Control Board, National Green Tribunal (NGT) committee members for river pollution, Delhi Pollution Control Committee, National Mission for Clean Ganga (NMCG) and Ministry of Jal Shakti. There is concurrence on the need to make sense of data being collected and utilize it in decision making. A willingness to disseminate data in the form of visualizations such as heatmaps for ease of understanding has been a welcome shift in outlook. The CPCB has, in fact, now created a color-coded, geospatial map based dashboard similar to Water-to-Cloud. The next step is to encourage agencies to share data between departments and with the civil society for increased transparency and accountability.

Various stakeholders have also come together to adopt a systems thinking approach to solving river pollution challenges holistically. The project has set up a Knowledge Management Taskforce as part of a Multi-Stakeholder Platform for Hindon Basin and Ganga Tributary Management by the Uttar Pradesh government on the Public-Private Partnership (PPP) model. 2030 Water Resources Group at World Bank, India-EU Water partnership, State Mission for Clean Ganga and and Uttar Pradesh Pollution Control Board (UPPCB) are some of the other partners to join hands. The project attempts to drive down the point that everyone is part of the problem (albeit to varying degrees) and hence, everyone should be involved in its solution with each stakeholder voluntarily setting targets towards pollution reduction in aligment with the overall vision. W2C will provide unbiased, authentic data for all players to base their decisions on.

Inter-comparison of data from different rivers has facilitated an understanding of pollution effects at various places and the possible linkages between urbanization, industrialization, impact on health and aquatic life, livelihoods and ecology with river water pollution. A study on socio-economic impact of river water pollution has been commissioned with National Council of Applied Economic Research. We held a number of consultations towards co-learning and knowledge sharing.

The vision is to leverage technological advancements to simplify and speed up data collection while automating data dissemination and analysis to facilitate collaborations towards interpreting and acting on the data for a water secure future.

# 7 Conclusion

Water quality continues to deteriorate with increasing population density which heightens the pressure on existing water resources and sanitation infrastructure. Its impacts on health, economic development and ecosystems are far-reaching. A multi-pronged water quality monitoring strategy is the need of the hour to address this challenge and undertake effective mitigation measures.

The design of a reliable and robust Water Quality Monitoring System (WQMS) beyond initial benchmarking is particularly useful to assess results of (planned and ongoing) interventions to mitigate pollution load entering the river at different stretches. In the context of long-term sustainable rejuvenation of any river basin, the WQMS would be expected to yield reliable and sufficient data to assess effectiveness of interventions to mitigate pollution, take corrective measures and guide future action. Consistency and trustworthiness of results and approach are thus crucial. The monitoring protocol should be a hybrid approach; whereby different modes of measurement are used with a definite purpose for each. Once the objective of each monitoring exercise is clear, the most effective approach can be selected from laboratory-based testing, satellite imagery, stationary sensors or mobile sensing technique. The different datasets should be put side by side to compare and complement in order to create a complete picture of river health. Water-to-Cloud system can form an integral part of this multi-pronged water quality monitoring program. With the dipping costs of sensing technologies, a shift to sensor-based monitoring systems is inevitable and has already begun in some developed nations. Reliance on technology and automation at various steps in data collection, processing, analysis and dissemination make it easier to scale up. The availability of visualized and analyzed data enables decision makers to understand and act upon it in a timely manner.

A judicious and surgical use of the Water-to-Cloud system can fill the gaps in the current water quality monitoring program, act as a decision support system for policymakers, and invoke public participation in regulation of water quality. This is, however, only an addition to the repertoire of tools to monitor water quality available with the government, CPCB, scientists and practitioners. Ultimately, the corrective measures taken, regulatory compliance enforced and policy changes implemented using such data help in making progress towards achieving the overarching objective of rejuvenation of any river.

# APPENDIX I: Dissemination of Knowledge and Learnings

The Water-to-Cloud team has published and presented its work at various platforms in India and internationally. We have been invited to several talks to present our work on river water quality monitoring and the Water-to-Cloud system. The team also hosted and participated in multiple workshops and panel discussions. All such publications, talks, presentation, workshops and publications have been listed.

*Conference Publications and Presentations*

*Using Cyber Physical Systems To Map Water Quality Over Large Water Bodies*, P. Hirani, S. Balivada, R. Chauhan, G. Shaikh, L. Murthy, A. Balhara, R. C. Ponduru, H. Sharma, S. Chary, G.B. Subramanyam, S. Randhawa, T. Dutta, H. P. Gupta, A. Gupta, A. Haldar, A. Sarkar, I. Khan, S. Guha, IEEE Sensors Conference, New Delhi, India, 2018.

*A novel approach to pinpoint pollution sources in rivers*, S. Balivada, R. Chauhan, G. Shaikh, L. Murthy, A. Balhara, R. C. Ponduru, H. Sharma, P. Hirani, S. Chary, G.B. Subramanyam, S. Randhawa, T. Dutta, H. P. Gupta, S. Guha, Innovation and Frontier Technology for Water Security and Scarcity - 3rd Regional IWA Diffuse Pollution Conference, Chiang Mai, Thailand, 2018.

*BLUEWATER EYE: Using satellite as a low-cost water pollution sensor: Analytics for deriving long-term pollution insights based on mapping water turbidity*, S. Randhawa, R. Guruprasad, S. Balivada, P. Hirani, S. Guha, SPIE Remote Sensing Conference, Berlin, Germany, 2018.

*Water to Cloud: A cyberphysical sensing system to monitor high frequency temporal and spatial variation in river water quality parameters*, P. Hirani, S. Balivada, R. Chauhan, G. Shaikh, L. Murthy, A. Balhara, R. C. Ponduru, H. Sharma, S. Chary, G.B. Subramanyam, S. Randhawa, T. Dutta, H. P. Gupta, A. Gupta, A. Haldar, A. Sarkar, I. Khan, S. Guha, Water, Waste water and Environmental Monitoring (WWEM) Conference, Telford, UK, 2018.

*Data Visualizations as a tool to bring social change*, R. Chauhan, W. Kent, S. Randhawa, S. Balivada, A. Balhara, H. Sharma, P. Hirani, S. Chary, G.B. Subramanyam, S. Guha, State of the Map (SoTM) Asia conference by Mapbox, Bangalore, India, 2018.

*Thoreau: A subterranean wireless sensing network for agriculture and the environment*, Xufeng Zhang, Arseniy Andreyev, Colleen Zumpf, M. Cristina Negri, Supratik Guha, and Monisha Ghosh, 1st IEEE Workshop on Wireless Communications and Networking in Extreme Environments (WCNEE), Atlanta, USA, 2017.

*Water-to-Cloud: a novel monitoring approach*, S. Balivada, P. Hirani, S. Randhawa, R. Chauhan, G.B. Subramanyam, S. Chary, S. Guha, American Water Resources Association (AWRA) conference, Portland, USA, 2017.

*Talks*

*Mapping Water Quality using Cyberphysical Sensing Systems*, Priyank Hirani, invited talk for a panel discussion on Geospatial IoT for Industrial Applications, IEEE World Forum on Internet of Things, Limerick, Ireland, 2019.

*Leveraging Technology for Informed Decision Making for our Rivers*, Priyank Hirani, guest lecture as part of Symbionts Week at BITS Pilani Goa Campus, Goa, India, 2019.

*Saving our Rivers - Working towards a shared Future*, Priyank Hirani, invited guest session at the UPES - University of Cambridge workshop on 'Sustainable Development: Technology and Beyond', Dehradun, India, 2019.

*Cyberphysical sensing networks for land and water*, Supratik Guha, plenary talk at Sensors Midwest 2017, Chicago, USA, 2017.

*Mapping of River Water Quality*, Priyank Hirani, invited talk at Symposium on 'Filling Data Gaps, Fueling Governance', New Delhi, India, 2017.

*Cyberphysical sensing networks for land and water*, Supratik Guha, invited lecture at Microsoft Corp., Mountain View, California, USA, 2017.

Work presented as part of the 2017 William G. Spitzer Lecture at the University of Southern California, Supratik Guha, 2017.

***Panel Discussions, Workshops and Events***

Himank Sharma moderated the discussion on "Creating avenues for sharing and accessing water quality data" at Water Dialogues, jointly hosted by UC Trust and Dalberg, New Delhi, India, 2019.

Priyank Hirani represented the project as a panelist at Telangana Jagruthi International Youth Leadership Conference on "Good Health, Clean Water", Hyderabad, India, 2019.

Tata Centre for Development (TCD) and International Innovation Corps (IIC) under the University of Chicago Trust held a Roundtable on "Water Quality Monitoring: Driving Decision Making through Actionable Data and Collaborations", New Delhi, India, 2019.

Srinivasa Balivada attended the Indo-UK workshop on Monitoring and analysis strategies for anthropogenic pollutants in environmental and waste waters by Ashoka Trust for Research in Ecology and the Environment (ATREE), Bangalore, India, 2018.

Ritambhara Chauhan and Priyank Hirani were invited to attend the Namami Gange Workshop - NMCG brainstorms on G-governance of Namami Gange programme through Geospatial technology on World GIS Day, New Delhi, India, 2018. part of Symbionts Week at BITS Pilani Goa Campus, Goa, India, 2019.

# APPENDIX II: Using Cyber Physical Systems to Map Water Quality over Large Water Bodies[21]

S. Balivada[1], R. Chauhan[1], G. Shaikh[1], L. Murthy[1], A. Balhara[1], R. C. Ponduru[1], P. Hirani[1], S. Chary[2], G. B. Subramanyam[2], S. Randhawa[3], T. Dutta[4], H. P. Gupta[4], K. Shukla5, M. Malik[5], N. Chowdhary[5], S. Sarkar[5], A. Haldar[6], A. Sarkar[6], I. Khan[6], A. Gupta[6], S. Guha[7]

*[1] University of Chicago Center in Delhi [2] Administrative Staff College of India, Hyderabad*
*[3] IBM Research, Bangalore [4] Indian Institute of Technology-BHU, Varanasi*
*[5] CCK, Ambedkar University Delhi [6] IIEST, Shibpur [7] IME, University of Chicago*

## Abstract

The world faces a grave water risk that affects all aspects of human life and ecology with implications for food security, energy production, industrial activity, and human health. India is particularly affected as it has 16% of the world's population but access to less than 4% of global freshwater resources. In this study, we use mobile (moving) sensors to spatially and temporally map river water quality based on in-situ data gathered in some of India's major rivers. Data visualizations generated are intended to pinpoint sources of pollution, ensure regulatory compliance and examine health of the water body. We show that such cyber physical sensing techniques can be a powerful and more practical (in terms of cost-effectiveness) way to dynamically monitor, predict and regulate the quality of large bodies of water. Keywords—Heatmaps, Water quality, Internet of Things(IoT).

## Introduction

Water plays a vital role in sustaining the Earth's ecosystem. United Nation's Sustainable Development Goal 6 discusses challenges related to water scarcity and security [1]. Availability of water, both in terms of quantity and quality, helps determine its utility function to fulfill the growing demands of an increasing population. Enhanced industrial and agricultural activities also pose a threat to water resources in the form of pollution, be it chemical, microbial or thermal. [2, 3] In order to have the defined utility value for the water source, one needs to take a close look at its quality parameters, which are important for its management. Without having reliable and consistent long-term water quality data, effective preservation and remediation won't be possible for any water body. [4] Traditionally, field measurements for water quality evaluation have depended upon on-site sample collection, and transport to land-based or shipboard laboratories for evaluation, which is a costly as well as time and labor-intensive process. [5] Such methods have been limited in temporal and spatial scales to adequately address factors influencing the development of events such as harmful algal blooms, oxygen depletion and fish kills. [1]

Recent advancements in technology in terms of mobile (moving) sensors can be leveraged for real-time online water quality monitoring to obtain a comprehensive picture. [6] While a stationary sensor-based approach would have ensured continuous monitoring, the costs of the large number of sensors required to monitor a given area do not justify the marginally greater insights that such data may present. Hence, a mobile sensor-based approach justifies lower costs and adequate data availability for a given area.

The intent is to empower researchers and practitioners with high resolution data and technology to capture the quality of river water using various parameters, in different seasons and at diverse locations across the river.

21 Balivada. (2018). IEEE Sensor 2018 Conference. Retrieved from https://www.semanticscholar.org/paper/Using-Cyber-Physical-Systems-to-Map-Water-Quality-Hirani-Balivada/c9eeda802fd36903a86775bdc72af68ece541bc4#extracted

We discuss such a Water-to-Cloud system, that encompasses, end-to-end, the gathering and curation of high resolution data on the Cloud in section II, explain our approach in detail in section III and present important initial results in section IV. Section V draws on important conclusions obtained so far and section VI briefly mentions our future plan of action.

## Relation with IoT and Cloud computing

Our novel approach lies in developing a scalable, low-cost, in-situ monitoring system using mobile sensing platforms carried around pre-selected routes to obtain high frequency temporally and spatially varying water quality data. In this study, we discuss results from the use of such systems across major rivers in India, namely Ganga, Yamuna and Godavari, based on non-stationary mobile sensors for mapping river water quality. Following curation of the geospatial and temporal water quality data, we intend to apply analytic and physics-based techniques to predict the spread of pollution, interpolate sparse data, and carry out inverse analysis to identify pollution sources.

## Methodology

Firstly, the general approach, showing our working model is described in the block diagram below. In section III A, tabulated information regarding the various parameters being considered and their classification is shown. Section III B describes the sensors being used. Section III C provides insights on how the collected data is used and managed to achieve our main objective. Section III D exhibits the process of data visualization and digital curation

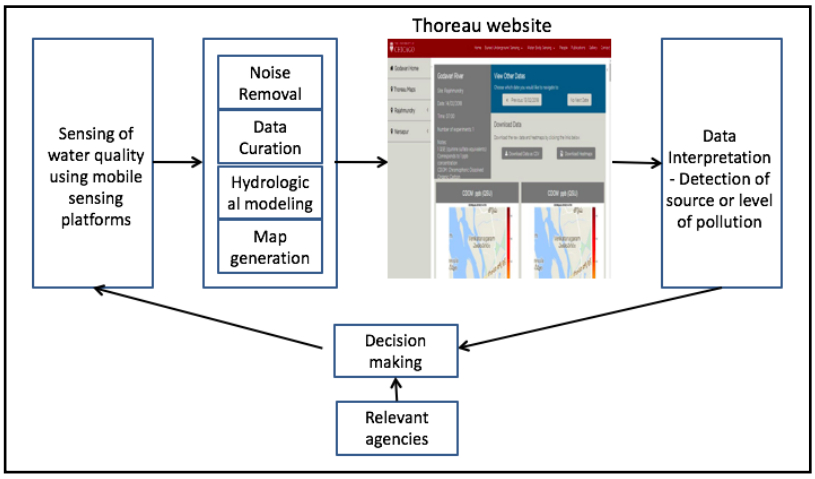



*Figure 1: Block diagram representation of water-to-cloud system*

## Parameters

A number of in-situ parameters being studied are listed in the table below. complete parameters and to study correlations, method is employed in addition to in-situ testing.

| In-situ parameters | Ex-situ parameters |
|---|---|
| pH, Turbidity, Electrical conductivity, Nitrate, Dissolved oxygen(DO), Chlorophyll-a, CDOM (Chromophoric Dissolved Organic Matter), Tryptophan, Temperature, Salinity | pH, Turbidity, Electrical conductivity, Nitrate, Dissolved oxygen(DO), Chlorophyll-a, CDOM (Chromophoric Dissolved Organic Matter), Tryptophan, Temperature, Salinity |

*Table 1: Parameters being considered for this study*

## About the sensors in use

We have employed Hanna Instruments' HI-9829 geo-tagged, time-stamped multi-parametric probes to monitor pH, turbidity, electrical conductivity, dissolved oxygen, temperature, salinity and nitrate ions. Turner Designs' C3 instrument is being used to measure chlorophyll-a content, tryptophan and CDOM.

## Working with data sets

| | |
|---|---|
| **Data Collection** | Sensor-based gathering of data<br>Geo-tagged files in csv format |
| **Data Cleansing** | Noise removal by both manual approach and using the average moving technique |
| **Data Visualization** | geo-tagged csv files used to generate the heatmaps to provide spatial and temporal information |
| **Analysis** | Correlations are determined between in-situ and ex-situ parameters to reduce reliance on labs. |

*Figure 2: Overall working process involved for data management*

## Maps generation

In-house algorithms are run on data to create heatmap visualizations that are digitally curated on an open-source platform, https://thoreau.uchicago.edu.

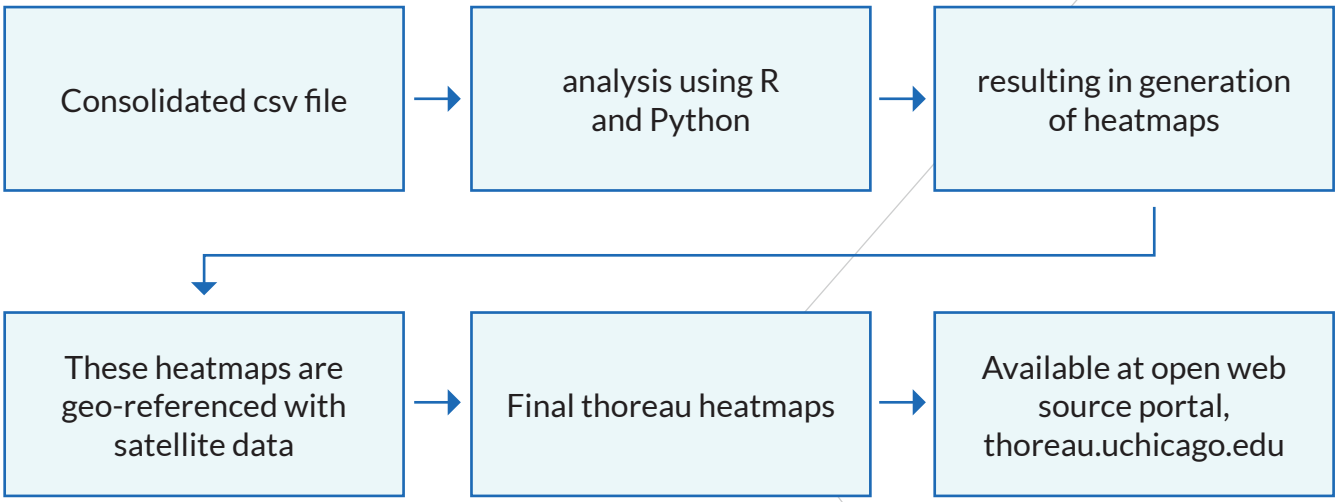



*Figure 3: Map generation and digital data curation process*

## Results and discussion

After making a routing map for the river at selected sites, experimental process is conducted with the help of sensors carried around in water attached to boats. Data collected at regular intervals is curated online, and the relevant interpretations are made upon the basis of heat maps generated from the above data set for a particular river system for selected sites. Over a period of time, several sides of river-water pollution were identified, as discussed below -

## Spatial and temporal quantification of data sets

The visual presentation of water quality parameters in the form of heatmaps helps in understanding the situation better as color gradient variation with respect to the standard scale based on color patterns is more intuitive for humans instead of large, technical tables of data, which restrict the analytical ability for a non-technical eye. In the figure below, the variation in electrical conductivity in River Godavari around Rajahmundry on different days is illustrated to identify local pollution hotspots for each day. Alternatively, one could also compare variations for a parameter across different days via normalization.

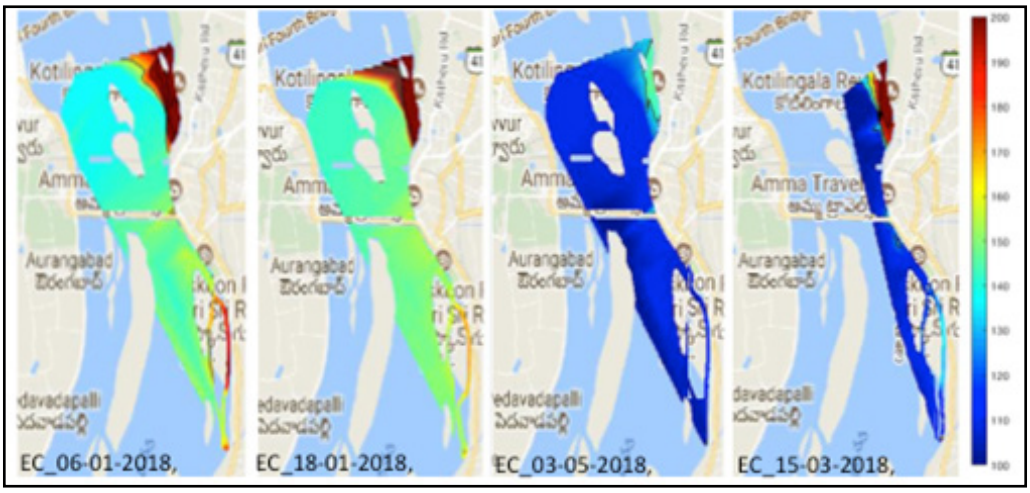



*Figure 4: Temporal Variability in electrical conductivity in River Godavari at*

## Pin-point pollutant sources

Experiments on river Ganga at Varanasi indicate the effect of effluent outlets such as drains as shown in figure 5. The effect of these contaminants occurs within less than the 200m radius in the river, and the pollution is significantly reduced by the dilution achieved beyond 200m and after a few hours of the drain being inactive, which could explain why this effect may be missed via a grab sampling

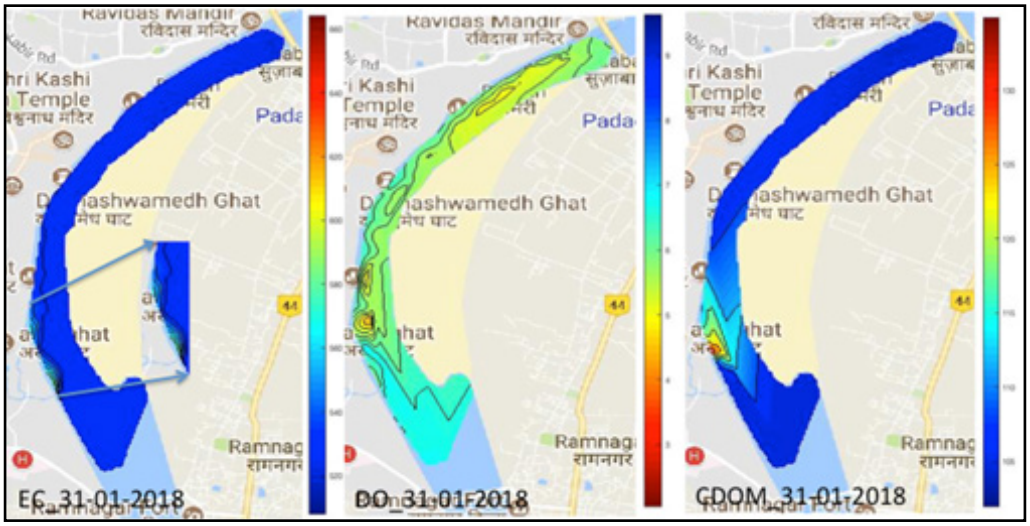



*Figure 5: Spatial variability as observed of EC, DO and CDOM in River Ganga at*

## Predicting Bio-chemical oxygen demand (BOD) and Chemical Oxygen Demand (COD)

Regression analysis and machine learning approaches are being examined for inferring BOD and COD values from parameters monitored in-situ. It has required the collection of large water sample

sets that are measured both via laboratory measurements as well as in-situ measurements so that models can be trained on this data. Such an approach, if successful, would eliminate the need for cost-intensive and time consuming (3-5 days) tests in lab once the prediction models are validated.

### Demonstrate effect of contaminant load on a water body

In order to conduct a one-week baseline study on the Yamuna, a 15 km route was identified at New Delhi. The flow of water from upstream is regulated, and water is released once every few days, as can be seen in Figure –6 from left to right in terms of electrical conductivity. Fresh water was released a day before the start of the experiment and not again till the end of it. Increasing trend in electrical conductivity values from day one to day four of experiments indicates the pollution load into the river. Foul smell increases progressively with incoming waste water from drains daily. Such information could help in effective water management and can aid in creating 2D and 3D pollutant dispersion models.

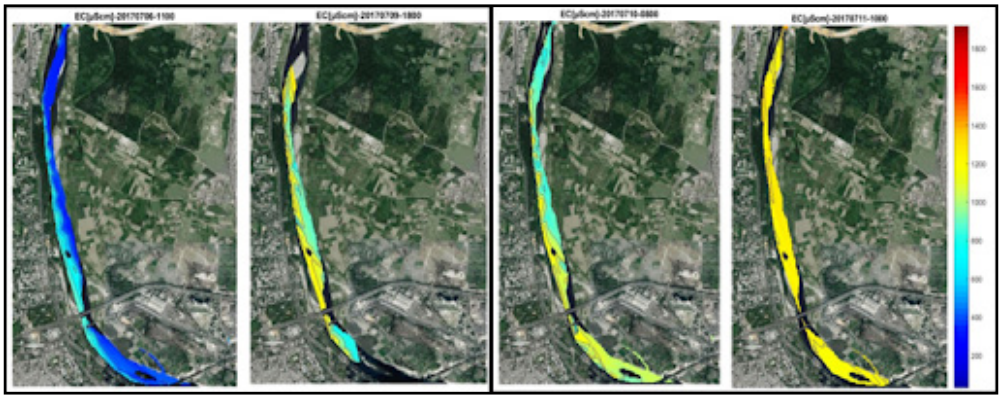


*Figure 6: Spatial and temporal variability of Electrical Conductivity in Yamuna at New Delhi (µS/ cm)*

### Understanding the concurrent effect of the different parameters

The visual presentation of the data helps in demonstrating the association or disassociation of the various parameters with each other, which helps in providing insights on how the river system is working in different dynamic situations. For example, during experiments at Narsapur in River Godavari, which is 10km from the sea, variation within Dissolved Oxygen (DO) levels in association with salinity and organic waste loading rate, which keeps on altering the river dynamics within the interval of few hours highlighted the effect of waste water drains when measured during morning or evening hours but failed to identify the same around noon due to high tides.

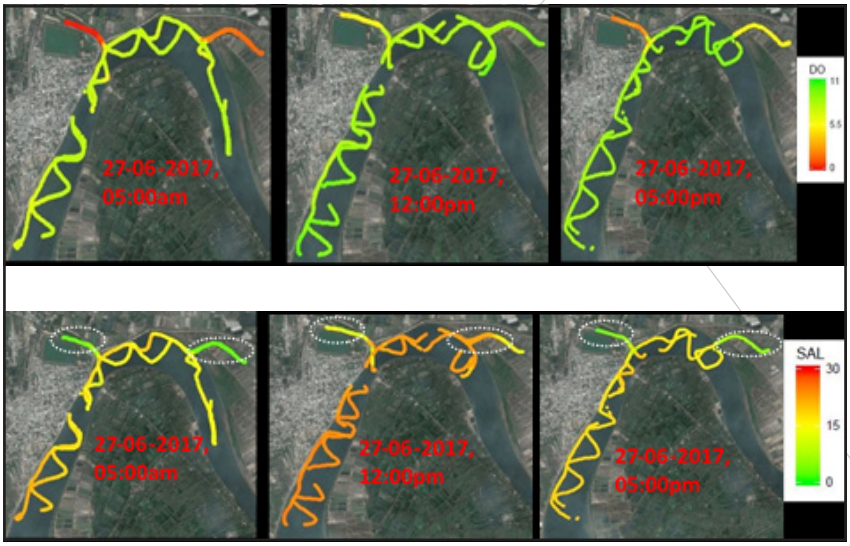



*Figure 7: Variation in salinity and DO levels within a single day in River Godavari at Narsapur*

## Defining water utility based on available data

Based on the concentration of pollutants at specific ghats in Varanasi or Rajahmundry, a layman could choose a safe site for the religious dip or swimming in these places. For example, Gauthami Ghat in Rajahmundry, a pilgrimage site associated with the holy dip, is often more polluted than nearby ghats due to high anthropogenic activities around it and should be avoided by people with sensitive bodies.

## Conclusion

From the above results, it can be concluded that Water-to-Cloud system can address key issues such as spatial distribution of pollutants, pinpointing contaminant sources, ensuring regulatory compliance for the concerned authorities, and benchmarking remediation measures; where traditional grab sampling method would fail to capture these features due to its limited approach. Moreover, the availability of raw data through an open source platform in near real time ensures transparency and accountability. Such data would contribute to effective policy discussions and ensure its use in decision making.

## Way forward

Above study could be scaled up in order to understand the seasonal fluctuation of any river system. We also plan to extend the representation of the water bodies in terms of deriving overall water quality indices for each river which could serve as a general method for easy comparison amongst different rivers and also facilitate quick actionable measures by relevant regulatory bodies. In order to predict a river's health status in association with measuring both longitudinal and lateral mixing of the pollutants within the system and to develop an associated hydrological model for the predication purpose in the future, collecting such high frequency spatial and temporal data is a first step.

We intend to validate our BOD and COD prediction models and create similar ones for total coliform and fecal coliform estimations over the next year.

## References

[1]Sustainable development goal 6 (2017), retrived on 2018, June 10, https://sustainabledevelopment.un.org/sdg6

[2]Vörösmarty CJ, Green P, Salisbury J, Lammers BR. Global Water Resources : Vulnerability from Climate Change and Population Growth. Science. 2000; 289: 284–288. PMID: 10894773

[3]Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, et al. Global threats to human water security and river biodiversity. Nature. 2010; 467: 555–561. doi: 10.1038/nature09440 PMID: 20882010

[4] T. P. Burt,*a N. J. K. Howden,b F. Worrallc and M. J. Whelanb (2010) Long-term monitoring of river water nitrate: how much data do we need? Journal of Environmental Monitoring 2010, 12, 71–79.[5]

C. Niu, Y. Zhang, Y. Zhou, K. Shi, X. Liu and B. Qin, "The Potential Applications of Real-Time Monitoring of Water Quality in a Large Shallow Lake (Lake Taihu, China) Using a Chromophoric Dissolved Organic Matter Fluorescence Sensor", Sensors, vol. 14, no. 7, pp. 11580-11594,2014.

[6] S. Randhawa, S. S. Sandha and B. Srivastava, "A Multi-sensor Process for In-Situ Monitoring of Water Pollution in Rivers or Lakes for High-Resolution Quantitative and Qualitative Water Quality Data," IEEE Intl Conference on

# APPENDIX III: Team Profile

Professor Supratik Guha | Principal Investigator

Dr Guha is the Professor at the Pritzker School for Molecular Engineering (PME), The University of Chicago; Director of the Nanoscience and Technology Division and the Center for Nanoscale Materials at the Argonne National Laboratory. He joined Argonne in 2015 after spending 20 years at IBM Research where he last served as the Director of Physical Sciences. At IBM, Dr. Guha pioneered the materials research that led to IBM's high dielectric constant metal gate transistor, one of the most significant developments in silicon microelectronics technology, and used in over 50% of smart phones in use today. He was also responsible for initiating or significantly expanding IBM's R&D programs in cyberphysical systems and sensor networks.

Dr. Guha is a member of the U.S. National Academy of Engineering and a Fellow of the Materials Research Society, American Physical Society and the recipient of the 2015 Prize for Industrial Applications of Physics. He received his Ph.D. in materials science in 1991 from the University of S. California, and a B.Tech in 1985 from the Indian Institute of Technology, Kharagpur. His research interests are focused on sensing, cyberphysical systems and future information processing.

Priyank Hirani | Program Director

Priyank Hirani received his MS in Integrated Circuit Design from Imperial College London in 2013. After graduation, he worked for Mullard Space Science Lab in UK before returning to India in 2015. Through various research endeavors, he has been exposed to the fields of water & sanitation, agriculture, energy, wireless communication, biotechnology and space science. He hopes to use this diverse experience in attempting to solve real world issues now. Priyank pursued the Young India Fellowship in 2015 and enhanced his understanding of contemporary India, its heritage, its problems and its international relations.

Having worked with the Andhra Pradesh government on sanitation issues in 2016-17, he has developed an understanding of the nuances of grassroots implementation of large-scale public projects. Priyank leads the operations for the project, coordinating between the science and engineering team in India and the US.

Himank Sharma | Project Lead

Himank graduated in Liberal Studies from the Young India Fellowship at Ashoka University in 2016. Prior to that he majored in Electronics and Electrical Communication engineering and worked as a Network Consulting Engineer with Cisco for a couple of years. Since 2016, Himank has been working with TechnoServe in the international development space to devise business solutions for poverty. He worked at the grassroots in India, Ghana and Kenya with a focus on improving livelihood opportunities for farmers through business and technology. Continuing his streak of solving real world problems, he is now working with Water-to-Cloud to explore practical applications of Thoreau technology to solve the challenge of sanitation in India.

Himank takes care of managing the team and operations.

### Dr Srinivasa Rao Balivada | Water Research Lead

Dr Srinivasa Rao Balivada holds a Ph.D in Chemistry from Andhra University, Visakhapatnam. Prior to joining this project, he was a research fellow at Centre for Studies on Bay of Bengal at Andhra University. He has also worked on Ministry of Earth Science projects on impact of oceanographic parameters on fishery resources. He has nine national and international journal publications and has presented his work at various conferences. He has research experience on water quality indices and instrumentation for water characterization along with handling lab & field equipment for the same. His research interests are focused on retrieval of water quality parameters from the satellite and significance of trace elements on metabolic disorders.

Dr Balivada's contribution to the project is in terms of his expertise in understanding water parameters and analyzing the data obtained for relevant science research.

### Ritambhara Chauhan | Technical Lead

Ritambhara Chauhan graduated in Computer Science and Engineering from IIT Hyderabad. During here association with Bank of America Continuum, Mumbai, she worked on Risk Management reports. She has served as a student counselor in IIT Hyderabad for 3 years. She specializes in writing problem-solving algorithms and using multiple technologies including C, C++, Python, R, Databases, etc. She takes interest in real world projects and is interested in using her technological and problem-solving skills in the social development sector. Ritambhara's other skills include sketching, caricature art, sculpture making and singing. She is also an avid reader and a blogger.

She supports the project by looking at data curation and web development activities along with automation of operations for a seamless end user experience.

### Gopi Shaik | Field Assistant

Gopi holds a MSc in Chemistry from Andhra University. With an understanding of various water parameters and experience with testing water quality, he is responsible for in-situ data collection and liaisoning with laboratories for sample analysis as required. He also calibrates various instruments and looks into maintenance issues from time to time.

### Lova Murthy | Field Assistant

Murthy holds an MSc in Environmental Chemistry from Andhra University. He is responsible for in-situ data collection and liaisoning with laboratories for sample analysis in this project.

### Ziqi Ma | Data Analyst

Ziqi is a rising third-year undergraduate student at the University of Chicago, majoring in Computer Science and Statistics (expected June 2020). She joined the Guha group in December 2017 and has been working on LoRa sensor data transfer and embedded systems programming.

### William Kent | Web Developer

William is an undergraduate student of Molecular Engineering at the University of Chicago. He is involved in developing software for data mapping and visualization, and in developing the web interface.

Jacob Gold | Web Developer

Jacob is an undergraduate student of Molecular Engineering at the University of Chicago. He is involved in developing software for data mapping and visualization, and in developing the web interface.

# APPENDIX IV: Use based Classification of Surface Waters in India as per CPCB

| Designated-best-Use/ Beneficial Use | Classification of water | Criteria |
|---|---|---|
| Drinking water source without conventional treatment but after disinfection | A | 1. Total Coliforms Organism MPN/100 ml shall be 50 or less<br>2. pH between 6.5 and 8.5<br>3. Dissolved Oxygen 6 mg/l or more<br>4. Biochemical Oxygen Demand 5 days 20 oC 2 mg/l or less |
| Outdoor bathing (organised) | B | 1. Total Coliforms Organism MPN/100 ml shall be 500 or less<br>2. pH between 6.5 and 8.5<br>3. Dissolved Oxygen 5 mg/l or more<br>4. Biochemical Oxygen Demand 5 days 20 oC 3 mg/l or less |
| Drinking water source after conventional treatment and disinfection | C | 1. Total Coliforms Organism MPN/100 ml shall be 5000 or less<br>2. pH between 6 and 9<br>3. Dissolved Oxygen 4 mg/l or more<br>4. Biochemical Oxygen Demand 5 days 20 oC 3 mg/l or less |
| Propagation of wildlife and fisheries | D | 1. pH between 6.5 and 8.5<br>2. Dissolved Oxygen 4 mg/l or more<br>3. Free Ammonia (as N) 1.2 mg/l or less |
| Irrigation, industrial cooling, controlled waste disposal | E | 1. pH between 6.0 and 8.5<br>2. Electrical Conductivity at 25 oC micro mhos/ cm maximum 2250<br>3. Sodium absorption ratio maximum 26<br>4. Boron maximum 2 mg/l |

# Appendix V: Water Quality Parameters

## pH

Based on the CPCB guidelines, the range of pH for surface water should be between 6.5 and 8.5. The pH exceeding the range might affect aquatic life, especially the adult fishes, as it may cause damage to the outer part including gills and eyes. However, alkaline pH might also affect ammonia concentration in river water. The low pH <6.5 could influence the growth of invasive aquatic species of planktons and disappearance of fishes such as smallmouth bass. If the pH goes <5, it is unsuitable for aquatic animals and increases the solubility of trace elements such as Aluminium (Al) in the soil close to the bank. High Al and other trace elements might restrict the growth and once accumulated can inter into the human food chain. In acidic condition the Calcium (Ca) concentration is found to be low which restricts the reproductive process in fish. If the water is used for drinking purpose, pH has an important effect on human metabolism and health as human body consists of 50–60 % water. The pH level of our body fluid should be in the range of 7–7.2. If pH is less than 5.3, it restricts the assimilation of vitamins or minerals.

## Electrical conductivity (EC)

Electrical conductivity (EC) is the amount of ions present in the water. It also indicates the amount of total dissolved solids and both of these are positively correlated i.e. more dissolved solid indicates more electrical conductivity. If the water is used for drinking and other domestic uses high EC can cause the laxative effect, however since drinking water is also a source of major ions in the human body and hence low EC might cause deficiency of ions. In agriculture, high EC might cause deposition of salt on the surface of soil which might reduce fertility and make the soil surface hard with low oxygen levels. In river ecosystems, high EC due to the presence of any organic chemical i.e. fuels, detergents, paints, solvents, etc. imparts objectionable and offensive tastes, odors and colors to fish and aquatic plants even when they are present in low concentrations. As noted earlier, dissolved ions affect the pH of water, which in turn may influence the health of aquatic species as discuss earlier.

## Dissolved Oxygen (DO)

An adequate supply of oxygen in dissolved form is essential for the survival of aquatic life. Hence, DO is considered as an indicator of aquatic health. Concentration of DO in surface water bodies depends on the temperature, turbulence of water, salinity, surface area and atmospheric pressure. DO in water also gets influenced by the atmospheric dissolution and autotrophic and heterotrophic process which consume or produce O2 in water systems. Concentration of DO also determines the changes made by abiotic or biotic organisms and hence measurement of DO is essential. A rapid fall in the DO levels indicates high organic pollution in the river. The optimum value for good water quality is 4 to 6 mg/l of DO, which ensures healthy aquatic life in a water body.

## Bio-chemical oxygen demand (BOD)

BOD is a measure of determining the requirement of oxygen needed by the aerobic micro-organism to stabilize the household and industrial wastes. To stabilize these organic wastes the minimum DO level should be 2-7 mg/L then only the BOD can be estimated in Laboratory condition at fixed temperature i.e. 20 °C for 5 days. High BOD indicates rapid fall in DO with potential threat on aquatic biodiversity. A major reason for high BOD

might be the high levels of organic pollution, caused usually by poorly treated waste water or high nitrate levels, which trigger high plant growth. For drinking water BOD should not exceed 6 mg/L however BOD > 3mg/L is not found feasible for the growth and survival of fish and other aquatic life.

### Chemical oxygen demand (COD)

Chemical oxygen demand (COD) is the amount of oxygen utilized for chemical oxidation of organic waste to inorganic end products. It is an indicative measure of the amount of oxidizable organic pollutant in water. Similar to BOD high COD is an indicator of high organic material and consumption of high DO to stabilize it resulting rapid depletion in DO level. A reduction in DO can lead to anaerobic conditions, which are harmful to aquatic life forms. The COD test is often used as an alternative to BOD due to shorter length of testing time. Hence COD values can be used to estimate the BOD and reduce the time.

### Tryptophan

Organic matter in the fresh water consists of plant materials including both fulvic and humic substances. The organic matter also has an active microbial community which consumes DO resulting into increase in the BOD concentration in water. The protein in the cell wall of these microbes has similar fluoresces like amino acids i.e. “Tryptophan” and hence tryptophan like fluorescence is used to measure the microbial health, fecal coliform concentration and BOD of surface water.

### Colored Dissolved Organic Matter (CDOM)

CDOM are a mixture of colored organic compounds essential for the biochemical cycle in the freshwater. These are also called yellow substances. They control the nutrient availability, light penetration and ecosystem productivity. High CDOM inhibits the growth of phytoplankton resulting in decrease in DO. Pure water absorbs the long wavelength red light whereas CDOM absorbs the short wavelength light ranging from blue to UV. So the water with no/less CDOM has blue color. CDOM represents the large proportion of dissolved total organic matter including terrestrially derived humic and fulvic acids. Variations in CDOM are a result of natural processes including changes in the amount and frequency of precipitation, human activities such as logging, agriculture, effluent discharge, and wetland drainage can affect CDOM levels in fresh water systems.

### Chlorophyll-a

Chlorophyll-a is the photosynthetic pigment responsible for the green color of the plant leaves and algal biomass. The concentration of Chlorophyll-a directly indicates the amount of photosynthetic community or algae in the water. Existence of chlorophyll-a concentration in phytoplankton and the phytoplankton spectra reflectance is related to the optical activity of the pigment, the composition and appearance of the algal cells. Chlorophyll, in various forms, is bound within the living cells of phytoplankton found in surface water and accounts for 1-2 percent of dry weight of planktonic algae. Chlorophyll is an important pigment in marine and freshwater algal species and has always been used as an indicator for phytoplankton biomass as well as bio-production calculation in water bodies.

# Appendix VI: Calibration of Sensor Probes

Calibration is the process that standardizes the electrical or optical signals from the sensors to reagent standards of known value. A major objective of calibration is to check the accuracy and efficiency of the instrument. In field research, the calibration of the instrument became more important as the system and surrounding change frequently, which may affect the precision of the instrument. To minimize the drift the instrument is calibrated against the references or known set of parameters.

The probes used in this study mostly have intuitive and menu driven calibration. HI 9829's calibration routines are accessed by highlighting "Calibration" and pressing <Select> from the main menu shown in the meters. All calibration data is stored in the non-volatile probe memory, allowing probes to be connected to different meters without recalibration. The user may also restore each parameter to a factory default calibration.

For calibration of HANNA probes we followed the steps and standard solutions provided in the user manual provided with the sensors[22]. We perform two types of calibration:

- Quick calibration is done every time on the site for parameters-pH, Dissolved Oxygen, Electrical Conductivity.

- Single parameter calibration which is done with the help of standard stock solutions provided by HANNA only as per the standard procedure stated in the user manual for the selected parameter probe.

### Frequency of calibration

Generally, individual parameter calibration is done while fixing the new sensors in the probe structure. While, we perform individual calibration for pH, Electrical Conductivity, Nitrate at every 8-9 months interval, for DO is calibration is done every 2-3 months. As DO is more sensitive, so to maintain the accuracy we calibrate the probe every time we change the location.

| Condition | Parameter | Calibration Type | Frequency |
|---|---|---|---|
| Fist time installation of probes | pH<br>Dissolved Oxygen<br>Conductivity<br>Turbidity\|<br>Temperature | Single point calibration | Once |
| Currently using sets of probes | pH<br>Dissolved Oxygen | Single point calibration | 8-9 months<br>2-3 months |
| Changing study site frequently | pH<br>Dissolved Oxygen | Single point calibration | 4 months<br>2 months |
| On experiment site | pH<br>Dissolved Oxygen<br>Conductivity | Quick Calibration | Always |

22 Hanna Instruments (2012). HI9289 Multiparameter meter Instruction Manuall. Retrieved from https://hannainst.in/downloads/dl/file/id/1035/hi9829_2012_04_13.pdf

## Quick calibration

As the field probes are sensitive to surrounding system, environment, atmospheric conditions, pressure, humidity, etc. the quick calibration of the instrument is performed every time before boat ride. The instrument is calibrated for pH, Dissolved Oxygen, Electrical Conductivity using the standard quick calibration solution provided by the HANNA and following are the prescribed guidelines as given below:

Solution used: HI 9828-25 (For both pH and electrical conductivity)

- Fill the calibration beaker 2/3 full with HI 9828-25 calibration solution
- Slowly place the sensors into the solution and dislodge bubbles that may adhere to the sensors
- Screw the calibration beaker completely on the probe body. (Some solution may overflow)
- Wait few minutes for the system to stabilize
- From the “Calibration” menu select “Quick calibration”
- A three item calibration menu will appear (pH, Conductivity and Dissolved oxygen) and “pH” will start to blink along with the “Not ready” message
- When the pH signal is stable, the “Ready” message appears. Press <Confirm> to store the calibration data.
- The “Storing” message will appear as the calibration proceeds to the next sensor. A checkmark will appear in the box next to “pH” to indicate Successful calibration
- Following the pH calibration, “Conductivity” will start to blink along with the “Not ready” message
- When the measurement is stable, “Ready” appears. Press <Confirm> to store the calibration data and the “Storing” message will appear
- For DO calibration the message “Empty the beaker” will appear
- Unscrew the calibration beaker and empty the solution
- Shake any remaining liquid off the probe and beaker. (Note: No droplets should remain on the DO sensor membrane)
- Screw the empty calibration beaker on the probe body. The beaker should not be dry.
- Precaution: Do not attempt to dry/wipe the DO sensor as damage to the membrane may occur.
- Press <Accept> to close the displayed message.
- When the measurement is stable, “Ready” appears. Press <Confirm> to store the calibration data and the “Storing” message will appear.
- Press <OK> to return to “Calibration” menu.

Quick Calibration is successfully done.

## Single parameter calibration - pH

Preparation:

- From the “Calibration” menu select “Single param. calibration” and then “pH calibration”.
- The display shows two options: “Calibrate pH” and “Restore factory calib”.
- If a new pH sensor has been installed use “Restore factory calib.” (Warning message from previous calibration).
- If “Restore Factory calib” is selected, all user calibration data will be deleted and the factory default calibration is restored. A user calibration should follow immediately.
- If “Calibrate pH” is selected, the user can perform a new calibration using up to 3 buffers (pH 4.01, 6.86, 7.01, 9.18, 10.01 or one custom buffer).When a three point calibration is performed, all old

data areoverwritten, while with a single or two-point calibration the meter will also use information from the previous calibration.

Precaution: To minimize cross contamination, use two beakers for each buffer solution: the first one for rinsing the sensor and the second one for calibration.

Procedure:
- The measured pH value is displayed, along with the temperature and the buffer value on the second level. If necessary, press the <Cal point> soft key and use the arrow keys to select the correct buffer.
- Immerse the pH sensor in the first buffer rinse solution and stir gently.
- Immerse the pH sensor and temperature probe into the selected buffer and stir gently. The temperature, pH buffer value and the "Not ready" message is displayed.
- Once the reading has stabilized the countdown timer will count down until the display shows the "Ready" message.
- Press <Confirm> to accept the calibration point.
- To avoid cross-contamination immerse the sensors in the next calibration buffer rinse solution and stir gently.
- Press <Cal Point> to select the next buffer (if necessary), and repeat the calibration procedure outlined above with the second and third buffers. The message "Storing" followed by "Calibration completed" will be displayed.
- Press <OK> to return to the Calibration menu. Press <Measure> to return to the measurement screen.

### Single parameter calibration-Dissolved Oxygen

The accuracy of dissolved oxygen measurements is directly related to membrane cleanliness and calibration technique. Oily coating and biological contaminants are the primary cause of calibration drift. However, use of brushes or other cleaning objects may damage the membrane. Replacing the membrane cap and electrolyte is the best way to perform periodic maintenance. Although it may be easier to calibrate the DO sensor prior to deployment, it is advised to calibrate at the site of deployment. Errors in measurement may result if altitude and barometric pressure differ between the calibration and measurement site. This is important for autonomously logging probes.

Note: Perform either the % DO Saturation or DO Concentration calibration.

We prefer the % DO saturation. Dissolved oxygen concentration values are based on % DO saturation, temperature, salinity and atmospheric pressure.

Choose "DO calibration" from the "Calibration" menu, select the DO calibration type using the arrow keys and press <Select> to confirm.

% DO saturation: The calibration of the % DO saturation range can be performed at a single or 2 standard points (0 % and 100 %), or at a single custom point (50 % to 500 %).

Procedure:
- To calibrate at 100 %, fill the calibration beaker with approximately 4 mm (5/32") of water and screw it onto the probe. The membrane should not be wet. This condition corresponds to air 100 % saturated with oxygen and water vapour.
- The reading, temperature, calibration point and "Not ready" message is displayed.
- Once stabilized the timer will count down until the display shows "Ready" message.
- Press <Confirm> to accept the calibration point, put the DO and temperature sensors into HI 7040L

zero oxygen solution and wait for stability to be reached. The stability timer will count down and <Confirm> will appear. Press <Confirm> to store the calibration.
•The “Storing” and “Calibration completed” message will appear. Press <OK> to return to the “Calibration” menu.
•Press ESC twice to return to the main menu. Press <Measure> to return to the measurement screen.

### Conductivity calibration

For conductivity calibration, choose “Conductivity calibration” from the “Calibration” menu, select the calibration type using the arrow keys and press <Select> to confirm.

Procedure:
•Select the “Conductivity” option and press <Select> to confirm.
•Fill the calibration beaker with a conductivity standard of various potassium chloride concentration solution.
•Pour additional standard into a second beaker to be used to rinse the sensor.
•Immerse the sensor into the rinse standard by raising and lowering the beaker a few times to ensure that the EC sensor channels are filled with fresh standard.
•Place the calibration beaker over the EC sensor and dislodge any trapped bubbles. Screw the beaker into place. Wait for the reading to stabilize.

### Turbidity calibration

The turbidity calibration is done by Custom calibration with standard Formazin standards, using the custom point. The standard has been prepared based upon published international protocols.
The first point is zero while the second point is 10 and 30 FNU. The third point used is in between 100 and 500 FNU.

Procedure:
•Pour standard solutions into clean beakers for rinse.
•Fill the HI 7698293 calibration beaker with the zero standard and follow the procedure for calibrating zero.
•After confirming zero, clean out the calibration beaker and refill with custom point 2 FNU standard. Immerse the sensor in the Point 2 custom FNU rinse beaker and then shake off excess solution. Place the sensor into the Point 2 custom FNU calibration beaker and take measures to eliminate bubbles from the optical area.
•Press <Cal. point> and T9 key pad to display the standards value. Press <Accept>.
•When the reading is stable the display shows the “Ready” message.
•Press <Confirm> to accept the second calibration point. The calibration can be stopped after the 2nd Custom FNU point.
•Press ESC. Storing will be displayed followed by Calibration completed. Press <OK> or <Measure>

### Temperature calibration

The probe is factory calibrated for temperature readings. The user can perform a single point temperature calibration or restore factory calibration. This procedure requires a reference temperature measuring instrument.

Procedure:
•Select “Temperature” from the “Calibration” menu.
•Select “Calibrate temperature”.

•Insert the probe in an isothermal bath with reference instrument and allow the probe to come to thermal equilibrium.
•Use the keypad to enter the known temperature and then press <Accept> to confirm.
•The stability timer will count down and the message "Ready" and <Confirm> will be displayed.
•Press <Confirm> to store the calibration point.

C3-turner calibration

The C-Soft Windows-based user interface allows for intuitive calibration, data logging, set up, and download functions[23]. Users can calibrate sensors in either the direct concentration mode, raw fluorescence mode, or blank subtracted raw fluorescence mode. We calibrated sensors by the direct concentration mode.

Raw Fluorescence Calibration-Blank Subtracted
Following steps 1-5 below will ensure values are reported as Blank subtracted Relative Fluorescence

*Step 1: Blanking*
•Place the Submersible Fluorometer in a blank solution.
•Wait until "Current RFU" readings have stabilized then select the "Set Blank" button to set the blank value.
•Wait for the blank value to be measured and set for all 3 gains.

Step 2: Standard Value and Units
•Place the Submersible Fluorometer in a standard solution or use a solid secondary standard.
•Enter the known concentration value of the standard solution or solid standard in the "Standard Value" box. If the concentration is not known, enter an arbitrary value.
•Enter manually or select desired units from the dropdown menu.

*Step 3: Set Standard*
•When readings in "Current RFU" box have stabilized, select the "Set Standard" button.

*Step 4: Optional Temperature Compensation*
Temperature compensation is optional; it enables the Submersible Fluorometer to compensate for changes in fluorescence due to varying temperatures as detected by the temperature sensor. Temperature compensation is available for in vivo Chlorophyll and Rhodamine sensors only. Temperature is displayed in degree Celsius.
•Click on the "Temperature Compensation" box.
•Select either in vivo Chlorophyll or Rhodamine from the dropdown menu.
•Data will now be corrected for temperature fluctuations. The temperature compensation coefficients for each fluorophore are listed in the table below.

*Step 5: Save Calibration*
•After calibration is complete, select "Save Calibration".
•Data are now displayed as concentration estimates.

Frequency of calibration
C3- turner calibration is one-time procedure. We preformed the calibration as per its study site requirements.

23 Turner Designs (2018). C3 & C6p submersible fluorometers User's Manual. Retrieved from http://docs.turnerdesigns.com/t2/doc/manuals/998-2300.pdf

## Stock solution preparation

Stock Solution preparation for standards absorbance values for calibration purpose for c3 turner:

***Step 1:*** Take a spinach leaf in 90% of Acetone solution overnight after crushing the spinach. To separate some amount of this solution in different stock (1ppb-10ppb) concentration to find out the absorbance of the standard values of RFU for these stocks.

***Step 2:*** Draw plot between the Concentration and Absorbance (RFU).

***Step 3:*** Then standard concentration values as added to the c-soft installed in the computer system. (Please refer to c3-turner manual for the installation and working of c-soft as per the system compatibility).

Total Organic Carbon (TOC) Calibration and Dissolved Organic Carbon (DOC) Calibration[24]:

- Firstly, it is recommended using TOC and DOC reagents and Photometer from manufactures such as Lovibond and Hach.
- Take at least 3 samples of the water and split into two.
- 3 Samples for the GO!
- 3 for Photometer/Spectrophotometer
- Perform the measurement of TOC using the procedure from the TOC and DOC reagent manufacture.
- Start the Go! by using the switch on the back of the device. Press Calibrate
- Select a location where you will store the calibration on for example Location1
- Select a Parameter TOC
- Once complete, select Generate Curve
- Prepare a reference sample of deionised water into a cuvette.
- Then, click the Reference button
- Keep the deionised water reference sample in the GO! and then click Measure
- Click measure and then enter the value for sample 1.
- Continue this process for sample 2 and sample 3.
- Three points will appear on the graph along with the point at 0.0mg/L from the reference.
- Once all calibration points have been measured click on Review
- Lastly, press Complete to store your calibration.

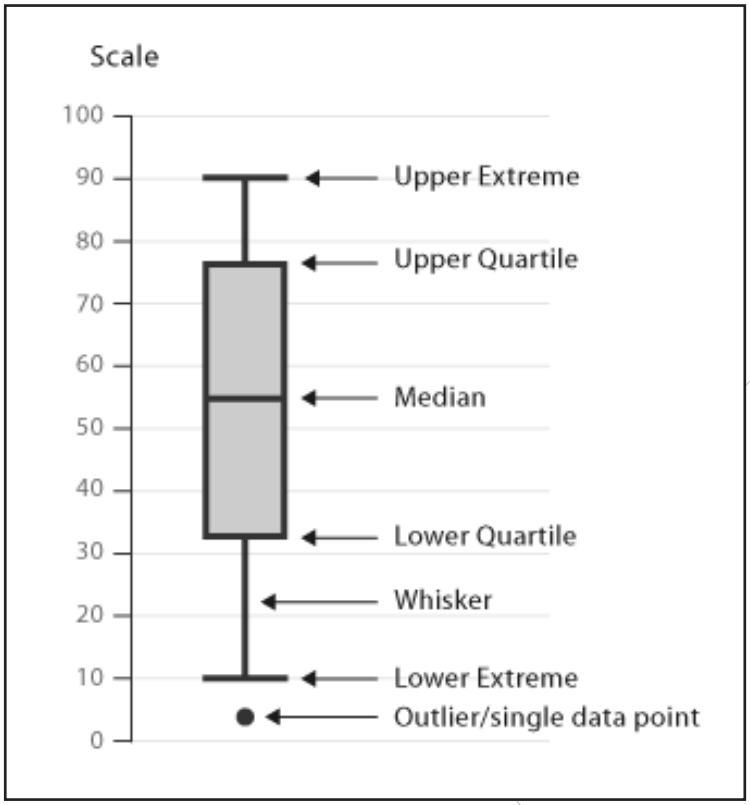
24 Photonic Measurements (2018). Photonic User's Manual. Retrieved from https://photonicmeasurements.com/products

# APPENDIX VII: Statistical Tools

## Box Plots

Box plots are a convenient way of visually displaying the data distribution through their quartiles. The lines extending parallel from the boxes are known as the "whiskers", which are used to indicate variability outside the upper and lower quartiles. Outliers are sometimes plotted as individual dots that are in-line with whiskers.

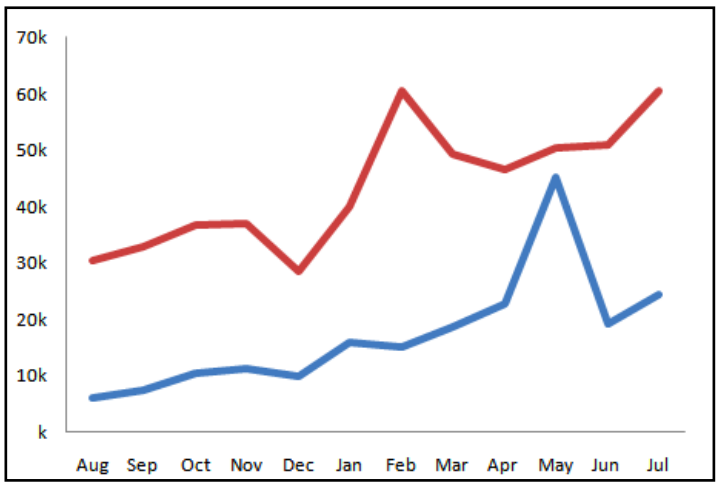
Here are the types of observations one can make from viewing a Box Plot:
•What are the key values such as: the average, median 25th percentile, etc?
•Are there any outliers and what are their values?
•How tightly is the data grouped?
•Is the data skewed, and if so, in what direction?




Box plots are a great way to understand water quality parameters collected in a varying frequency over a substantial time period. A comparison between different monitoring locations shows how a parameter behaves at the site. This helps us understand the sites that are more critical and plan for necessary action.

## Time Trends

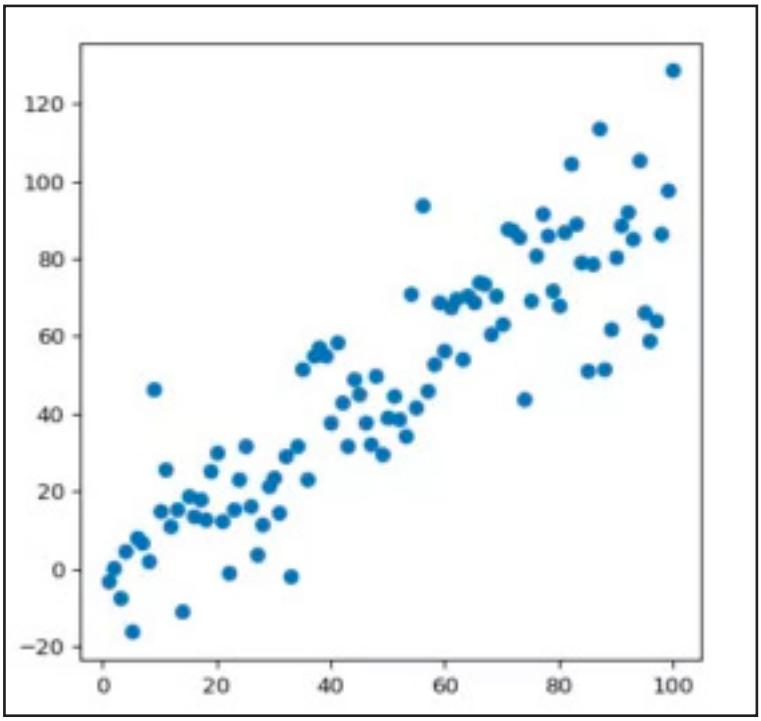
A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive points in time. Thus, it is a sequence of discrete-time data.




Analysis of time series data on water quality can help in understanding the characteristics of such data. We can understand the following:

- Patterns of change in an indicator over time– for example whether DO or EC increase or decrease in a water body with days or seasons.
- Comparing one time period to another time period-for example, evaluating the impact of a sanitation intervention by comparing water quality results before and after the event.
- Comparing one geographical area to another – for example, how the time trends for a parameter compare between Narsapur and Rajahmundry stretch of Godavari
- Making future projections-for example to aid the planning of healthcare services by estimating likely resource requirements.

## Scatter Plots




A scatter plot is a two-dimensional data visualization that uses dots to represent the values obtained for two different variables - one plotted along the x-axis and the other plotted along the y-axis.

Scatter plots are used when you want to show the relationship between two variables. Scatter plots are sometimes called correlation plots because they show how two variables are correlated. For example, when we collect data from same location at same time using two different methods (sensor and lab sampling), drawing a scatter plot helps us understand if both the data sets correlate with each other. This is a powerful tool used to validate the accuracy of sensor data by comparing it to lab data.

Scatter plots are also useful to understand how two different parameters are correlated with each other. For example, are DO and Chlorophyll related to each other in anyway; inversely or directly? Such questions can be answered using scatter plots.

# APPENDIX VIII: Heavy Metal Analysis

Drinking water guidelines by BIS 2014 for heavy metals (NR=No relaxation)

| Parameters | Acceptable (ppb) | Permissible (ppb) |
|---|---|---|
| Aluminium (Al) | 30 | 200 |
| Barium (Ba) | 700 | NR |
| Calcium (Ca) | 75000 | 200000 |
| Copper (Cu) | 50 | 1500 |
| Iron | 300 | NR |
| Magnesium (Mg) | 30000 | 100000 |
| Manganese (Mn) | 100 | 300 |
| Seleneium (Se) | 10 | NR |
| Silver (Ag) | 100 | NR |
| Zinc (Zn) | 5000 | 15000 |
| Cadmium (Cd) | 3 | NR |
| Lead (Pb) | 10 | NR |
| Mercury (Hg) | 1 | NR |
| Molybdenum (Mo) | 70 | NR |
| Nickel (Ni) | 20 | NR |
| Total Arsenic (As) | 10 | 50 |
| Chromium (Cr) | 50 | NR |

Irrigation water quality guidelines from Central Water Commission 2018

| Element | Long-term (mg/L) | Short term (mg/L) |
|---|---|---|
| Al | 5000 | 20000 |
| Be | 100 | 5000 |
| Cd | 10 | 50 |
| Cr | 100 | 1000 |
| Co | 50 | 5000 |
| Cu | 200 | 5000 |
| Fe | 5000 | 20000 |
| Pb | 5000 | 10000 |
| Li | 2500 | 2500 |
| Mn | 200 | 10000 |
| Mo | 10 | 50 |
| Ni | 200 | 2000 |
| Se | 20 | 20 |
| V | 100 | 1000 |
| Zn | 2000 | 10000 |

In Narsapur, heavy metal sampling was carried out at the following locations as shown in the picture: Location 1: Channel 1 of the river-coming from Palacole (contains waste water from palacole and agricultural run-off); Location 2: Channel 2 meets River Godavari; Location 3: Close to the solid waste dumping site.

The results from these locations for samples exceeding limits have been analyzed below.

**Heavy Metals exceeding limits in Narsapur**

| | Heavy metals violating limits | Al | Fe | Mn | Ni | Zn | Pb | Cu |
|---|---|---|---|---|---|---|---|---|
| Location 1 | N | 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| | Max Value | 1263.5 | 1872.5 | 1709.3 | 5.7 | 31927.8 | 46.4 | 12712.1 |
| | Median Value | 221.9 | 618.1 | 212.2 | 1.6 | 505.5 | 11.4 | 2647.8 |
| | % Violation (drinking) | 56.5 | 73.9 | 30.4 | 0 | 4.3 | 43.4 | 43.4 |
| | % Violation (irrigation) | 0 | 0 | NA | 0 | 4.3 | 0 | 43.4 |
| Location 2 | N | 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| | Max Value | 2491.3 | 3996.9 | 856.9 | 3351.8 | 1833.1 | 26.6 | 8506.9 |
| | Median Value | 221.6 | 1264.3 | 175.9 | 1.6 | 697.4 | 11.3 | 4934.3 |
| | % Violation (drinking) | 47.8 | 65.2 | 45.4 | 4.3 | 0 | 39.1 | 43.4 |
| | % Violation (irrigation) | 0 | 0 | NA | 0 | 0 | 0 | 43.4 |
| Location 3 | N | 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| | Max Value | 1326.8 | 4414.8 | 471.5 | 101.6 | 13427.4 | 55.3 | 7786.4 |
| | Median Value | 195.3 | 724.2 | 147.5 | 0.9 | 740.9 | 9.8 | 4664.4 |
| | % Violation (drinking) | 43.4 | 69.5 | 18.1 | 4.3 | 0 | 30.4 | 47.8 |
| | % Violation (irrigation) | 0 | 0 | NA | 0 | 0 | 0 | 47.8 |

In Kovvur- Rajahmundry stretch, the sampling was carried at the followinglocations as shown in the picture:

Location 1: Close to the point source i.e. discharge from paper mill; Location 2: Middle of the river; Location 3: Close to Ava drain

The results from these locations for samples exceeding limits have been analyzed below.

| Heavy Metals exceeding limits in Kovvur - Rajahmundry | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Heavy metals violating limits | Al | Fe | Mn | Ni | Zn | Pb | Cu |
| Location 1 | N | 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| | Max Value | 1478.95 | 1524.67 | 107.61 | 2.38 | 32.2656 | 121.38 | 1.8526 |
| | Median Value | 21.98 | 82.82 | 11.89 | 0.3 | 1.2444 | 0.1 | 0.793 |
| | % Violation (drinking) | 13.6 | 9 | 0 | 0 | 0 | 4.5 | 0 |
| | % Violation (irrigation) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Location 2 | N | 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| | Max Value | 1892.3 | 1965.61 | 66.79 | 13.41 | 99.633 | 2.9 | 7.7041 |
| | Median Value | 33.4 | 76.66 | 11.85 | 0.29 | 8.0165 | 0.07 | 0.884 |
| | % Violation (drinking) | 22.7 | 13.6 | 0 | 4.5 | 0 | 0 | 0 |
| | % Violation (irrigation) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Location 3 | N | 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| | Max Value | 1498.44 | 2864.14 | 371 | 6.28 | 83.4225 | 5.99 | 34.2013 |
| | Median Value | 25.33 | 78.85 | 22.69 | 0.47 | 4.1236 | 0.12 | 0.9084 |
| | % Violation (drinking) | 18.1 | 36.3 | 4.5 | 0 | 0 | 0 | 0 |
| | % Violation (irrigation) | 0 | 0 | NA | 0 | 0 | 0 | 0 |

Water-To-Cloud

www.thoreau.uchicago.edu/water/home

THE UNIVERSITY OF CHICAGO | Center in Delhi

*UChicago Center in Delhi, Ground Floor, DLF Capitol Point Building, Baba Kharak Singh Marg, Near Hanuman Temple, Connaught Place, New Delhi -*

PRITZKER SCHOOL OF MOLECULAR ENGINEERING
THE UNIVERSITY OF CHICAGO

*Pritzker School of Molecular Engineering at University of Chicago, 5640 S Ellis Ave, Chicago, IL 60637, USA*