सत्यमेव जयते

National Mission for Clean Ganga

Ministry of Jal Shakti

Department of Water Resources, River Development & Ganga Rejuvenation
Government of India

NOVEL SENSOR-BASED WATER QUALITY MEASUREMENTS FOR PUBLIC PURPOSES: HOW RELIABLE ARE THEY?

cGanga

Centre for Ganga River Basin Management and Studies

National Mission for Clean Ganga (NMCG)

NMCG is the implementation wing of National Ganga Council which was setup in October 2016 under the River Ganga Authority order 2016. Initially NMCG was registered as a society on 12th August 2011 under the Societies Registration Act 1860. It acted as implementation arm of National Ganga River Basin Authority (NGRBA) which was constituted under the provisions of the Environment (Protection) Act (EPA) 1986. NGRBA has since been dissolved with effect from the 7th October 2016, consequent to constitution of National Council for Rejuvenation, Protection and Management of River Ganga (referred to as National Ganga Council).

www.nmcg.in

Centre for Ganga River Basin Management and Studies (cGanga)

cGanga is a think tank formed under the aegis of NMCG, and one of its stated objectives is to make India a world leader in river and water science. The Centre is headquartered at IIT Kanpur and has representation from most leading science and technological institutes of the country. cGanga's mandate is to serve as think-tank in implementation and dynamic evolution of Ganga River Basin Management Plan (GRBMP) prepared by the Consortium of 7 IITs. In addition to this it is also responsible for introducing new technologies, innovations and solutions into India.

www.cganga.org

Suggested Citation

@cGanga and NMCG, 2022, Novel Sensor-Based Water Quality Measurements for Public Purposes: How Reliable are They?

Contacts

Centre for Ganga River Basin Management and Studies (cGanga)
Indian Institute of Technology Kanpur, Kanpur 208 016, Uttar Pradesh, India

or

National Mission for Clean Ganga (NMCG)
Major Dhyan Chand National Stadium, New Delhi 110 002, India

Author

Vinod Tare, Founding Head, cGanga, IIT Kanpur

Contributors

Gautam Roy, cGanga, IIT Kanpur
Vishal Kapoor, cGanga, IIT Kanpur
Rahul Ramachandran, cGanga, IIT Kanpur
Tushar Deshpande, cGanga, IIT Kanpur
Akanksha Chaudhari, IIT Kanpur

NOVEL SENSOR-BASED WATER QUALITY MEASUREMENTS FOR PUBLIC PURPOSES: HOW RELIABLE ARE THEY?

DECEMBER 2022

cGanga

Centre for Ganga River Basin Management and Studies

PREFACE

Pursuant to the Ganga River Basin Management Plan (GRBMP-2015) submitted by IIT Consortium to the National Mission for Clean Ganga (NMCG), Ministry of Jal Shakti (then Ministry of Water Resources, River Development and Ganga Rejuvenation) in 2015, the Centre for Ganga River Basin Management and Studies ("cGanga") was established in IIT Kanpur to provide state-of-the-art inputs to specific problems faced by the government in implementing the GRBMP for River Ganga's rejuvenation and allied issues via a Memorandum of Agreement between the Ministry of Water Resources, River Development and Ganga Rejuvenation (MoWR, RD & GR), GoI (now Ministry of Jal Shakti) and IIT, Kanpur signed in March 2016 for "Continual Scientific Support in the Implementation and Dynamic Evolution of the Ganga River Basin Management Plan".

In keeping with this goal of cGanga, the National Mission for Clean Ganga (NMCG), among other works, had assigned the task of assessing the effectiveness of implementation of the Central Pollution Control Board's (CPCB's) 2015 Charter for Water Recycling & Pollution Prevention in Pulp and Paper Industry (PPI) in Uttar Pradesh and Uttarakhand. cGanga's report on the same had been satisfactorily submitted in June 2019, but the study had raised many doubts about the reliability of real-time sensor-based monitoring of PPI discharges. To clear the doubts and issues concerned, further work was carried out on technical and statistical scrutiny of the PPIs' data plus the sensor-based flow and water quality measurements for different measuring stations on River Ganga and its major tributaries and drains over several years procured from CPCB.

This report describes the objectives and strategic plan, methodology, site information, data collection and analysis, results, suggestions and recommendations of the efficacy and reliability of the CPCB-mandated novel sensor-based measurements for water quality data of rivers, waterbodies and effluent discharges. The overall assessment is viewed from the perspective of authentic data needs for comprehensive natural resource assessment in India.

There are two associated aspects to the outcome of this report that need mentioning. On the one hand, dedicated members of cGanga spent many months diligently studying, surveying, analysing and discussing various aspects of the monitoring data. On the other hand, many premier institutions including IITK, IITR, IITD, PPIs, CPCB, SPCB - UK & UP, and NMCG's interacted with us and contributed to this report during various phases of the study. This report is therefore very much the outcome of a cooperative effort of cGanga with the numerous stakeholders of Ganga River Basin's. This collaborative dedication led to the comprehensiveness of this report to a degree that may extend its usefulness well beyond its immediate purpose.

VINOD TARE
Emeritus Fellow & Founding Head,
cGanga, IIT Kanpur

EXECUTIVE SUMMARY

Natural resource management in India and other countries depends on authentic resource data. Comprehensive and reliable data for different aspects of renewable natural resources of rivers and waterbodies are fundamental to optimal natural resource use while ensuring healthy aquatic ecosystems. In the present study, CPCB data of optical sensor-based water quality parameters for 36 Real-Time Water Quality Monitoring Stations on River Ganga and its major tributaries and drains over a 4½ year period between 2017 and 2021 were analyzed and evaluated for their validity and reliability based on scientific scrutiny and statistical tests. The results showed that, firstly, the plots of BOD versus COD data show high levels of scatter, poor correlation, and high root mean square errors (RMSE) at most stations, as well as large variations in correlation coefficients (and of COD:BOD ratio) between different river stations, which indicate poor data quality and erroneous data. Likewise, analysis of paired TOC and COD (as well as paired TOC and BOD) data showed wide scatter, poor correlation and high RMSE even for river stations. More surprisingly, the COD vs. TOC scatter plots showed them to lie on one (or a few) perfect straight lines in many cases, suggesting either TOC or COD may have been simply computed as some proportion of the other instead of any actual measurement of that parameter.

The above doubts on the quality of sensor-based data were reinforced by direct comparison of such measurements of effluent discharges from Pulp and Paper Industries in U.P. and Uttarakhand states (mandated by CPCB) and measurements using standard methods by cGanga. Not only were the sensor-based measurements of BOD and COD found to have high scatter and RMSE values in comparison to standard method measurements, in many cases they were also found to plot in a narrow horizontal band with little variation in range whatsoever. These findings clearly show that the sensor-based monitoring data deviate strongly from measurements by standard methods and are largely erroneous.

The above results confirm what may be surmised from basic considerations, viz.: (i) real-time optical sensor-based measurement may be flawed from scientific principles for not just BOD (which is a slow biological-process parameter), but for many water quality parameters of non-homogeneous and variable quality natural waters, (ii) the new methods adopted are neither recommended by international agencies nor used by advanced countries for such measurements, and (iii) no information has been released by CPCB, the regulatory agency that introduced these methods, divulging any test reports or validity of the methods. The generation of likely spurious natural resource data by such novel measurement methods at public expense and for public purposes is certainly unwarranted.

In the broader context of natural resource management in India – not just for sensibly managing our rivers and waterbodies – the above issue calls for an urgent and clearly defined policy/ protocol for introducing new and non-standard methods of natural resource measurements in India.

NOVEL SENSOR-BASED WATER QUALITY MEASUREMENTS FOR PUBLIC PURPOSES: HOW RELIABLE ARE THEY?

1.0 Introduction

Renewable natural resources such as water, soil, nutrients (organic and inorganic), and energy are fundamental needs of all terrestrial ecosystems besides their secondary human usage having significant economic value. Comprehensive natural resource management is essential in modern times since the anthropogenic exploitation of these resources needs to be optimal or sub-optimal; otherwise, if they become scarce or are degraded due to human impacts, then all ecosystems – urban, agricultural and other manmade systems included – would be threatened, and both human lives and other terrestrial lives would be affected in return. To varying extents natural resource management is therefore being carried out today in most countries including India, both for the purpose of optimal use of healthy ecosystems as well as to revive degraded ecosystems. In both of these cases, the first basic need to enable natural resource management is that of comprehensive and reliable data pertaining to different renewable resources in order to estimate the quantity and quality of resource availability for human usage.

Rivers and freshwater bodies are key ecosystems that support a multitude of human needs such as agriculture, industry, urban and rural habitats, livelihoods, healthcare, pollution mitigation, and protection against natural disasters. These ecosystems embody a host of renewable resources including water, sediments, nutrients, and biodiversity, of which water is the most visible and basic resource that enables other natural resources to develop and/or accumulate. The measurement of different aspects of the major natural resources of rivers and waterbodies is, therefore, fundamental to generating the information needed to adopt suitable measures for comprehensive natural resource use while ensuring healthy aquatic ecosystems.

Data collection through field measurements is a time-consuming task, conventionally involving large-scale human engagement. With rapid technological advancements in recent times, data collection too has become much more rapid and economic by saving on the need for human involvement through increasingly mechanized and/or automated data measurement devices, making it feasible to collect much more data within short timeframes than was possible earlier. This may have also increased the reliability of data insofar as the scope for human error gets reduced. But it also introduces the possibility of measurement errors which may be more difficult to detect due to the complexities of such instruments/ technologies. Hence, automated and semi-automated devices need to be

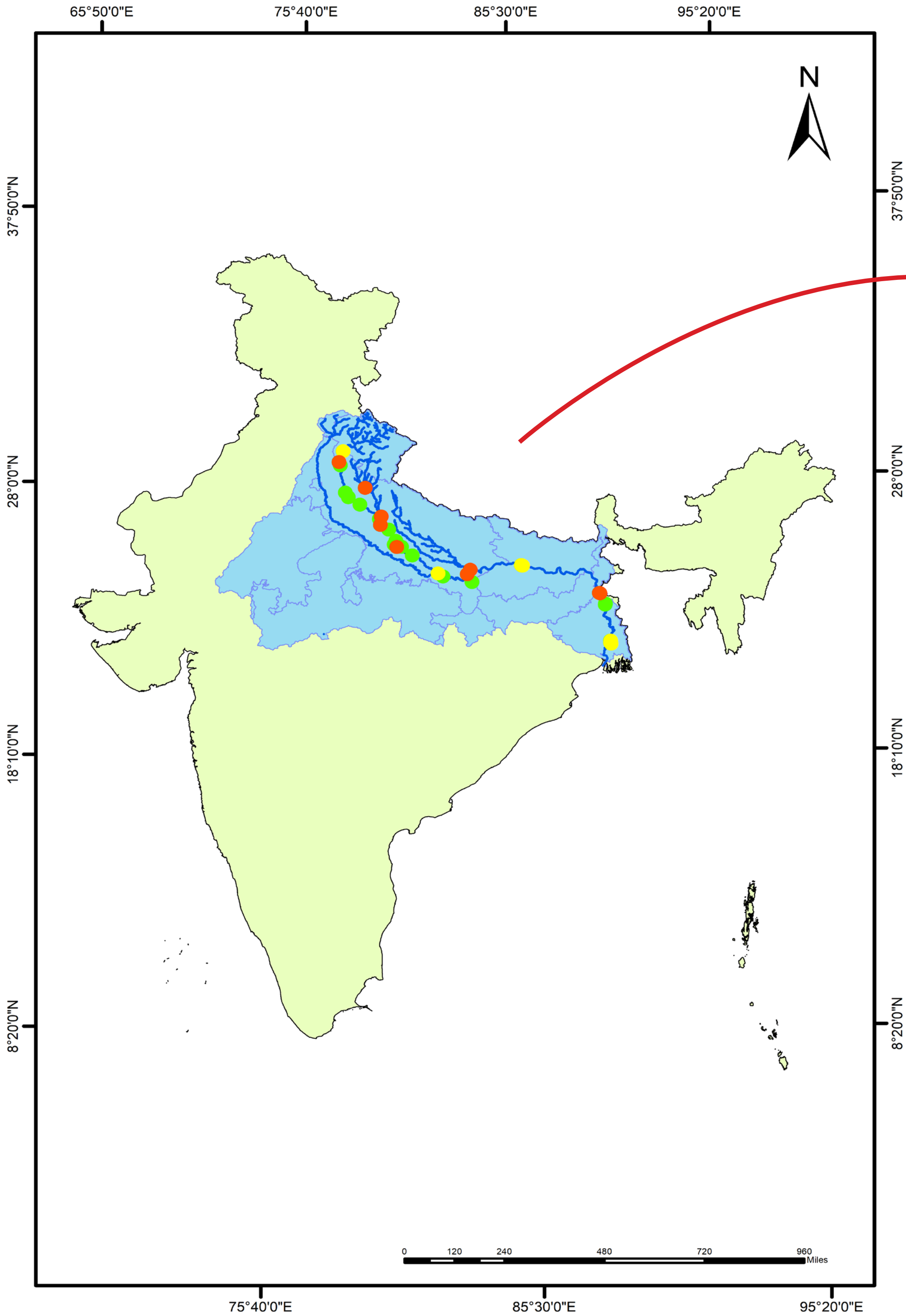
The automated sensors can transfer the measured data immediately to specific computers/ servers, but they leave open the question of reliability of the data generated

introduced with caution, by verifying their applicability for different types of data and establishing suitable protocols for their use. They also may need to be periodically cross-checked with standard data measurement procedures until the technology and its limitations are well understood and documented.

It is in the above backdrop that water quality measurements by real-time optical sensors (based on UV spectrometry or fluorescence) introduced by the Central Pollution Control Board (CPCB) for real-time monitoring of river waters and industrial effluents in the last decade that there was a felt need to examine the appropriateness and reliability of the new measurement systems adopted for water quality parameters for regulatory or public needs. The automated sensors used for this purpose can transfer the measured data immediately to specific computers/ servers for real-time recording and analysis, but they leave open the question of reliability of the data generated since information about the verification and standardization of the measurement technologies and their applications were unavailable from CPCB.

2.0 Automated Sensor-Based Data Acquisition for River/ Drain Waters

For the present study, CPCB data has been procured for 36 Real-Time Water Quality Monitoring Stations (RTWQMS) for the period 11th March 2017 to 30th September 2021. The frequency of the data is hourly, and data were obtained for 17 water quality parameters such as DO, BOD, COD, TSS, TOC, Color, BTX, Temperature, NH_4-N, NO_3, Turbidity, pH, Potassium, Flouride, Chloride, Conductivity and Level , Details of the data acquisition method, sensors, parameters, and stations are given in the following sub-sections. Overview of stations under this study in the Ganga River basin are shown in Figure 1, and sample photos of the monitoring stations are shown in Figures 2 and 3. Details of all 36 RTWQMS (18 Ganga River stations + 9 tributary rivers stations + 9 Drain stations) are presented in Table 1.



65°50'0"E
75°40'0"E
85°30'0"E
95°20'0"E
N
37°50'0"N
28°0'0"N
18°10'0"N
8°20'0"N
0
120
240
480
720
960
Miles

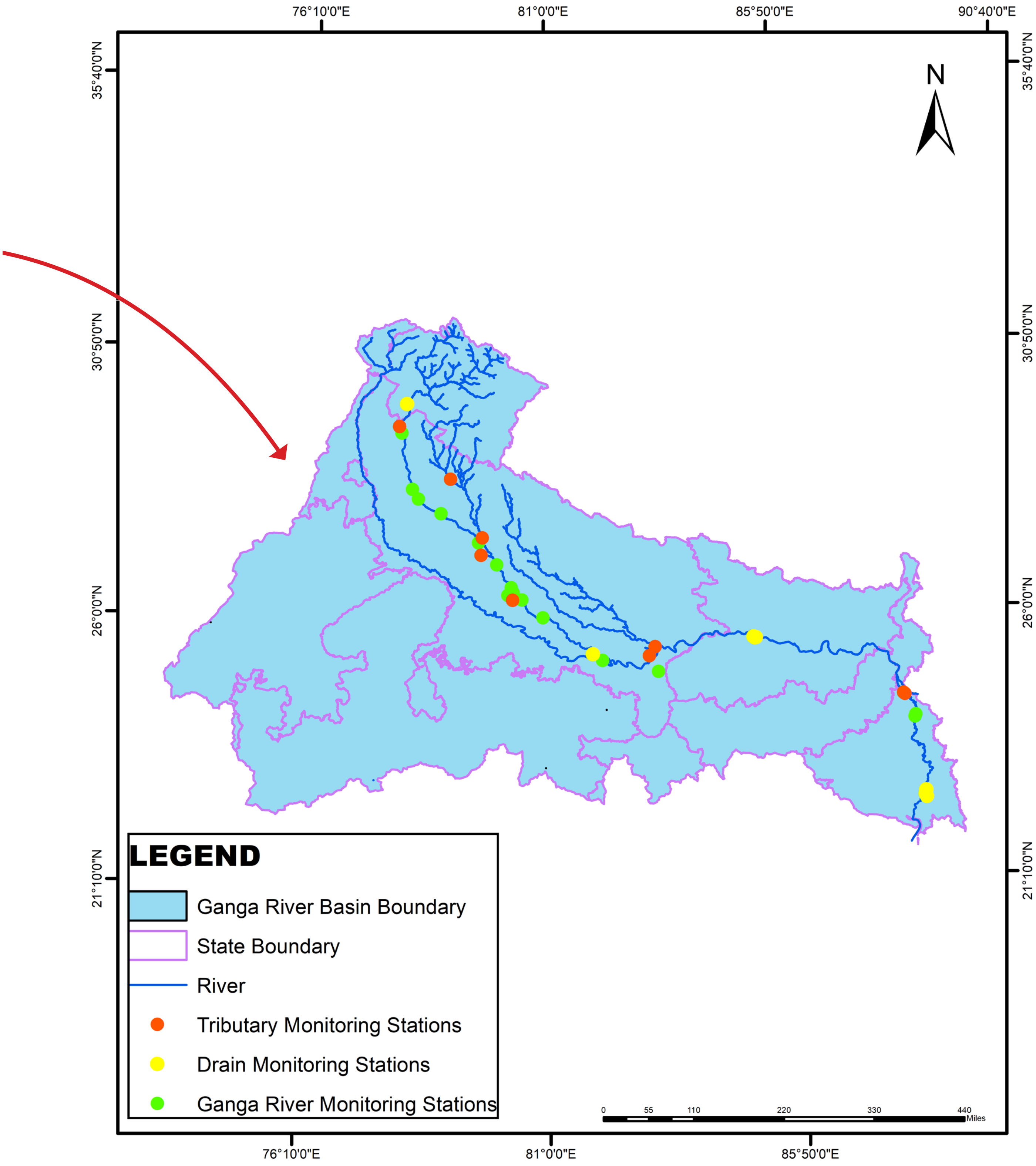


Figure 1: Location of CPCB's river/drain water monitoring stations in Ganga basin

Table 1: List of 36 Real Time Water Quality Monitoring Stations under study

TYPE – 1 RIVER GANGA STATIONS (18)					
Station No.	Station code	Location name	Installed at	Latitude N	Longitude E
Station - 1	UP-02	Madhya Ganga Barrage, Bijnore	River Ganga	29.37379	78.04072
Station - 2	UP-06	Anupshahar Ghat, Anupshahar	River Ganga	28.36452	78.27184
Station - 3	UP-08	Narora Barrage, Narora	River Ganga	28.19036	78.39535
Station - 4	UP-09	Kachla Ghat Bridge, Badaun	River Ganga	27.93106	78.85529
Station - 5	UP-14	Ghatiyaghat Bridge, Farrukhabad	River Ganga	27.40900	79.61833
Station - 6	UP-16	Manimau Bridge (Mehendi Ghat), Kannauj	River Ganga	27.01279	79.98865
Station - 7	UP-18	Pariyal Bridge, Bithoor, Kanpur	River Ganga	26.60028	80.27889
Station - 8	UP-19	Ganga (Luv Kush) Barrage, Kanpur	River Ganga	26.50825	80.31645
Station - 9	UP-26	Shuklaganj Bridge, Kanpur	River Ganga	26.46188	80.20968
Station - 10	UP-24	Dhondhi Ghat (Maharajapur)	River Ganga	26.37634	80.49249
Station - 11	UP-32	Bridge at Ansi, Fatehpur	River Ganga	26.05505	80.90953
Station - 12	UP-40	Pontoon Bridge Sirsa, Allahabad	River Ganga	25.27100	82.09300
Station - 13	UP-56	Ghazipur (D/s)	River Ganga	25.05603	83.19933
Station - 14	WB-10	Bridge at Behrampore, Behrampore (U/s)	River Ganga	24.10038	88.24428
Station - 15	WB-11	Bridge at Behrampore, Behrampore (D/s)	River Ganga	24.06172	88.22758
Station - 16	WB-21	Ghat at Srirampore, Srirampore (D/s)	River Ganga	22.72577	88.35612
Station - 17	WB-23	Intake Pumping Station at Belgharia, Belgharia	River Ganga	22.67095	88.35973
Station - 18	WB-27	Millenium Park, Howrah Bridge, Howrah	River Ganga	Not Available	Not Available

TYPE – 2 TRIBUTARY STATIONS (9)					
Station - 19	UP-03	Sukratal Ghat, Bijnore	River Banganga	29.49350	77.99003
Station - 20	UP-10	Shahbad Bridge, Moradabad (D/s)	River Ramganga	28.55350	79.04748
Station - 21	UP-13	Khudaganj Bridge, Farrukhabad	River Kali	27.18447	79.67083
Station - 22	UP-17	Allahganj Bridge, Kannauj-Farrukhabad	River Ramganga	27.49797	79.69614
Station - 23	UP-29	Bhimgave Bridge, Hamirpur Road, Kanpur	River Pandu	26.37063	80.30699
Station - 24	UP-54	Bathing Ghat 1, Varanasi	River Varuna	25.34212	83.02291
Station - 25	UP-55	Bridge at River Gomti, Varanasi	River Gomti	25.50696	83.14097
Station - 26	WB-05	1_Bridge on NH-34, Farrakha (U/s)	River Falgu	24.50535	88.03008
Station - 27	WB-06	2_Bridge on NH-35, Farrakha (U/s)	River Maya	24.48237	88.05554
TYPE – 3 DRAIN STATIONS (9)					
Point - 1	UK-08	Jagjeetpur STP drain, Haridwar	Drain	29.90054	78.14071
Point - 2	UP-46	Mawaiyanala, Allahabad	Drain	25.38990	81.90133
Point - 3	BH-07	Kurzi Nalla, Patna	Drain	25.64092	85.10539
Point - 4	BH-09	Rajapul Nalla, Rajapur old pump house	Drain	25.62358	85.12460
Point - 5	BH-10	Mandiri Nalla, Patna 3a	Drain	25.62226	85.13366
Point - 6	BH-11	Anta Ghat Nalla, Patna 3a	Drain	25.62230	85.15043
Point - 7	WB-22	Nalla opposite Ghat (D/s), Srirampur	Drain	22.72629	88.36413
Point - 8	WB-24	Ballykhal Nalla, Ballykhal bridge	Drain	22.65503	88.34764
Point - 9	WB-26	Chitpur Nalla, Circular Canal, Chitpur	Drain	22.60748	88.36977

2.1 Data Acquisition Technique and Sensor Information

All the 17 water quality parameters are measured by innovative sensors, preferred optical, that is reagent-free and operate almost without maintenance. Monitoring stations were installed by Austria-based company s::can (refer- https://www.s-can.at/) in partnership with local Indian companies such as Aaxis Nano, Tritec, and Techspan, etc. The monitoring stations comprised of –

- Up to 4 to 6 sensors each to measure 17 parameters.
- Station terminal with SQL postgres database, interfaces for - almost any number of analog and digital sensor inputs, SDI12, Modbus, USB, TCP/IP-Ethernet, 4-20 mA, and other interfaces.
- The con::cube – a compact versatile terminal for acquiring data and controlling station.
- The moni::tool – a powerful terminal for compact station control and data management, data validation, and event detection software.
- Battery charging system (battery, solar charger, solar panel).
- Auto brush cleaning for energy-optimized cleaning of sensors.
- VPN access for remote control 8 Cameras and alarm sirens, security cages, and other protection against vandalism.

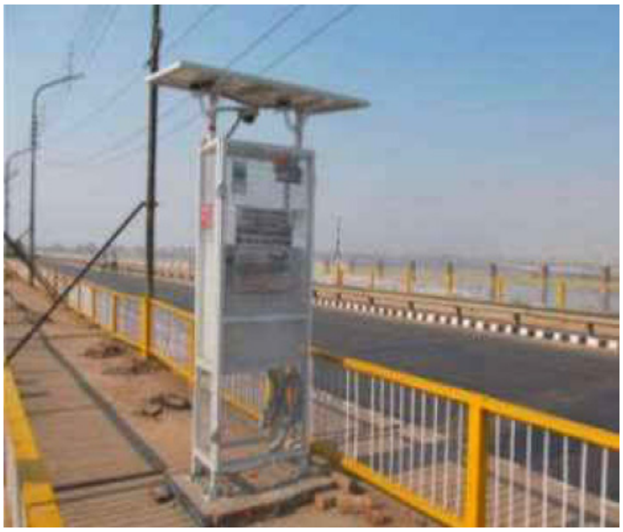


Figure 2: Real-time water quality monitoring station on river Ganga at Allahabad

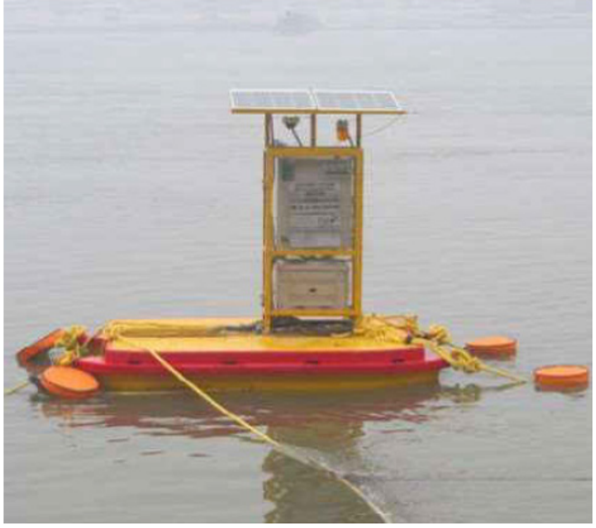


Figure 3: Floating type real-time water quality monitoring station

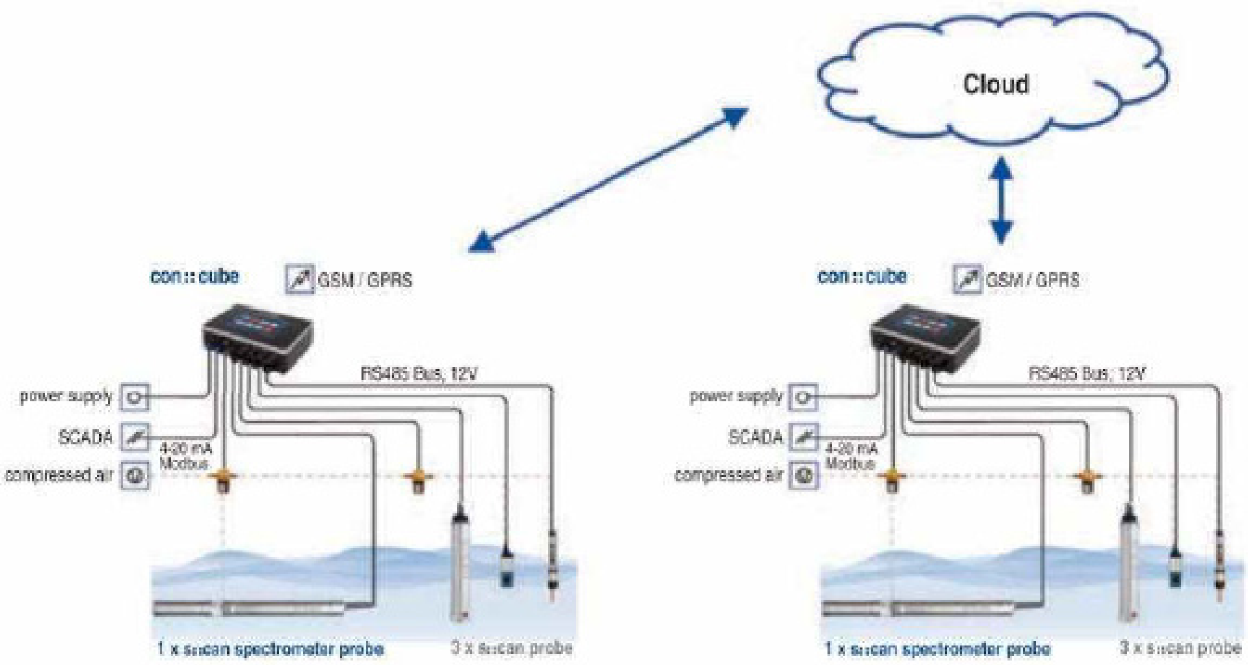


Figure 4: Overview of monitoring network

All real-time data is automatically transferred via a General Packet Radio Service network and a secure shell protocol to a receiving cloud server, and thereafter to the CPCB central office in New Delhi as shown in Figure 4.

Probes used for measuring the 17 water quality parameters data, their measurement principles, and their brief specification are given in Table 2.

Table 2: Summary of the Probes/Sensors

Specifications	Oxi::lyser	Spectro::lyser	Condu::lyser	Amno::lyser
Measuring Principle	Fluorescence	UV Spectrometry 190 - 390 nm	4-Electrode, Direct-Contact	Ion Selective Electrodes (ISE)
Monitoring Parameters	Dissolved Oxygen & Temperature	NO_2-N, TSS, Turbidity, NO_3-N, COD, BOD, TOC, UV254, BTX	Conductivity, Temperature & Salinity	NH_4-N [mg/l] NO_3-N [mg/l] pH, Temperature
Measuring Range	25 mg/l O_2		Conductivity 0-500,000 µS/cm	1-1000 mg/l NH_4-N and Cl
Accuracy	1% of reading	(l) NO_3-N: +/- 2%+1/OPL [mg/l]* , COD-KHP: +/-2% +10/OPL[mg/l]* (*OPL-optical pathlength in mm)	1% of reading	NH_4-N: +/-3% of Measuring Range or +/-0.5mg/l* (*whichever is greater)
Operating Temperature	0 to 60 °C	0 to 45 °C	20-60 °C (immersed), 20-110 °C (in flow cell)	0 to 60 °C
Reference Standard	Saturated Sodium Sulphite Solution	Distilled Water		

Principles on which Sensors are based:

a) Spectrophotometry – The spectrometer probe uses the principle of UV Spectrometry (Lindon et al. 2016). It estimates the light absorbed with the help of a chemical substance by measuring the intensity of light as a monochrome beam of light that passes through a sample solution. The basic principle in this method based on the property of absorbing or transmitting light over a certain range of wavelengths for each chemical compound. It can measure optical spectra from 200 to 750 nm directly in liquid media. Spectrometer produces the required range of wavelengths of light. When the required range of wavelength of light passes through the sample solution, the photometer detects the number of photons that have been absorbed, and then transmits a signal to a digital display. Studies have found that organic matter has maximum absorption value in the ultraviolet region of 254 nanometers.

b) Fluorescence quenching – The Dissolved Oxygen (DO) sensor, based on the fluorescence quenching principle, is composed of excitation light sources, a substrate film attached to fluorescence-sensitive substances, and an optoelectronic detection element (Wei et al. 2019). After a collision with DO molecules, the fluorescent substance absorbs visible or ultraviolet light of a specific wavelength, its electrons gain energy and become excited and release energy to return to the ground state by emitting fluorescence. Since the collisions between oxygen molecules and excited fluorescent substances interfere with the excitation process of fluorescent substances, the content of oxygen molecules in the water samples can be determined according to the fluorescence intensity or the fluorescence lifetime generated at the sensitive interface.

c) Ion-selective electrode (ISE) – It is an example of an electrochemical sensor utilizing the principle of potentiometry (Hanrahan et al. 2005). It measures the cell potential difference across two electrodes, i.e., ISE against a standard reference electrode at near-zero current. The boundary potential at the ISE–solution interface is governed by the laws of electrochemical thermodynamics and is compliant with the Nernst equation.

d) Conductivity measurement with conductive 4-electrode sensors – In this instrument, four electrode sensors measure the potential difference in a medium (Hyldgard et al. 2005). Each sensor has two electrodes that have no current and are, therefore, not affected by the polarization effect. A connected transmitter uses the measured potential difference and current to calculate the conductivity value. The polarisation effect defines a mutual repulsion of the ions due to a high ion concentration in the medium, which leads to reduced current and, hence, possible influence on the measuring accuracy of the probes. The 4-electrode method of measuring increases the accuracy by avoiding the polarization effect.

All real-time data is automatically transferred via a General Packet Radio Service network and a secure shell protocol to a receiving cloud server, and thereafter to the CPCB central office in New Delhi

2.2 Water Quality Parameters of Study

CPCB's RTWQMS was established to monitor seventeen water quality parameters. The seventeen parameters of interest are shown in Table 3. They are categorized by CPCB into 4 categories (A, B, C, and D) on the basis of the importance of the parameter to the client.

Table 3: 17 Water Quality Parameters categorised on the basis of their importance

Category A	Category B	Category C	Category D
Biochemical Oxygen Demand (BOD)	Ammonia	Colour	BTX
Dissolved Oxygen (DO)	Chemical Oxygen Demand (COD)	Fluoride	Total Organic Carbon (TOC)
Chemical Oxygen Demand (COD)	Total Suspended Solids (TSS)	Nitrogen nutrients (NH_4-N, NO_3)	Water level
pH	Chloride	Potassium	
Temperature	Turbidity		

3.0 Methodology

3.1 Data Processing

CPCB collects the above flow data on an hourly basis. Depending on the use of the water quality parameters, hourly or daily average data may be calculated. The data for the long timeline was received from CPCB in various excel files containing multiple sheets and workbooks. Firstly, the raw data was converted from the excel workbooks and worksheets into a uniform time series format for each parameter. The time series conversion of physico-chemical parameters was carried out for all stations individually. Datasets were arranged and sorted as per the need for the analysis of the data.

3.2 Removal of Outliers

Outliers are observations or data points that lie at abnormally distant from other values in a random sample from a population (Hodge and Austin 2004). These outliers can be due to data entry errors, instrument errors, or measurement errors. They can be problematic for many statistical analyses, resulting in increased error variance, insignificant findings or distorted results, reduced accuracy of statistical tests, contradict statistical assumptions, or give biased estimates (Osborne and Overbay 2004). There is no strict statistical rule for identifying and eliminating outliers. Removing outliers mainly depends on parameters knowledge. Standard Deviation and Box Plots are the most common methods for outlier detection (Seo 2006). In the present study, three methods were considered to omit outliers from the data sample, viz. (a) Average ± 2x standard deviation, (b) Box plots with a coefficient of 1.5, and (c) Box plots with a coefficient of 3. Using these methods a few certain outliers were removed manually. Abrupt and significantly higher or lower values lying above the entire time series were deleted.

3.3 Correlation Analysis

In this study large datasets have been taken for analysis comprising more than 55 months' of hourly data for various parameters with some data missing on specific days. To ensure parity in analysis, all data for the missing days or hours were deleted from the datasets. The analyses are based primarily on comparing the dataset through correlation analysis for any two pairs of variables (water quality parameters) that are likely to be related to each other. Correlation analysis provides an overview of the correlation between various water quality parameters (Restrepo and González 2007). Karl Pearson's test is a parametric test, whereas the Spearman correlation coefficient is a non-parametric test. Since it is assumed that the data available is normally distributed, the Karl Pearson method is used in this study to determine correlations among various water quality parameters. Pearson's Correlation Analysis is a bivariate analysis that measures linear correlation or association between the two variables in a population, and the direction of the relationship. The correlation coefficient denoted as r is given by following formula (Lee Rodgers and Nicewander 1988):-

$$r = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma(x_i - \overline{x})^2(y_i - \overline{y})^2}}$$

where, r = correlation coefficient,
x_i = magnitudes of the x-variables in a sample,
$\overline{x}$ = mean of the magnitudes of the x-variables
y_i = magnitudes of the y-variables in a sample,
and $\overline{y}$ = mean of the magnitudes of the y-variables

Correlation can be positive (r>0), negative (r<0), or no correlation (r=0) based on the correlation coefficient ranging between +1 (perfect positive correlation) and –1 (perfect negative correlation). The lower the numerical value of r, the poorer is the correlation between the variables, whether r be positive or negative.

The analyses are based primarily on comparing the dataset through correlation analysis for any two pairs of variables (water quality parameters) that are likely to be related to each other

4.0 Evaluating the Reliability of BOD, COD and TOC data from Typical Relations and Regression Analyses

To evaluate the reliability of the data collected by the sensors installed at CPCB's stations, the dataset was assessed against known principles, phenomena, and expected relationships between parameters. Since sensor-based measurements of many water quality parameters are relatively new for many such measurements, many of the parameters measured have been evaluated by Chaudhury (2021). In this report, however, the focus is only on Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Total Organic Carbon (TOC), since their measurement by optical sensors is novel and not an established technique in water quality assessment. BOD by definition measurable only after biological processing, obviously defies merely physical or chemical measurement. For COD and TOC, too, there are no definite scientific principles by which they can be measured optically in non-homogeneous flows having variable compositions of different types of ionic components and molecular structures. Hence their measurement by optical sensors raises fundamental doubts about the resulting data.

To evaluate the reliability of BOD and COD sensor-based data, COD vs BOD relationship graphs for each station have been plotted to examine if the data indicate any meaningful relationship between the two variables. Similarly, the validity of sensor-based TOC data has been evaluated by plotting COD vs. TOC graphs and BOD vs. TOC for each station. In general, a fair correlation is known to exist between COD and BOD for river water data. However, a definite regression relationship (linear or non-linear) between any two of these parameters (BOD and COD, COD and TOC, or BOD and TOC) may not exist if the water quality varies due to significant ingress of different wastewaters/ pollutants at different times, as may happen due to discharge of industrial effluents of variable quality or amounts into drains or small rivers, and also if considerable municipal sewage flows in at variable treated and untreated states in such channels. But this is unlikely to be significant for large rivers due to the homogenising effect of the much larger river flow volumes.

A definite regression relationship (linear or non-linear) between any two of these parameters (BOD and COD, COD and TOC, or BOD and TOC) may not exist if the water quality varies due to significant ingress of different wastewaters/ pollutants at different times

4.1 BOD and COD

In several studies, BOD to COD ratio has been found to serve a reliable and useful critical indicator for pollution measurement in rivers and for biodegradability of the organic matter in the flow. BOD to COD ratio can also indicate the toxicity of rivers or water bodies and can be used as an important attribute to characterize the river. Generally, COD is 1 to 3 times the BOD in biodegradable streams. However, hourly COD data in the present datasets have often been found to be much higher, even 100 times or 1000 times higher than BOD. Such data cannot be considered as outliers, since there are too many such data of very high COD relative to BOD. This is a clear indication of the unreliability of the dataset for BOD and COD considered together.

To further test the hourly data for BOD and COD, the entire dataset of COD vs. BOD were plotted for each measuring station as shown in Figure 5. A careful look at the figure shows that the data points are generally widely scattered and they do not suggest any definitive relation. While this is possible for drains receiving significant and variable amounts of pollutants from domestic sewage and industrial effluents, it is unlikely for large rivers for reasons stated earlier. While the scatter plots do not suggest very definitive relation between BOD and COD, linear regression analysis was applied to the data to check for relatedness between BOD and COD. The regression lines thus obtained are shown in the respective plots for each measuring station.

Looking at the regression plots of Figure 5, the following observations can be clearly made for the river stations:

(i) The data are positively correlated ($r>0$), but with slopes of the regression lines varying from about 1.4 to nearly 5 between different river stations with corresponding large variations in the intercepts. These large variations of flope between river stations are inexplicable. They also suggest vastly different average ratios of COD:BOD between stations which is also unrealistic.

(ii) BOD and COD are poorly correlated for most river stations, with absolute values of correlation coefficient $|r|<0.5$ in many cases.

(iii) The scatter around the regression lines is very large in many plots.

(iv) The errors in the data (root mean square error of COD values) are high in most cases even when the correlation is fair, which are beyond the acceptable range of measurement errors by standard methods.

All four observations indicate that the above measurements are much less reliable than standard methods and are evidently erroneous to a high degree. It may be noted that since Pearson's correlation coefficient is distribution-dependent, the non-parametric Spearman's Rank Correlation Coefficient (Montogomery and Runger, 2016) may be used instead to indicate any possible monotonic association between BOD and COD. The values of Spearman's rank correlation coefficient, rs, (computed using MATLAB software) were found to be comparable to Pearson's coefficient, confirming the poor correlation between

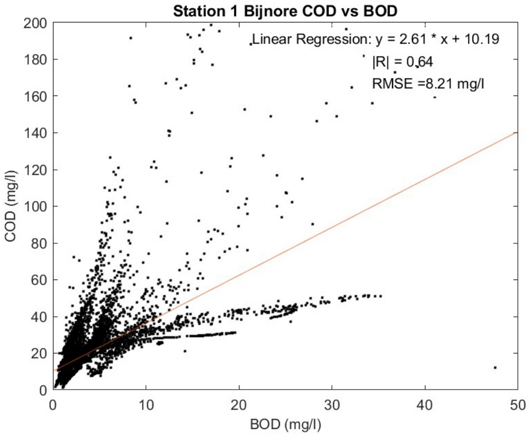
BOD and COD. The values of Pearson's "r" and Spearman's "rs" are presented in Table 4.

The four points of concern noted above are even more pronounced for measurements at drain stations, but as mentioned earlier, at least part of the difference may be attributable to possible inflows of variable water quality into the drains. Hence, without knowing the quantum and quality of such inflows with respect to drain flows, definitive comments cannot be made about measurement errors for drain water quality measurements.

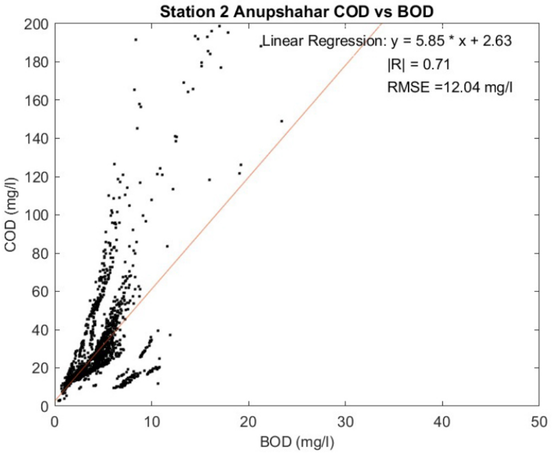
Table 4: Pearson's and Spearman's correlation coefficients between BOD and COD of sensor-measured water quality data

Station ID	COD vs BOD	
	Pearson's "r"	Spearman's "r_s"
Station 1	0.64	0.67
Station 2	0.71	0.87
Station 3	0.30	0.56
Station 4	0.67	0.65
Station 5	0.46	0.48
Station 6	0.57	0.48
Station 7	0.45	0.36
Station 8	0.35	0.57
Station 9	0.51	0.42
Station 10	0.88	0.67
Station 11	0.84	0.59
Station 12	0.49	0.57
Station 13	0.29	0.46
Station 14	0.56	0.76

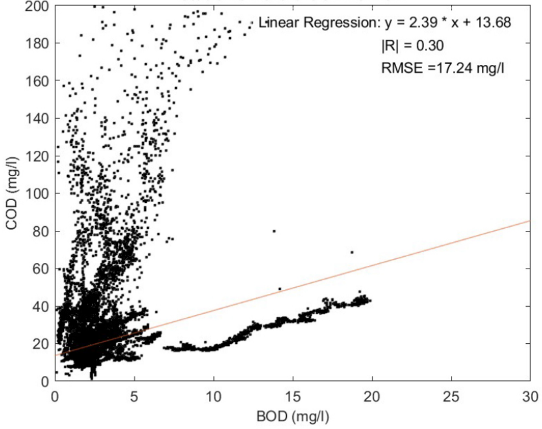
Station ID	COD vs BOD	
	Pearson’s “r”	Spearman’s “r_s”
Station 15	0.80	0.83
Station 16	0.48	0.52
Station 17	0.79	0.73
Station 18	0.47	0.68
Station 19	0.80	0.63
Station 20	0.64	0.55
Station 21	0.38	0.32
Station 22	0.59	0.50
Station 23	0.70	0.65
Station 24	0.69	0.75
Station 25	0.63	0.65
Station 26	0.75	0.41
Station 27	0.41	0.25
Point 1	0.53	0.58
Point 2	0.04	0.13
Point 3	0.33	0.29
Point 4	0.20	0.43
Point 5	0.52	0.63
Point 6	0.31	0.54
Point 7	0.56	0.59
Point 8	0.72	0.81
Point 9	0.51	0.57



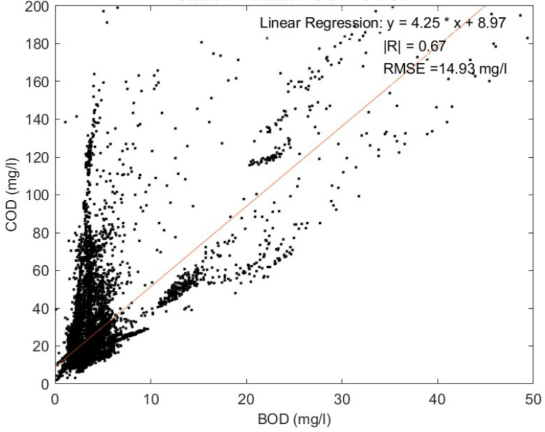
Station 1 Bijnore COD vs BOD



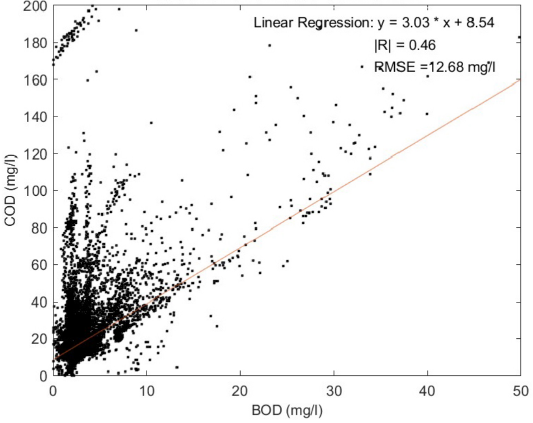
Station 2 Anupshahar COD vs BOD



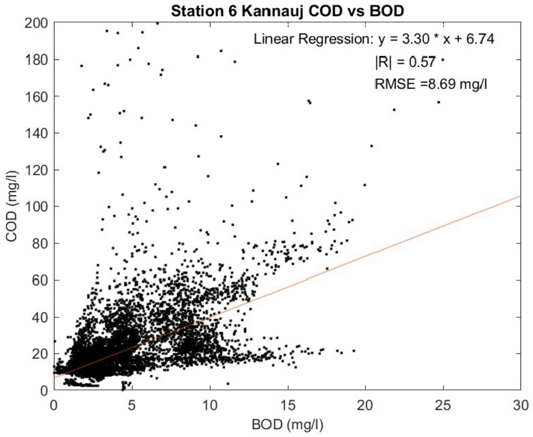
Station 3 Narora COD vs BOD



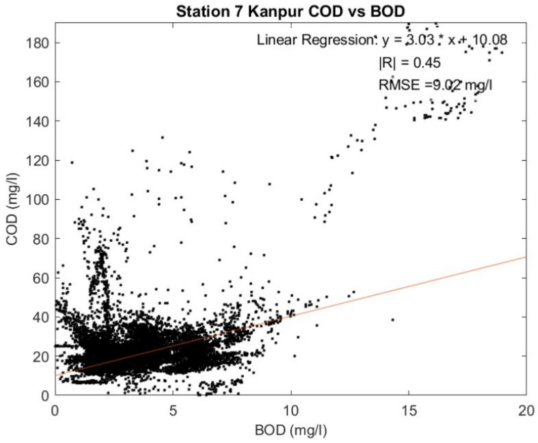
Station 4 Badaun COD vs BOD

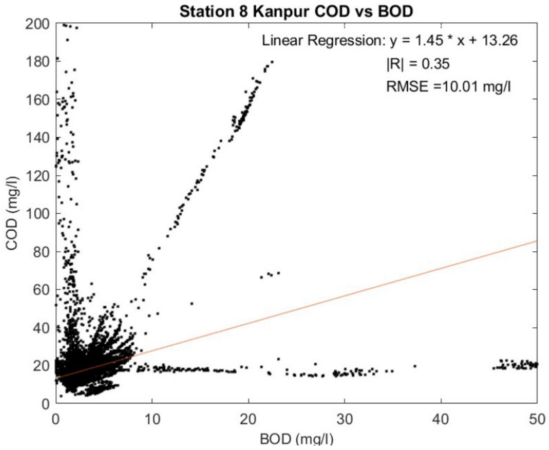


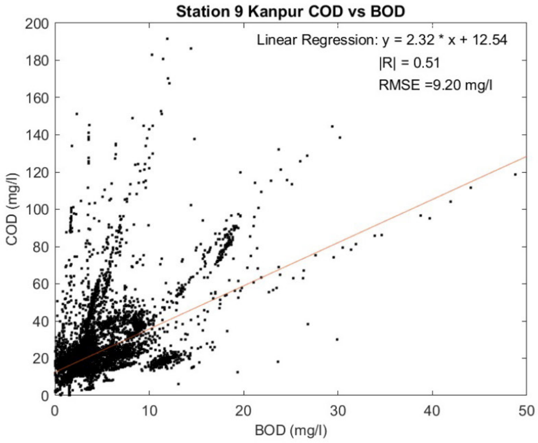
Station 5 Farrukhabad COD vs BOD

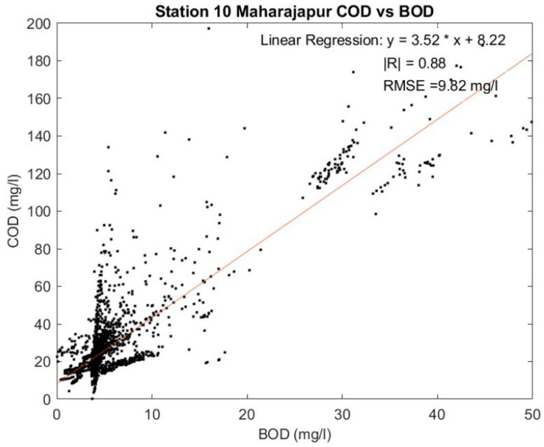


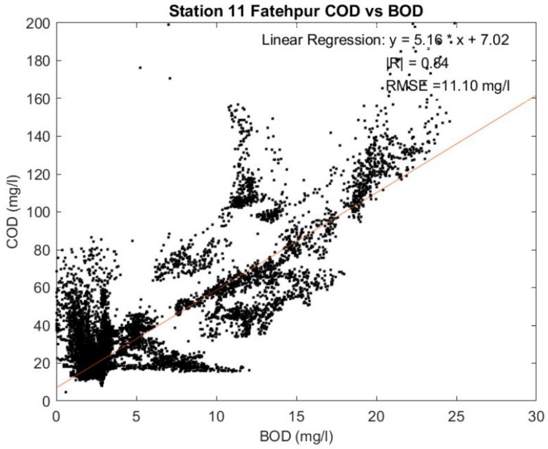
Station 6 Kannauj COD vs BOD

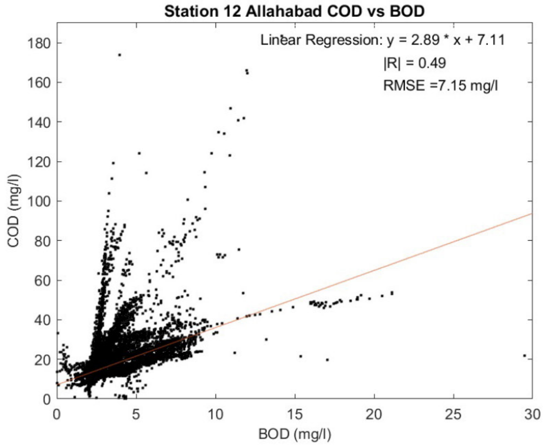

Station 7 Kanpur COD vs BOD

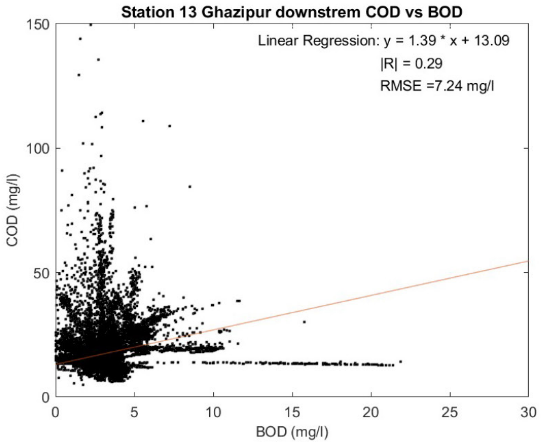

Station 8 Kanpur COD vs BOD

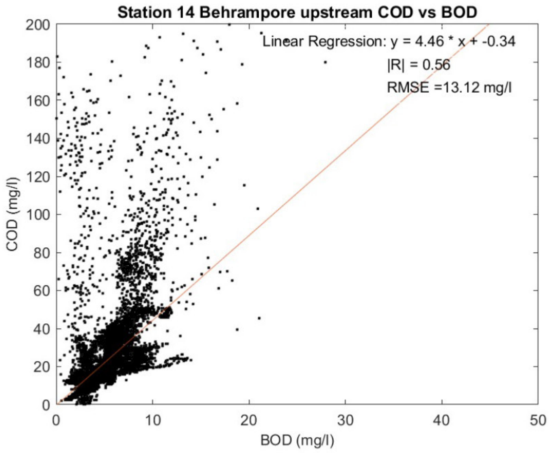

Station 9 Kanpur COD vs BOD

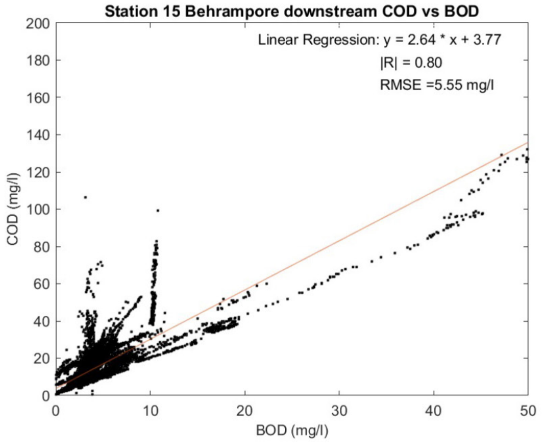

Station 10 Maharajapur COD vs BOD

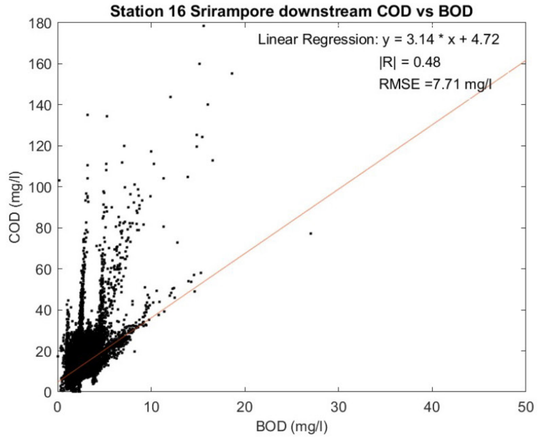

Station 11 Fatehpur COD vs BOD

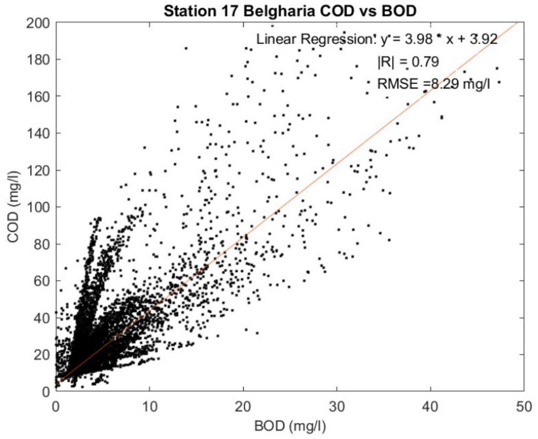

Station 12 Allahabad COD vs BOD

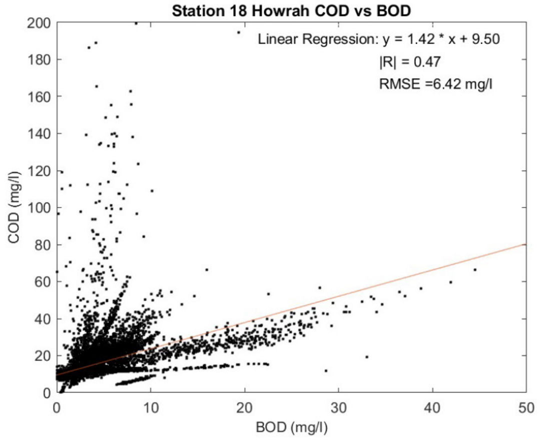

Station 13 Ghazipur downstrem COD vs BOD

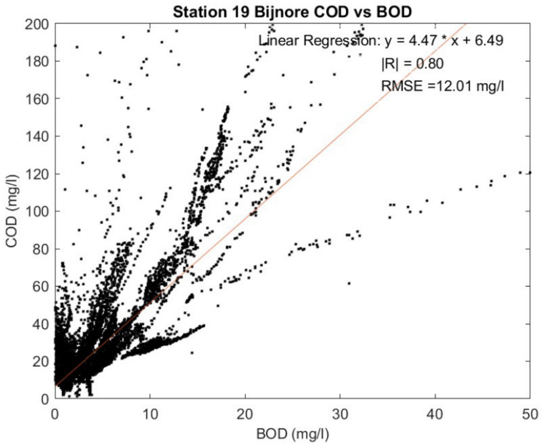

Station 14 Behrampore upstream COD vs BOD

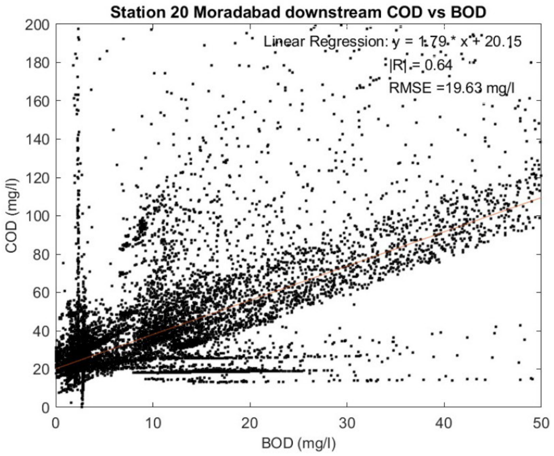

Station 15 Behrampore downstream COD vs BOD

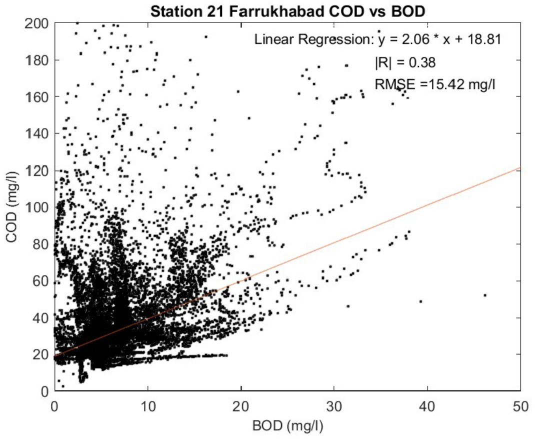

Station 16 Srirampore downstream COD vs BOD

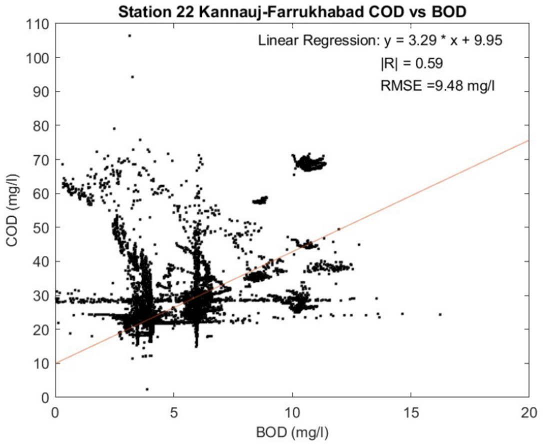

Station 17 Belgharia COD vs BOD

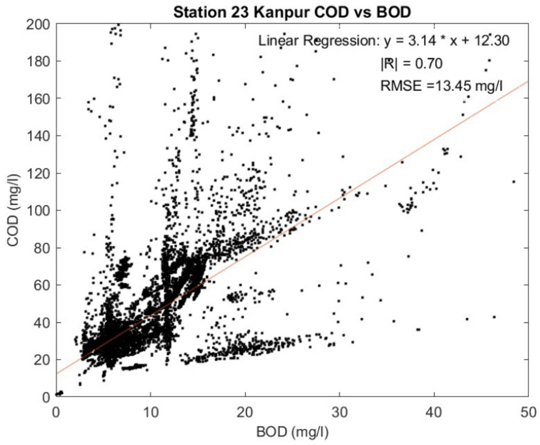

Station 18 Howrah COD vs BOD

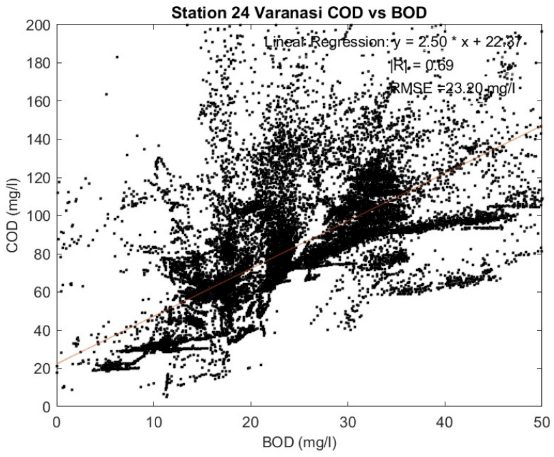

Station 19 Bijnore COD vs BOD
Linear Regression: y = 4.47 * x + 6.49
|R| = 0.80
RMSE =12.01 mg/l
COD (mg/l)
BOD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50

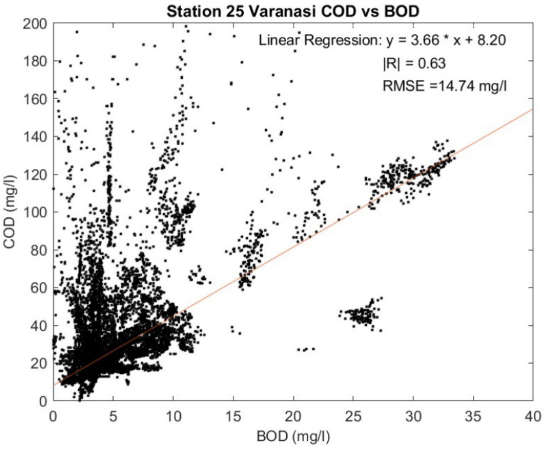

Station 20 Moradabad downstream COD vs BOD
Linear Regression: y = 1.79 * x + 20.15
|R| = 0.64
RMSE =19.63 mg/l
COD (mg/l)
BOD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50


Station 21 Farrukhabad COD vs BOD
Linear Regression: y = 2.06 * x + 18.81
|R| = 0.38
RMSE =15.42 mg/l
COD (mg/l)
BOD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50

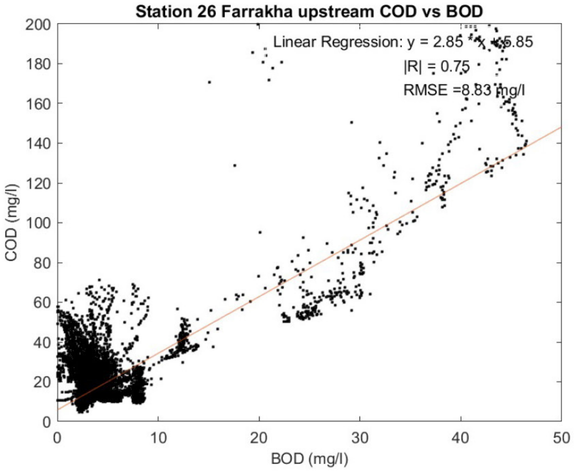

Station 22 Kannauj-Farrukhabad COD vs BOD
Linear Regression: y = 3.29 * x + 9.95
|R| = 0.59
RMSE =9.48 mg/l
COD (mg/l)
BOD (mg/l)
0
10
20
30
40
50
60
70
80
90
100
110
0
5
10
15
20


Station 23 Kanpur COD vs BOD
Linear Regression: y = 3.14 * x + 12.30
|R| = 0.70
RMSE =13.45 mg/l
COD (mg/l)
BOD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50

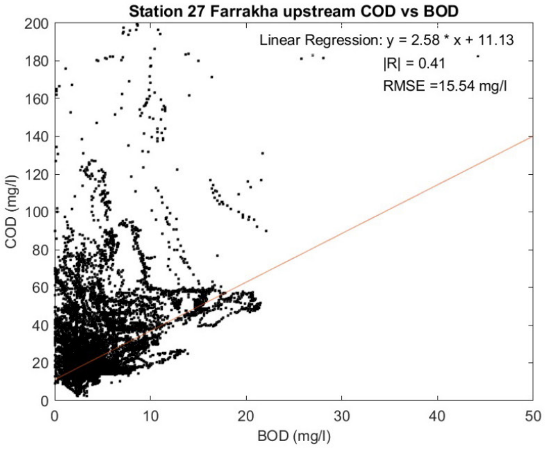

Station 24 Varanasi COD vs BOD
Linear Regression: y = 2.50 * x + 22.37
|R| = 0.69
RMSE =23.20 mg/l
COD (mg/l)
BOD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
10
20
30
40
50



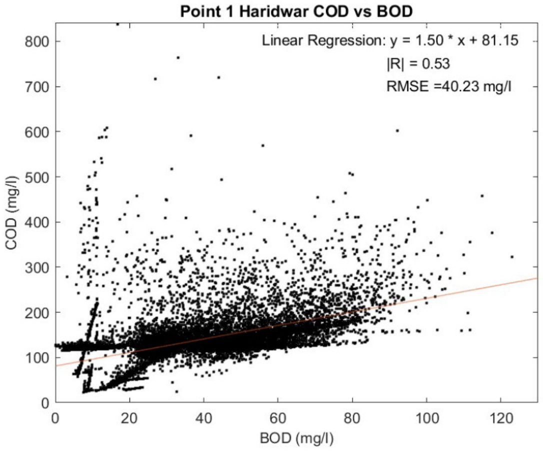
Station 25 Varanasi COD vs BOD



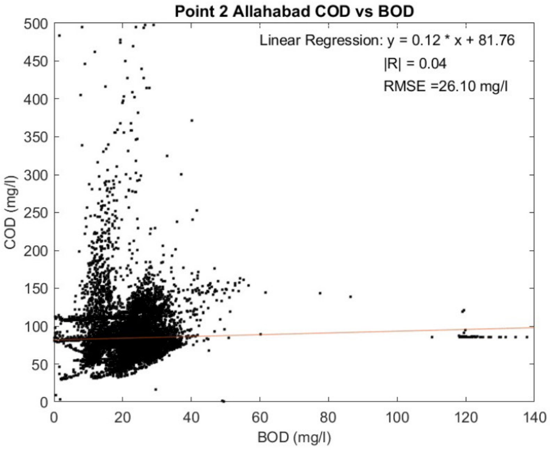
Station 26 Farrakha upstream COD vs BOD



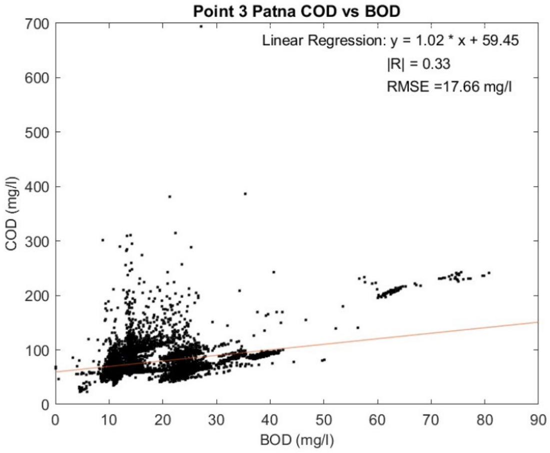
Station 27 Farrakha upstream COD vs BOD



Point 1 Haridwar COD vs BOD



Point 2 Allahabad COD vs BOD



Point 3 Patna COD vs BOD

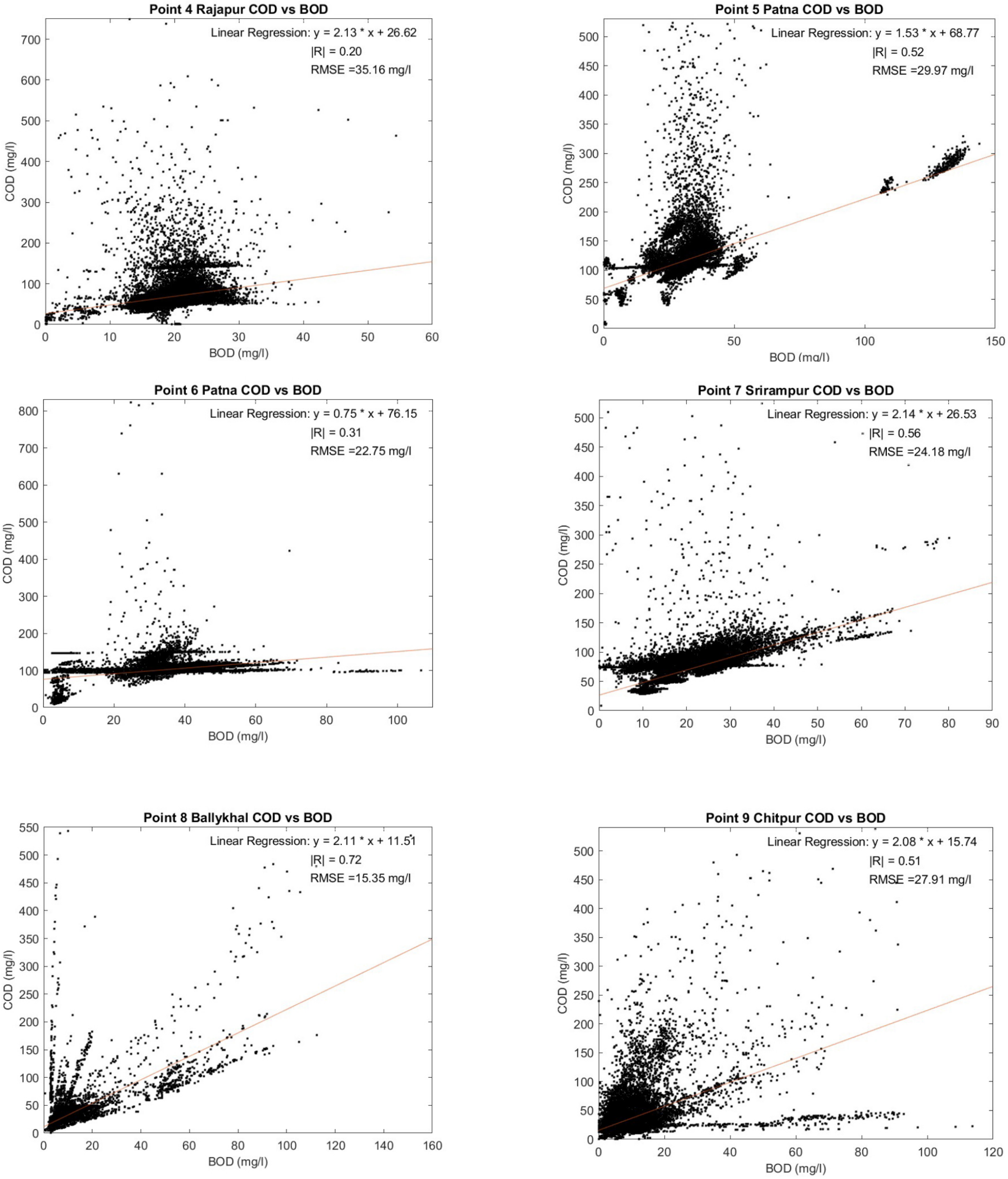


Figure 5: Regression plots of COD vs. BOD

4.2 COD and TOC, BOD and TOC

TOC is the total amount of carbon content of organic compounds whereas COD is the oxygen equivalent that acts as the electron donor during full oxidation of those compounds. The ratio of TOC: COD depends on the oxidation state of carbon. Since organics with similar amounts of carbon may have vastly different molecular structures, therefore TOC does not determine the full impact of complex organics on treatment. Moreover TOC shows only organic carbon. COD shows not only organic carbon, but also metals, which can change their oxidation state. Thus COD:TOC ratio generally ranges from 2:1 to 6:1. But in the data sets in many cases TOC is found to be greater than COD, which is evidently erroneous.

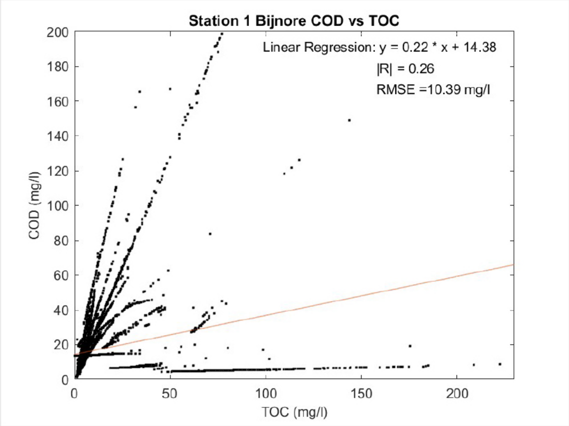
In industrial wastewaters with variable organic loads COD and TOC (hence, also BOD and TOC) may not have a very good relation, but in natural wastewaters they will generally show a high degree of correlation unless subjected to large and variable wastewater inflows. Thus fair to good correlation between COD and TOC may be expected in large rivers but not perfect correlation. Figure 6 shows the linear regression plots of COD vs. TOC and BOD vs. TOC for different measuring stations. It is observed from the figure that COD vs. TOC data seem to lie on one or two (or a few) perfectly straight lines in most cases, which is certainly confounding. It may be surmised that TOC data has been simply derived as a proportion (or one of a few randomly varying proportions) of COD instead of any actual TOC measurement whatsoever.

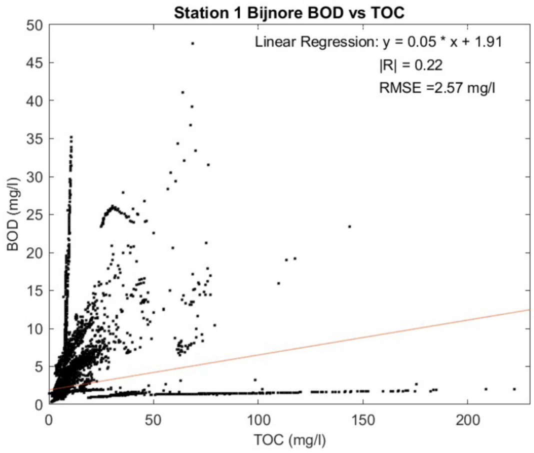
Finally, as in the case of Figure 5 (vide Section 4.1), similar discrepancies are observed indicating the data points to be widely scattered, with poor correlation, and high values of root mean square error. While this is possible for drains receiving significant and variable amounts of pollutants from domestic sewage and industrial effluents, it is unlikely for large rivers for reasons stated earlier. These discrepancies further indicate the erroneous nature of the measurements of BOD, COD and TOC.

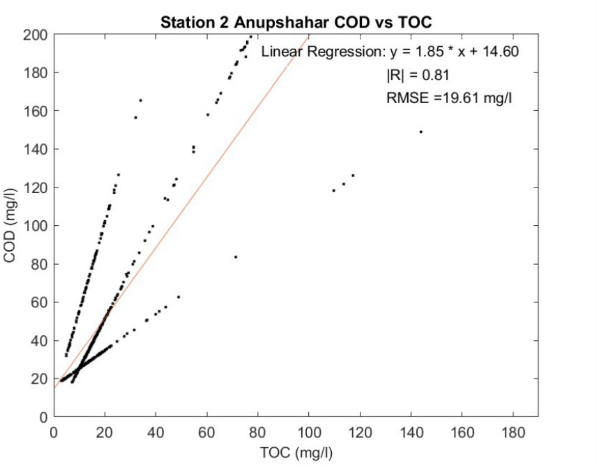
As for BOD & COD correlations, in the case of COD & TOC and BOD & TOC variables too the Pearson's "r" and Spearman's "rs" were found to be comparable, confirming the poor correlations between these pairs of variables.confirming poor correlations.

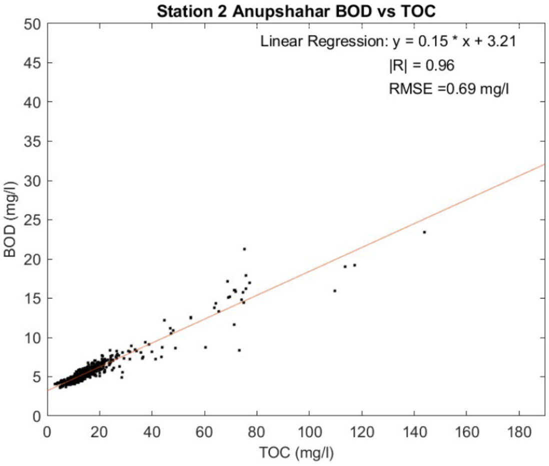
COD vs. TOC data seem

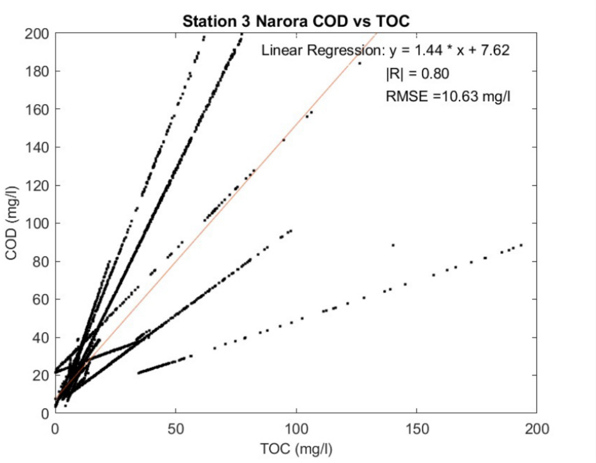
to lie on one or two (or a few) perfectly straight lines in most cases, which is certainly confounding. It may be surmised that TOC data has been simply derived as a proportion (or one of a few randomly varying proportions) of COD instead of any actual TOC measurement

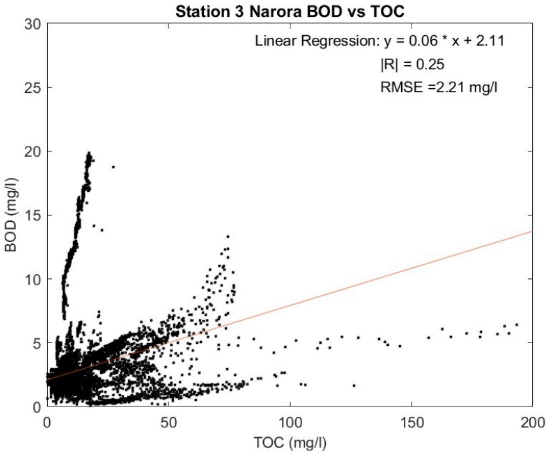

Station 1 Bijnore COD vs TOC

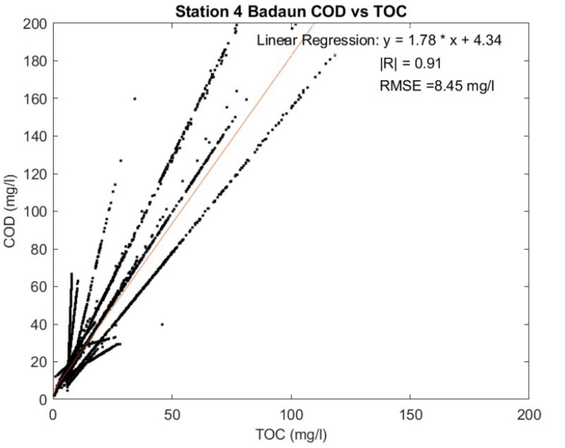

Station 1 Bijnore BOD vs TOC

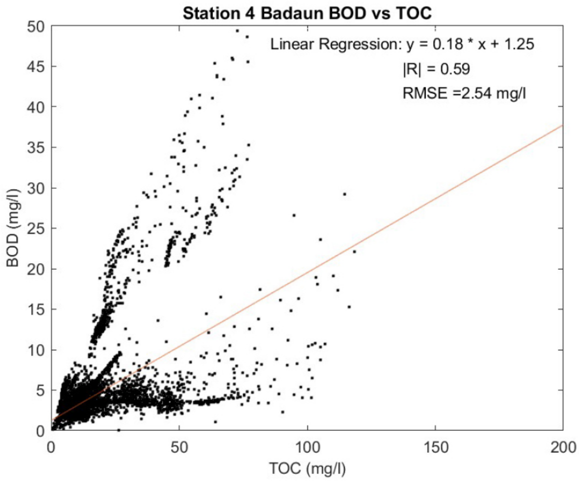

Station 2 Anupshahar COD vs TOC

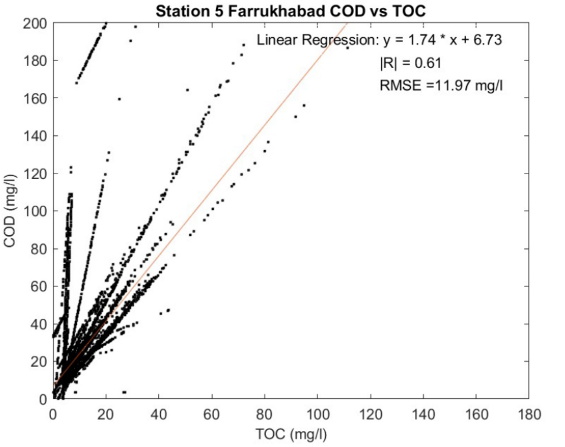

Station 2 Anupshahar BOD vs TOC

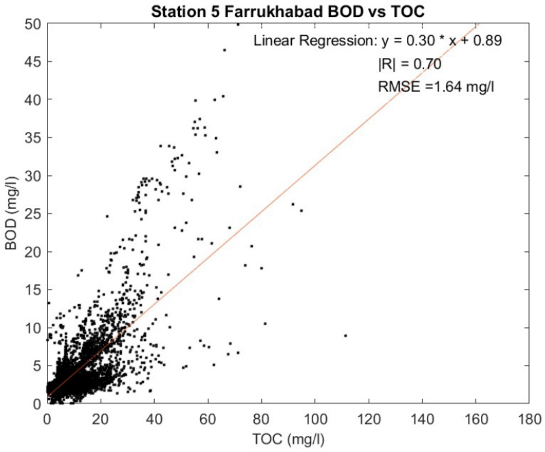

Station 3 Narora COD vs TOC

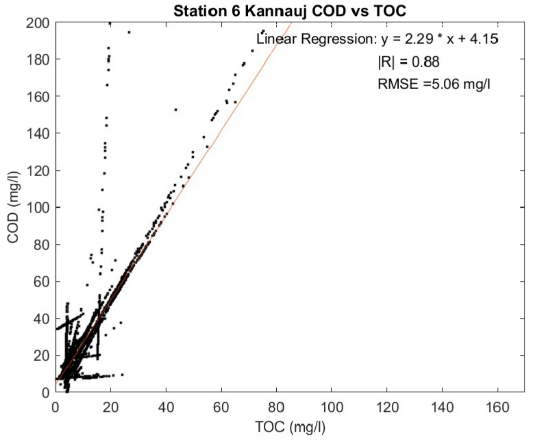

Station 3 Narora BOD vs TOC

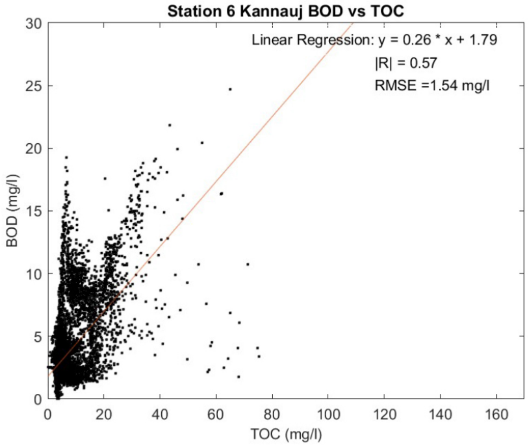

Station 4 Badaun COD vs TOC
Linear Regression: y = 1.78 * x + 4.34
|R| = 0.91
RMSE =8.45 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
50
100
150
200

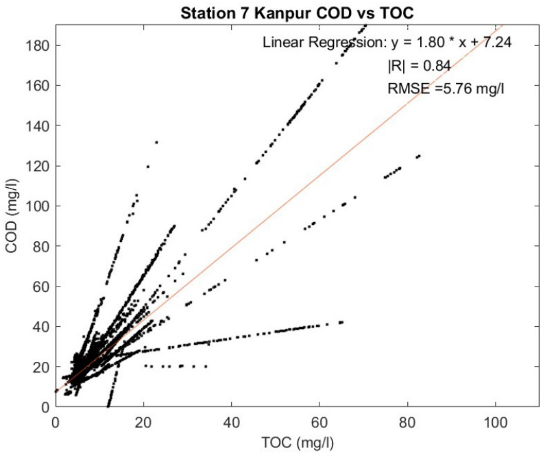

Station 4 Badaun BOD vs TOC
Linear Regression: y = 0.18 * x + 1.25
|R| = 0.59
RMSE =2.54 mg/l
BOD (mg/l)
TOC (mg/l)
0
5
10
15
20
25
30
35
40
45
50
0
50
100
150
200

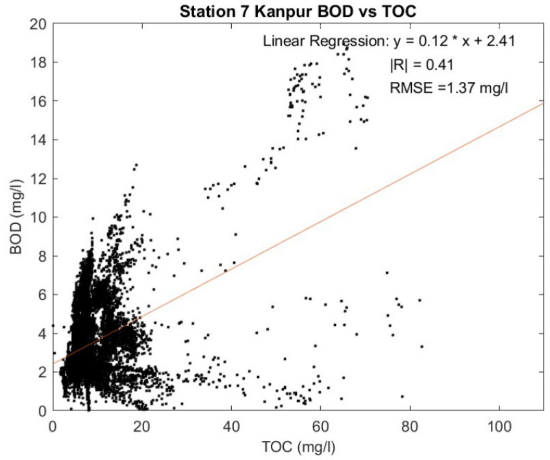

Station 5 Farrukhabad COD vs TOC
Linear Regression: y = 1.74 * x + 6.73
|R| = 0.61
RMSE =11.97 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
20
40
60
80
100
120
140
160
180

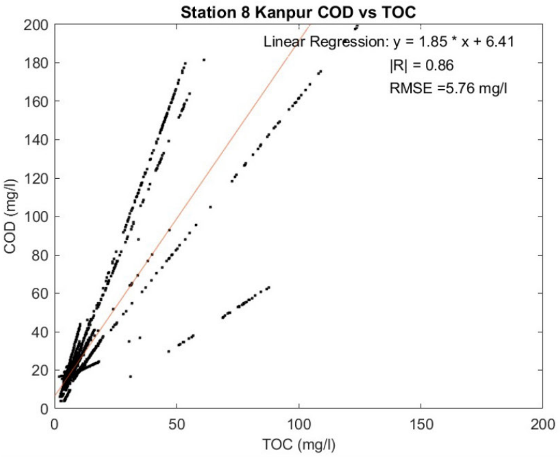

Station 5 Farrukhabad BOD vs TOC
Linear Regression: y = 0.30 * x + 0.89
|R| = 0.70
RMSE =1.64 mg/l
BOD (mg/l)
TOC (mg/l)
0
5
10
15
20
25
30
35
40
45
50
0
20
40
60
80
100
120
140
160
180

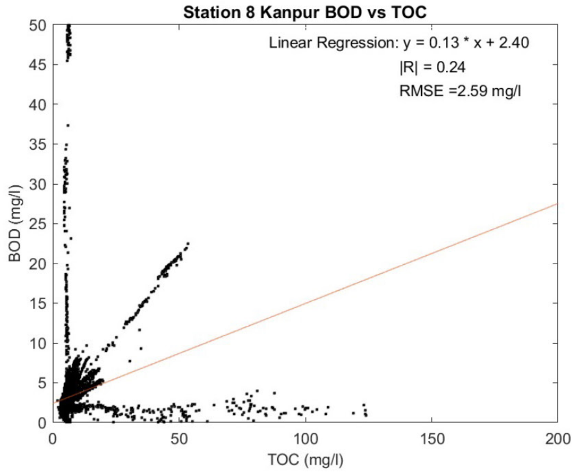

Station 6 Kannauj COD vs TOC
Linear Regression: y = 2.29 * x + 4.15
|R| = 0.88
RMSE =5.06 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
20
40
60
80
100
120
140
160

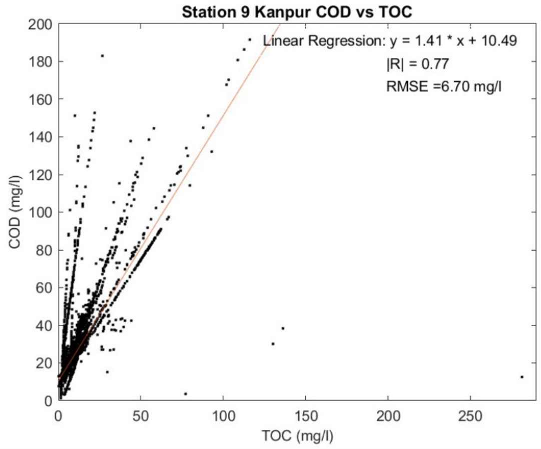

Station 6 Kannauj BOD vs TOC
Linear Regression: y = 0.26 * x + 1.79
|R| = 0.57
RMSE =1.54 mg/l
BOD (mg/l)
TOC (mg/l)
0
5
10
15
20
25
30
0
20
40
60
80
100
120
140
160

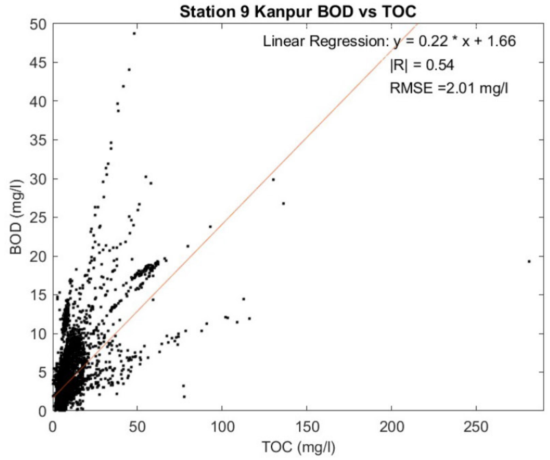

Station 7 Kanpur COD vs TOC
Linear Regression: y = 1.80 * x + 7.24
|R| = 0.84
RMSE =5.76 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
0
20
40
60
80
100

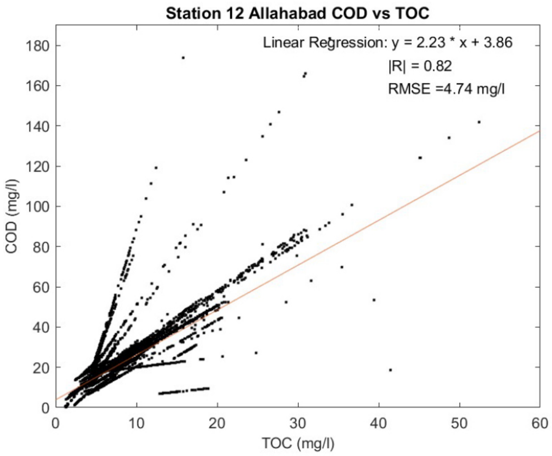

Station 7 Kanpur BOD vs TOC
Linear Regression: y = 0.12 * x + 2.41
|R| = 0.41
RMSE =1.37 mg/l
BOD (mg/l)
TOC (mg/l)
0
2
4
6
8
10
12
14
16
18
20
0
20
40
60
80
100

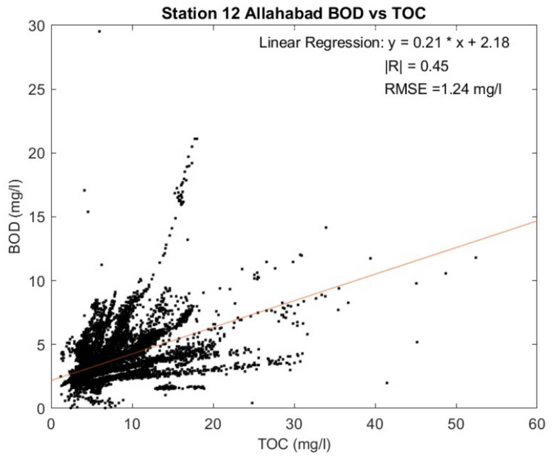

Station 8 Kanpur COD vs TOC
Linear Regression: y = 1.85 * x + 6.41
|R| = 0.86
RMSE =5.76 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
50
100
150
200

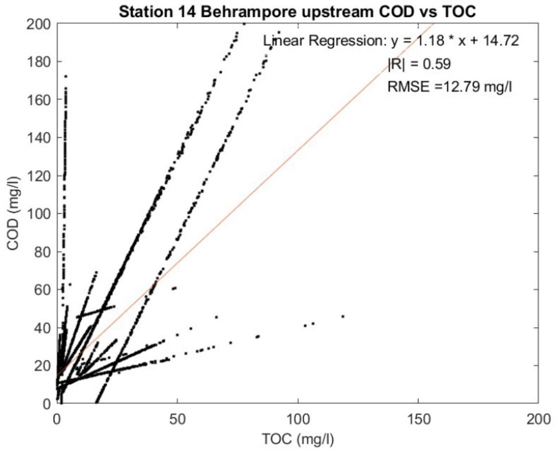

Station 8 Kanpur BOD vs TOC
Linear Regression: y = 0.13 * x + 2.40
|R| = 0.24
RMSE =2.59 mg/l
BOD (mg/l)
TOC (mg/l)
0
5
10
15
20
25
30
35
40
45
50
0
50
100
150
200

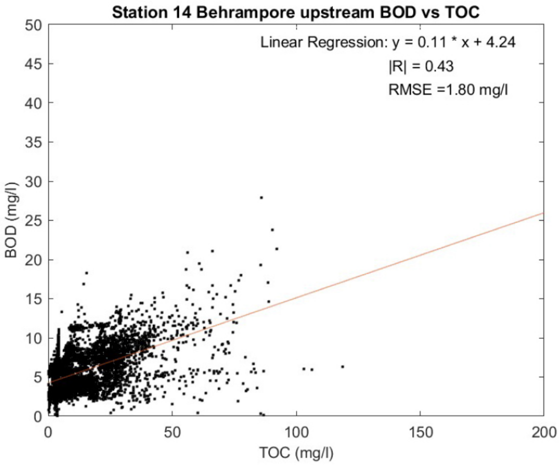

Station 9 Kanpur COD vs TOC
Linear Regression: y = 1.41 * x + 10.49
|R| = 0.77
RMSE =6.70 mg/l
COD (mg/l)
TOC (mg/l)
0
20
40
60
80
100
120
140
160
180
200
0
50
100
150
200
250

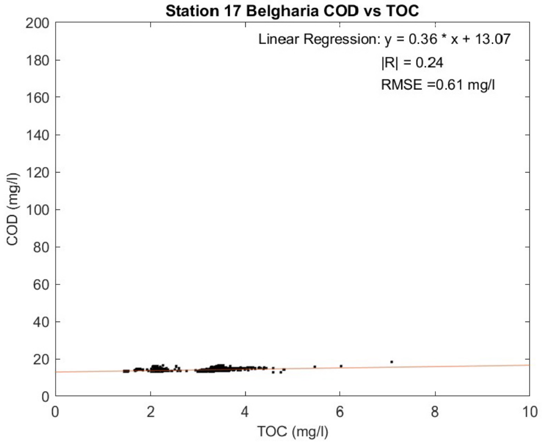

Station 9 Kanpur BOD vs TOC
Linear Regression: y = 0.22 * x + 1.66
|R| = 0.54
RMSE =2.01 mg/l
BOD (mg/l)
TOC (mg/l)
0
5
10
15
20
25
30
35
40
45
50
0
50
100
150
200
250

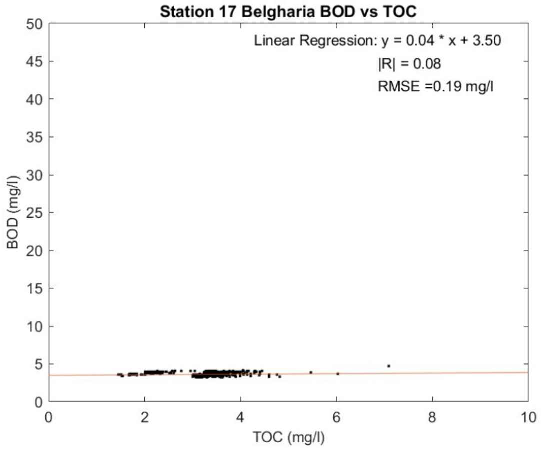

Station 12 Allahabad COD vs TOC

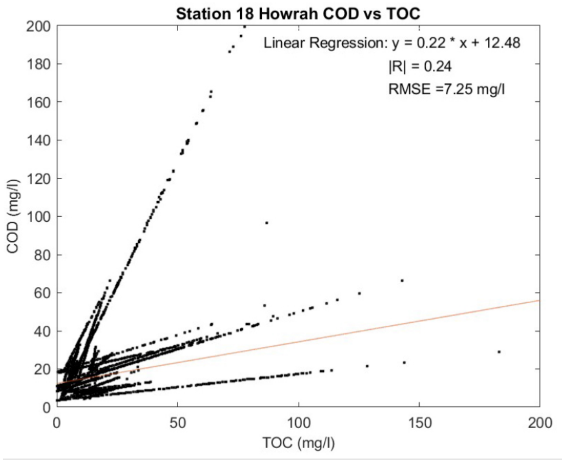

Station 12 Allahabad BOD vs TOC

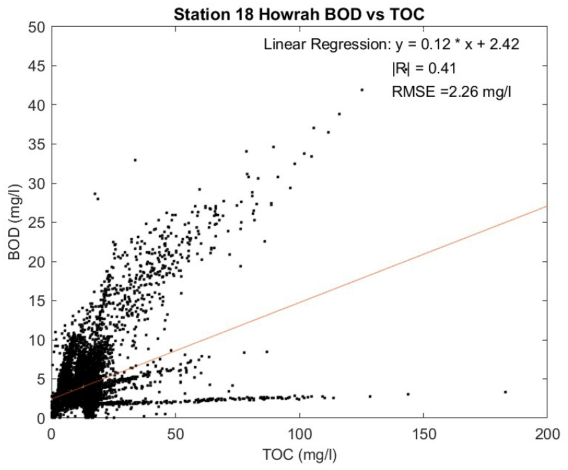

Station 14 Behrampore upstream COD vs TOC

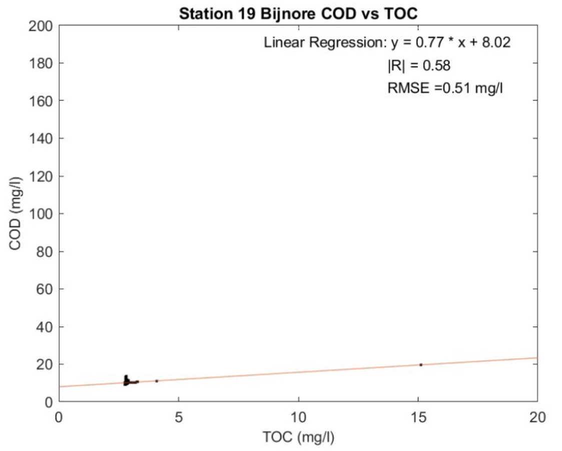

Station 14 Behrampore upstream BOD vs TOC

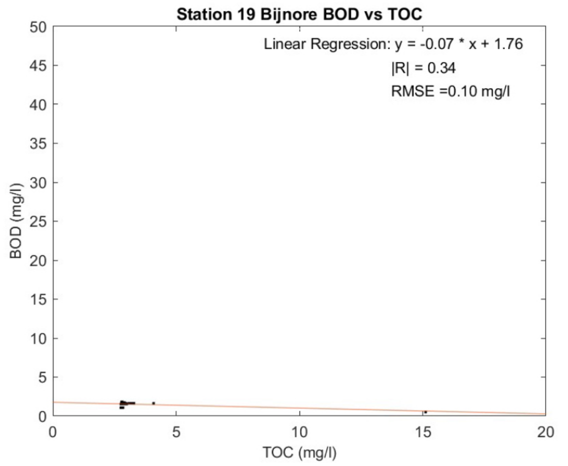

Station 17 Belgharia COD vs TOC

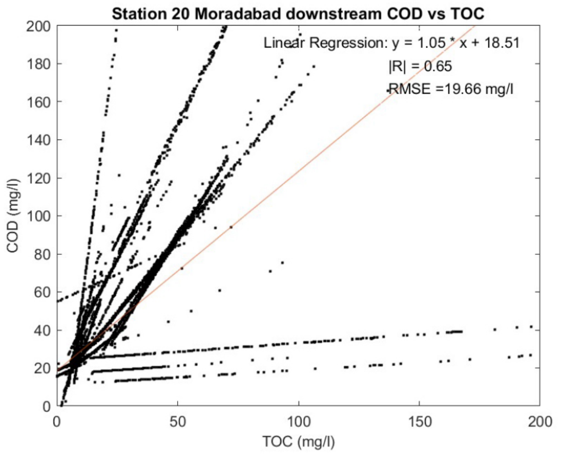

Station 17 Belgharia BOD vs TOC

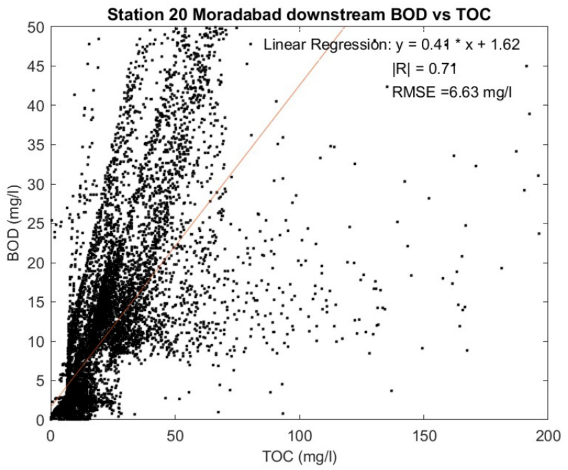

Station 18 Howrah COD vs TOC
Linear Regression: y = 0.22 * x + 12.48
|R| = 0.24
RMSE =7.25 mg/l
COD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
TOC (mg/l)
0
50
100
150
200

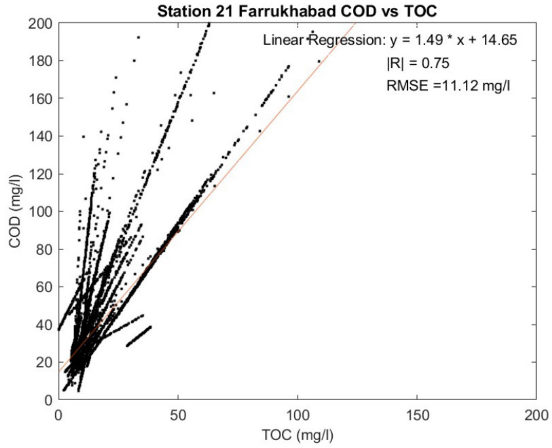

Station 18 Howrah BOD vs TOC
Linear Regression: y = 0.12 * x + 2.42
|R| = 0.41
RMSE =2.26 mg/l
BOD (mg/l)
0
5
10
15
20
25
30
35
40
45
50
TOC (mg/l)
0
50
100
150
200


Station 19 Bijnore COD vs TOC
Linear Regression: y = 0.77 * x + 8.02
|R| = 0.58
RMSE =0.51 mg/l
COD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
TOC (mg/l)
0
5
10
15
20

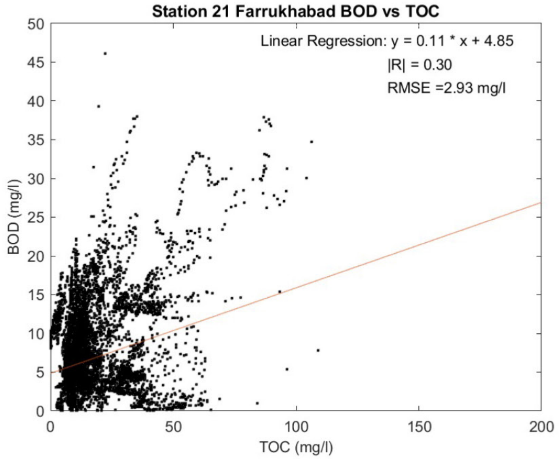

Station 19 Bijnore BOD vs TOC
Linear Regression: y = -0.07 * x + 1.76
|R| = 0.34
RMSE =0.10 mg/l
BOD (mg/l)
0
5
10
15
20
25
30
35
40
45
50
TOC (mg/l)
0
5
10
15
20


Station 20 Moradabad downstream COD vs TOC
Linear Regression: y = 1.05 * x + 18.51
|R| = 0.65
RMSE =19.66 mg/l
COD (mg/l)
0
20
40
60
80
100
120
140
160
180
200
TOC (mg/l)
0
50
100
150
200

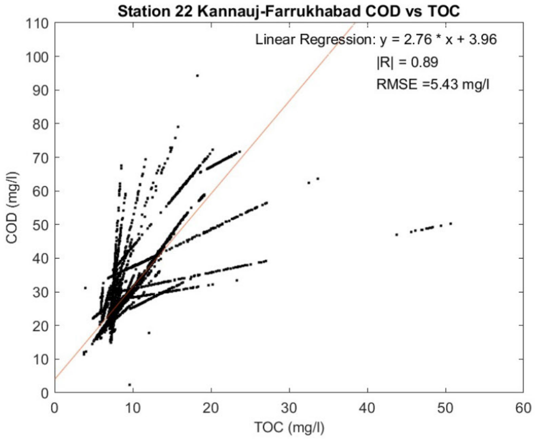

Station 20 Moradabad downstream BOD vs TOC
Linear Regression: y = 0.41 * x + 1.62
|R| = 0.71
RMSE =6.63 mg/l
BOD (mg/l)
0
5
10
15
20
25
30
35
40
45
50
TOC (mg/l)
0
50
100
150
200



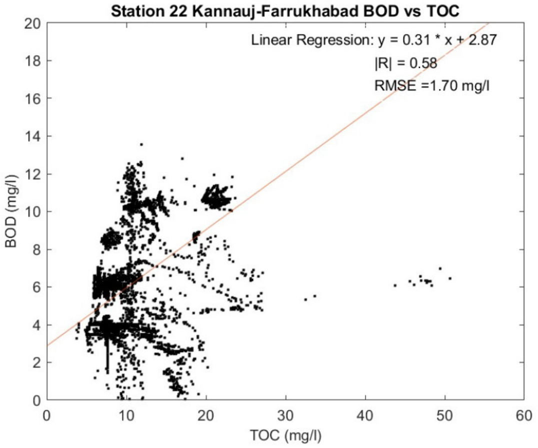
Station 21 Farrukhabad COD vs TOC



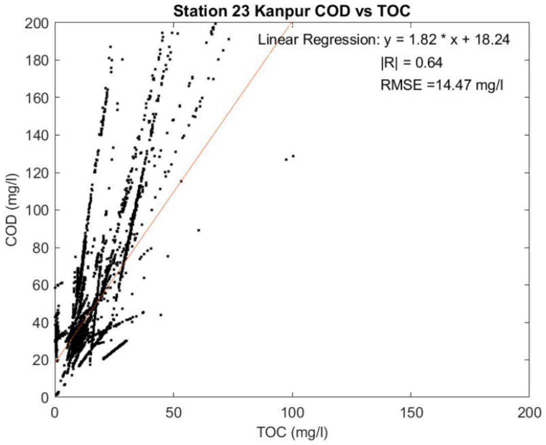
Station 21 Farrukhabad BOD vs TOC



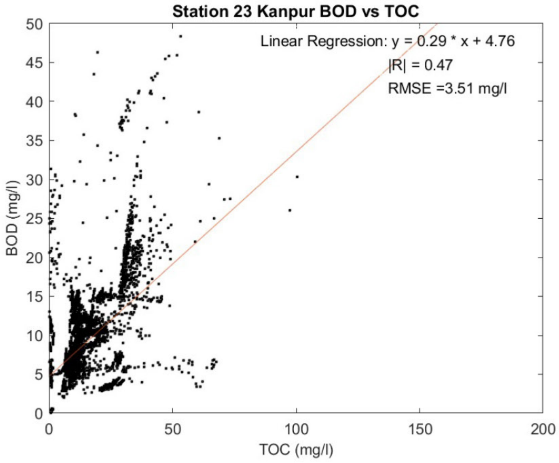
Station 22 Kannauj-Farrukhabad COD vs TOC



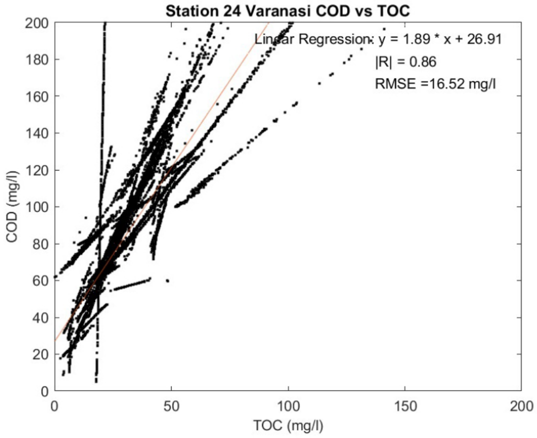
Station 22 Kannauj-Farrukhabad BOD vs TOC

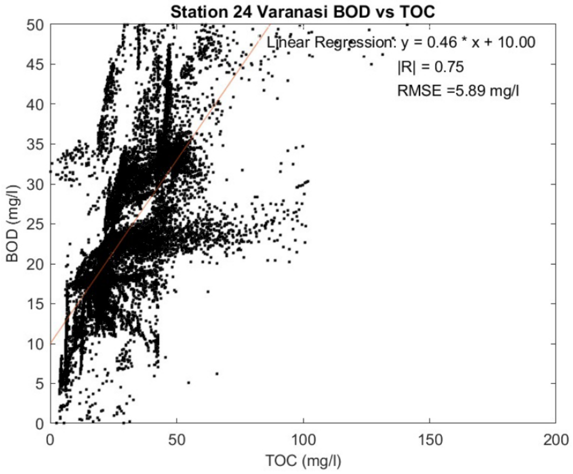


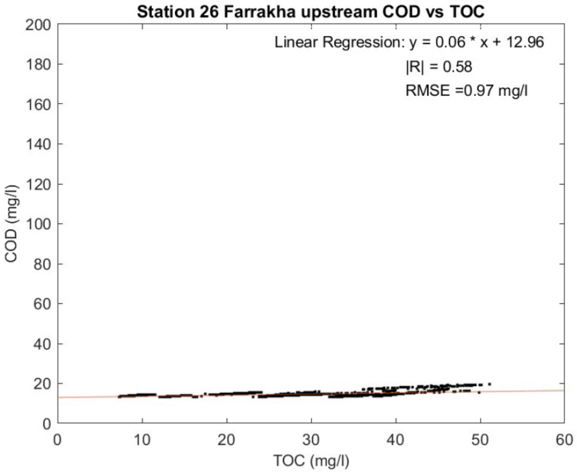
Station 23 Kanpur COD vs TOC

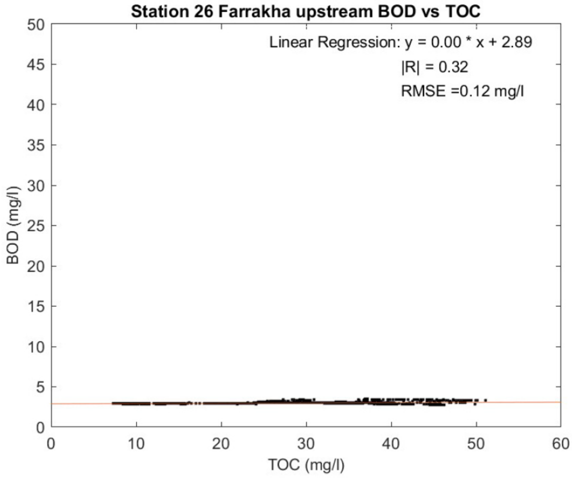


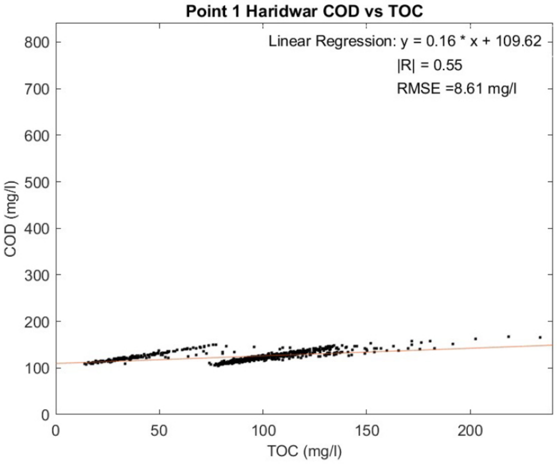
Station 23 Kanpur BOD vs TOC

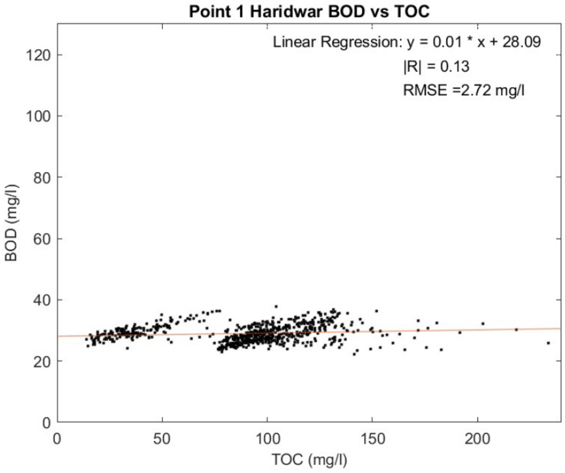

Station 24 Varanasi COD vs TOC


Station 24 Varanasi BOD vs TOC


Station 26 Farrakha upstream COD vs TOC


Station 26 Farrakha upstream BOD vs TOC


Point 1 Haridwar COD vs TOC


Point 1 Haridwar BOD vs TOC

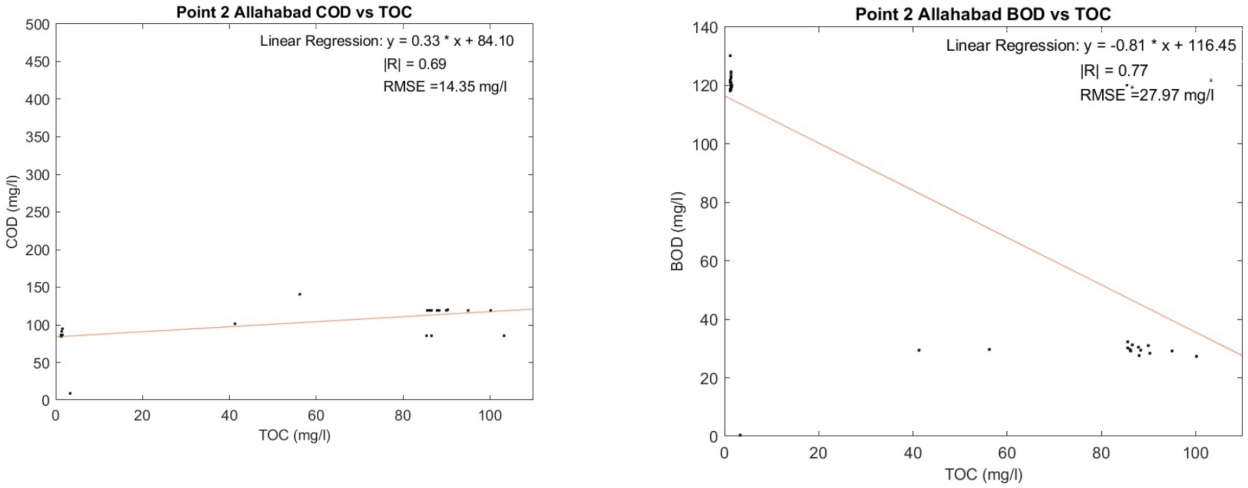


Figure 6: Regression plots of COD vs. TOC and BOD vs. TOC

5.0 Automated Sensor-Based Data Acquisition for Industrial Effluents

While statistical and other tests on river and drain water quality measurements clearly indicated grave discrepancies and errors in the sensor-based data for BOD, COD and TOC, further confirmation about the unreliability of the measurement techniques was obtained from direct comparisons of such measurements with those of standard methods carried out for wastewaters of Pulp and Paper Industries (PPIs). In general, PPIs are among the most polluting industries in India, particularly in the water sector. Based on the raw material (e.g. wood, bamboo, recycled fiber and rice husk) usage, industries can mainly be divided as wood pulp based, agro based and recycled chemical fiber (RCF) based, all of which may produce highly polluting wastewaters. Even when recycled paper is used for making pulp, significant chemicals are used in the de-inking process, which result in polluting effluents. The effluents contain color, high BOD and COD (due to presence of lignin and its derivatives from the raw cellulosic materials), chlorinated compounds, total suspended solids, fatty acids, tannins, resin acids, sulfur and its derivatives, etc. For effective implementation of CPCB's Charter on "Water Recycling & Pollution Prevention in Pulp and Paper industry", real time monitoring sensors had been installed for measuring five parameters of the industrial discharges – namely flow, pH, BOD, COD and TSS – in PPIs in the Ganga River basin. Of these, 94 PPIs located in Uttar Pradesh and 37 PPIs located in Uttarakhand have been studied to assess the pollution status of these industries and their likely effects on nearby waterbodies (cGanga and NMCG, 2019). The study was planned, surveyed, and sampled from April 2017 to July 2018. This report presents an evaluation of the reliability of the online sensor-based water quality measurement systems installed in these industries at the behest of CPCB for measuring BOD, COD and TSS vis-à-vis their measurements by standard methods.

The industries monitored in the above work are mainly clustered in four geographically proximal groups as shown in Figure 7, viz.:

1. Cluster 1 (Kashipur district, U.P.): 24 industries.
2. Cluster 2A (Meerut district, U.P.): 14 industries.
3. Cluster 2B (Muzaffarnagar district, U.P.): 33 industries.
4. Cluster 3 (Others): 60 industries.

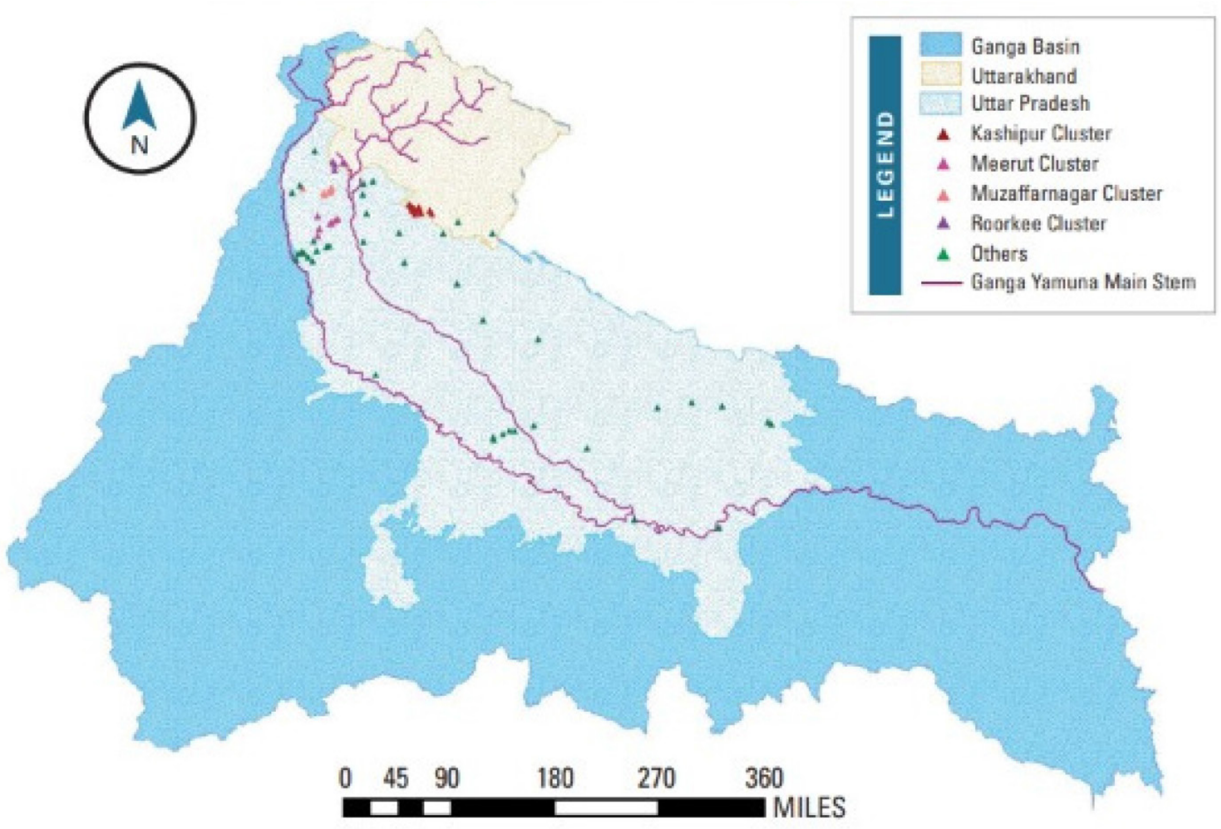


Figure 7: Location of 4 clusters within the administrative boundaries of the states of Uttarakhand and Uttar Pradesh

5.1 Evaluating the Reliability of Sensor-based BOD, COD and TSS data

At each industry real time monitoring sensors were installed for measuring flow, pH, BOD, COD and TSS of the industrial effluents. For trade effluents sampled during the same times these parameters were also independently analyzed by cGanga for each PPI using standard methods (for details, refer cGanga and NMCG, 2019).

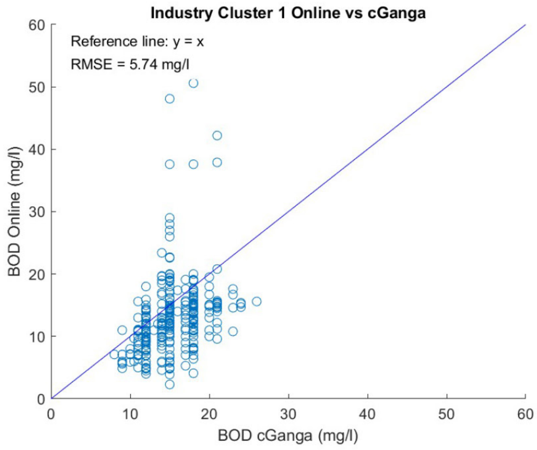
To evaluate the reliability of the data collected by the sensors installed at the industries, a direct comparison between them and the data measured by standard methods was made for BOD, COD and TSS respectively. Figure 8 presents the scatter plots for each of these variables for each industrial cluster. Ideally, each plot should show the data points falling on a straight line passing diagonally through the origin (at 45° angle) denoting the equation Y=X, subject to measurement errors. It is these measurement errors that are the subject of scrutiny from the plots.

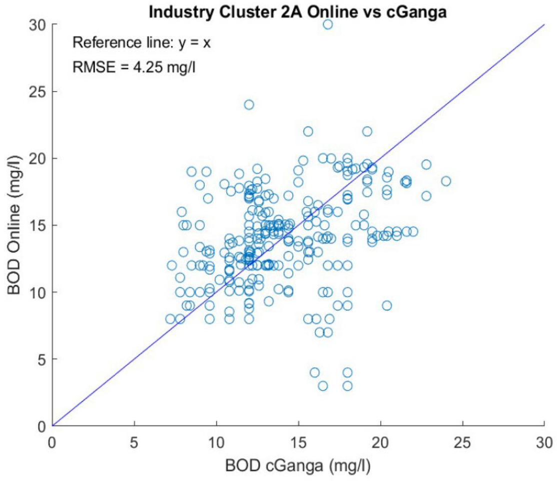
On going through the plots of Figure 8 the following clear discrepancies are observed:

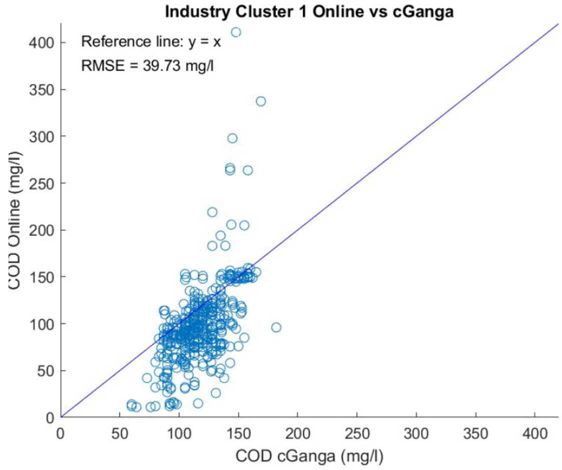
1) The scatter is very high in most plots.
2) In most plots the data do not lie along or close to the diagonal line, nor are they evenly distributed (being biased towards the right or left) around the diagonal.
3) The standard error of sensor-based measurements (root mean square error of the “y-variable” from the diagonal line) are very high in many cases.
4) In several plots the data for BOD and COD from sensor-based measurements (“y-variables”) tend to fall in a horizontal band, showing little variation in range.

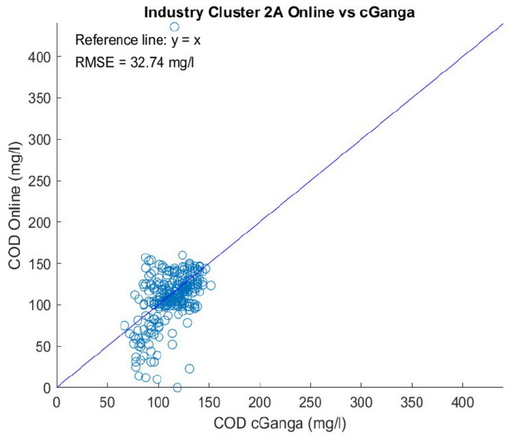
Overall, the above-mentioned four definitive inferences clearly show that the sensor-based monitoring data deviate strongly from measurements by standard methods and, hence, are largely erroneous.

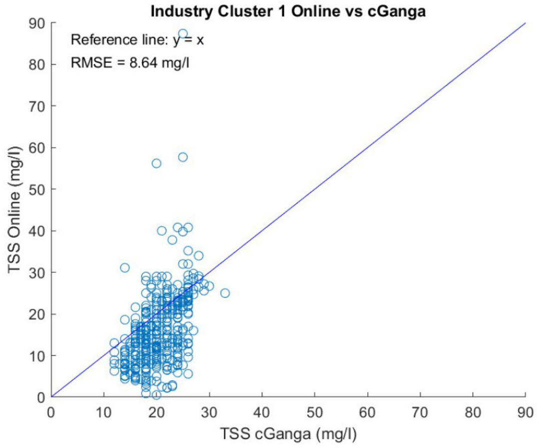
The industry- obtained data may not be very reliable since the industries themselves may consider their own test reports as superfluous in view of the real-time monitoring systems installed by regulatory authorities

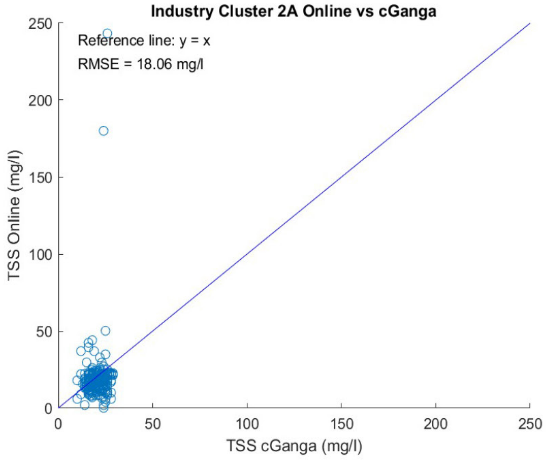

Industry Cluster 1 Online vs cGanga
Reference line: y = x
RMSE = 5.74 mg/l
BOD Online (mg/l)
BOD cGanga (mg/l)
0
10
20
30
40
50
60


Industry Cluster 2A Online vs cGanga
Reference line: y = x
RMSE = 4.25 mg/l
BOD Online (mg/l)
BOD cGanga (mg/l)
0
5
10
15
20
25
30


Industry Cluster 1 Online vs cGanga
Reference line: y = x
RMSE = 39.73 mg/l
COD Online (mg/l)
COD cGanga (mg/l)
0
50
100
150
200
250
300
350
400


Industry Cluster 2A Online vs cGanga
Reference line: y = x
RMSE = 32.74 mg/l
COD Online (mg/l)
COD cGanga (mg/l)
0
50
100
150
200
250
300
350
400


Industry Cluster 1 Online vs cGanga
Reference line: y = x
RMSE = 8.64 mg/l
TSS Online (mg/l)
TSS cGanga (mg/l)
0
10
20
30
40
50
60
70
80
90


Industry Cluster 2A Online vs cGanga
Reference line: y = x
RMSE = 18.06 mg/l
TSS Online (mg/l)
TSS cGanga (mg/l)
0
50
100
150
200
250

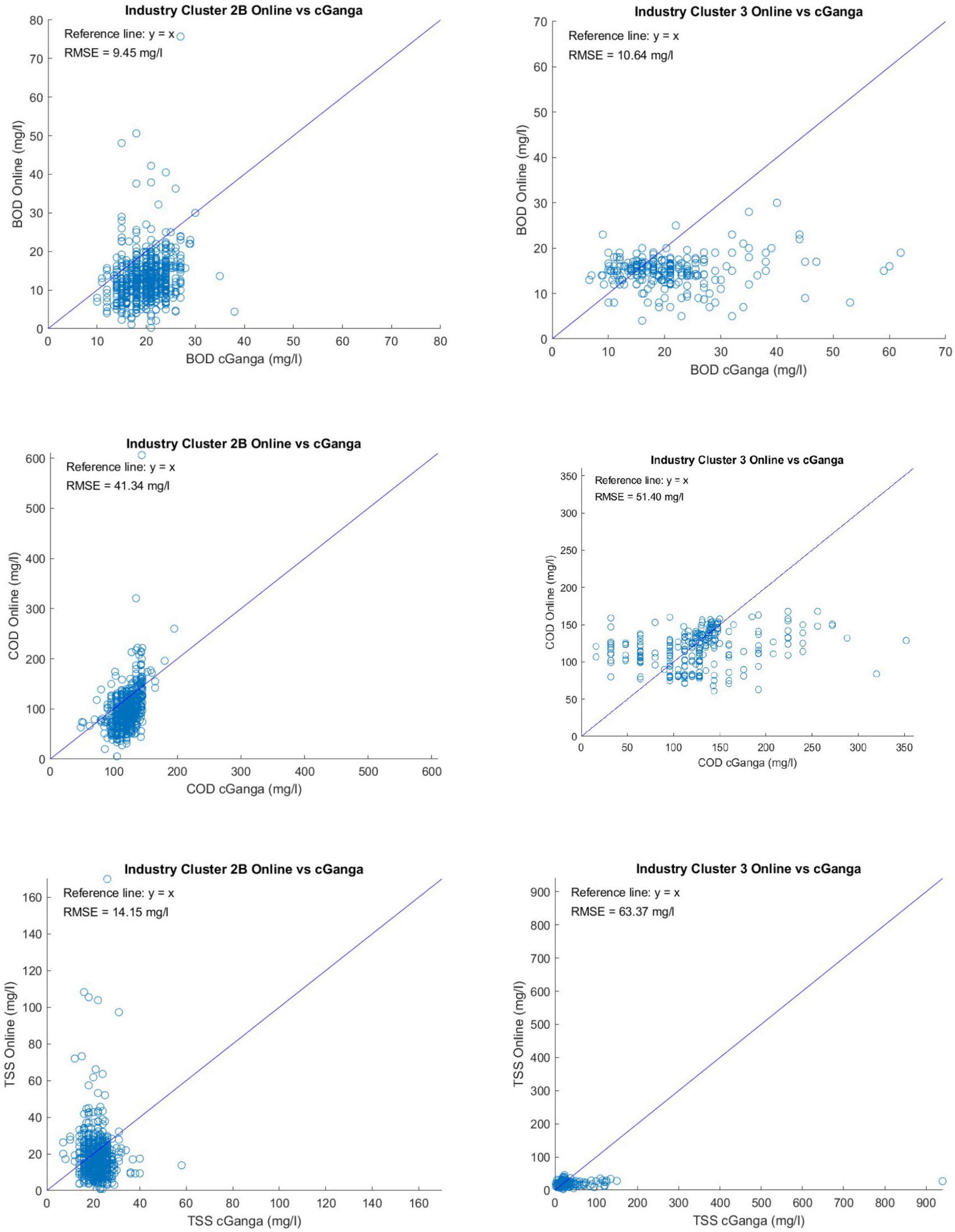


Figure 8: Scatter plots between BOD, COD and TSS from sensor-based measurements and by cGanga using Standard Methods

6.0 Conclusions

The analyses and discussions in the previous sections clearly show that the novel optical sensor-based measurements of natural watercourses (rivers and drains) and industrial effluents have apparently grave defects and almost certainly produce erroneous data. The main reasons for this conclusion are summarized below:

1. The scientific principles of optical sensor-based assessment of such parameters as BOD, COD and TOC (besides probably several others for which they have been applied) are non-existent or at best tenuous. In fact, how real-time optical assessment of slow biological process-based parameters such as BOD is possible defies imagination.
2. These methods are not standard methods recommended by any internationally renowned agency for testing the quality of natural waters.
3. These methods are not used in advanced countries (including even the parent countries of the manufacturers/ suppliers of the sensor instruments adopted by our regulatory agency) for assessing the water quality of rivers or natural waterbodies.
4. No information has been released by the regulatory agency on the validity of these novel methods, their test reports in Indian conditions (or even elsewhere), or their accuracies and limitations for natural waters (and other non-homogeneous flows).
5. The statistical tests on the sensor-generated data presented in the previous two sections show too many discrepancies and large errors, clearly indicative of defective data produced by such measurements.

Thus the data produced from such measurements may be spurious and extremely misleading. Spending significant public money and resources for generating such data may be itself bad in practice. But if such misleading data generated through public money are not used for any internal purpose by an agency, but for public use (such as river/drain monitoring or industrial discharge monitoring) instead, then it can be vastly harmful for managing our rivers and waterbodies. Hence, in the broader context of natural resource management in India, the issue calls for an urgent and clearly defined policy/ protocol for introducing new and non-standard methods of natural resource measurements.

In the broader context of natural resource management in India, the issue calls for an urgent and clearly defined policy/ protocol for introducing new and non-standard methods of natural resource measurements

References

1. APHA (195): "Standard Methods for the Examination of Water and Wastewater", American Public Health Association (APHA), Wasington DC.
2. cGanga, NMCG (2019): "Strategy for Improving Condition of Water Bodies in the vicinity of Pilp and Paper Industries in Ganga River Basin."
3. Chaudhary, A (2021): "Assessment of Real Time Water Quality Data of Rivers and Drains Acquired Using Sensors in the Middle Segment of the Ganga Basin", M.Tech. Thesis, IIT Kanpur (October 2021).
4. Montgomery DC, Runger GC (2016): "Applied Statistics and Probability for Engineers (6th Edn.)", Wiley Student Edition.
5. Hanrahan G, Casey H, Worsfold PJ (2005): WATER ANALYSIS | Freshwater. In: Worsfold P, Townshend A, Poole C (eds) Encyclopedia of Analytical Science (Second Edition), Elsevier, Oxford, pp 262–268.
6. Hodge V, Austin J (2004): A survey of outlier detection methodologies. Artif Intell Rev 22:85–126.
7. Hyldgard A, Olafsdottir I, Olesen M, et al (2005): FISH & CHIPS: four electrode conductivity/salinity sensor on a silicon multi-sensor chip for fisheries research. IEEE, pp 1124–1127.
8. Lee Rodgers J, Nicewander WA (1988): Thirteen ways to look at the correlation coefficient. Am Stat 42:59–66.
9. Lindon JC, Tranter GE, Koppenaal D (2016): Encyclopedia of spectroscopy and spectrometry. Academic Press.
10. Osborne JW, Overbay A (2004): The power of outliers (and why researchers should always check for them). Practical Assess Res. Eval. 9:6.
11. Restrepo LF, González J (2007): From pearson to Spearman. Rev Colomb Cienc Pecu 20:183–192.
12. Seo S (2006): A review and comparison of methods for detecting outliers in univariate data sets. University of Pittsburgh. Pittsburgh, USA.
13. Wei Y, Jiao Y, An D, et al (2019): Review of Dissolved Oxygen Detection Technology: From Laboratory Analysis to Online Intelligent Detection. Sensors 19: https://doi.org/10.3390/s19183995.

cGanga

Centre for Ganga River Basin Management and Studies