LIMNOTEK Perairan Darat Tropis di Indonesia
(Tropical Limnology and Water Resources)
p-ISSN: 0854-8390 e-ISSN: 2549-8029
https://ejournal.brin.go.id/limnotek

Assessing Data Quality in Real-Time River Water Quality Monitoring Using Multi-Parameter Indicators: A Case Study of Ciliwung and Cisadane, Indonesia

Satmoko Yudo[1*)], Dwi Nowo Martono[2], Dwita Sutjiningsih[2], and Budi Kurniawan[3]

[1] *Research Center for Environmental and Clean Technologies, National Research and Innovation Agency*
[2] *Graduate School of Sustainable Development, University of Indonesia*
[3] *Research Center for Limnology and Water Resources, National Research and Innovation Agency*

Article Info

Keywords:
River water quality monitoring
Internet of Things (IoT)
Online data quality evaluation
Data quality assessment
Data governance
Stability index

***Corresponding author:**
E-mail address: satm001@brin.go.id
(Satmoko Yudo)

https://doi.org/10.55981/limnotek.2026.14216

Received: 26 November 2025; Accepted: 5 June 2026; Published: 20 June 2026

Abstract

Water pollution has become a complex environmental challenge in rapidly urbanizing Southeast Asia. Indonesia has deployed over 347 online river water quality monitoring stations (ORWQM) in several watersheds to provide real-time data for pollution control and informed ecological decision-making. However, problems have arisen with some of the longstanding stations, including equipment degradation, data loss, sensor malfunctions, and inconsistencies, raising concerns about the reliability of these datasets. This study aims to address these issues by using a hybrid Total Data Quality Management (TDQM), Online Data Quality Evaluation Model (ODQEM) framework to evaluate and improve the data quality at four stations (Manggarai, Kelapa Dua, Pasar Baru, and Empang Dam) representing the upstream and downstream areas of the Ciliwung and Cisadane Rivers. The workflow consists of four steps: (i) defining the relevant data quality dimensions; (ii) measuring data quality using completeness index, range validity, accuracy, uniqueness, and stability index; (iii) analyzing cross-parameter and cross-station patterns to identify potential failure modes and governance gaps; and (iv) translating findings into targeted calibration, cleaning, and maintenance actions. The results show strong structural integrity; completeness indicators reach 100%, and uniqueness exceeds 99.9%, indicating robust acquisition and temporal consistency. However, functional reliability varies widely across parameters and station locations. Completeness indicators for non-zero values show systematic zero values for the DO probe (1.02%) and nitrate (0.19%) at Empang Dam Station; TDS (9.57%) at Pasar Baru Station; ammonia (60.02%) at Kelapa Dua Station. Overall stability is high for physical probes (temperature, pH), but low for chemical/ion probes at some stations (salinity 1.47%–6.62%; TDS 5.49%–10.92%; ammonia 30.20%–71.40%). Ion sensors also show higher risks, including low validity or accuracy for nitrate at Kelapa Dua Station (61.67%). These findings indicate that the dataset appears complete and valid, but still contains substantial functional uncertainty due to several zero-value parameters and unstable sensor behavior. Because the evaluation is limited to four stations, these findings should be interpreted as an in-depth case study rather than a complete representation of the ORWQM network in Indonesia. The proposed TDQM–ODQEM Data Governance Model offers a replicable method for improving online and real-time environmental monitoring, helping bridge the gap between technological performance and policy design in sustainable river management.

1. Introduction

River pollution, caused by various pollutants, significantly impacts aquatic ecosystem degradation, reduces water quality, and poses a threat to human health (Lin et al., 2022). One preventative measure to control pollution is the implementation of continuous monitoring and strict supervision to maintain the river water quality status. The online river water quality monitoring (ORWQM) system consists of monitoring stations installed along riverbanks. This system is equipped with multiparameter sensors that continuously measure and transmit water quality data in real-time to a central database via the internet. As of 2024, 347 ORWQM stations have been established across several watersheds (DAS) in Indonesia (KLHK, 2024). This monitoring plays a crucial role in water resource management, providing insight into river water conditions and tracking changes over time (Soares et al., 2020). Furthermore, long-term water quality monitoring is crucial for detecting ecological changes in river ecosystems and ensuring compliance with water quality standards (Meyer et al., 2019). The real-time data obtained from this monitoring system is invaluable for developing innovative studies that capture dynamic temporal variations, including water quality prediction, assessment, and environmental management (Zhang & Thorburn, 2022). This information provides a crucial foundation for developing policies and planning strategies for water quality management at the national, provincial, and local levels of government (Adu-Manu et al., 2020). Therefore, the main challenge is not only ensuring the continuous transmission of data but also establishing trust in the ORWQM data, analytically, and in its suitability for regulatory interpretation, early warning, and evidence-based environmental management. This sharper focus addresses the practical need to assess whether the available data in the system is sufficiently reliable to support decision-making in environmental policy.

A common problem in real-time water quality monitoring is data loss, often caused by equipment failure, limited network coverage, or data corruption (Zhang & Thorburn, 2022). As Internet of Things (IoT)-enabled water quality monitoring expands, the likelihood of system and network failures also increases significantly (Zhang et al., 2021). These failures can lead to lost data, duplicate records, or outliers, which reduce data quality and affect the integrity of analysis results. Data quality is defined as the extent to which data is suitable for its intended use in operations, decision-making, and planning (Mansouri et al., 2023). Inaccurate sensor data can cause misinterpretation of environmental conditions, resulting in inappropriate policy responses or inefficient resource allocation (Teh et al., 2020). Furthermore, poor data quality directly impacts the effectiveness of organizational processes. It results in misinformation, affects reporting accuracy, increases operational risks, and often requires costly data reprocessing or cleansing efforts from a strategic perspective (Wijayanti et al., 2018; Bowo et al., 2019). Data quality is a key factor in evidence-based decision-making, especially in fields such as environmental protection, urban water management, and public health. High-quality data supports accurate modeling, risk forecasting, and automation systems in innovative environmental initiatives. However, maintaining this quality in an IoT environment demands robust system architecture, real-time validation protocols, standardized data formats, and coordination among stakeholders (Karkouch et al., 2016; Pezoulas et al., 2019). In Indonesian river systems, particularly in urban areas like Jakarta and Tangerang, pollution levels remain high due to increasing discharges of domestic waste, industrial waste, and urban runoff (Sulthonuddin et al., 2019). Therefore, high-quality monitoring data is essential for sustainable watershed (DAS) management and mitigation programs. However, systematic research on the quality of multidimensional ORWQM data in Indonesia remains limited, especially studies that distinguish between aspects such as structural data integrity (including completeness and timestamp uniqueness) and the reliability of operational sensors (including stability and parameter anomaly detection) in IoT-based monitoring systems. Most previous studies have focused more on the technical aspects of water quality monitoring or analysis itself (Meyer et al., 2019; Soares et al., 2020), but have not explicitly addressed the integrity and reliability of the data generated by the system. Although the literature on water quality monitoring systems is extensive, most focus on hydro chemical interpretation, pollution source analysis, and sensor placement strategies, rather than on evaluating monitoring data quality as an asset supporting decision-making (Chapman, 1996; Behmel et al., 2016; Meyer et al., 2019; Soares et al., 2020). Many IoT-based environmental monitoring studies highlight sensing architecture and communication systems but provide less of a multidimensional framework to assess whether the generated time-series data remains reliable during long-term operation (Karkouch et al., 2016; Adu-Manu et al., 2020; Teh et al., 2020). As a result, operational data issues such as stagnation, instability in certain parameters, and records that appear complete but are unusable are often treated as technical disruptions rather than methodological and governance problems that affect policy credibility.

Data quality is crucial for the effective use of information in policy formulation, managerial decision-making, and responding rapidly to pollution and changing environmental conditions. Furthermore, the knowledge gap is evident in the lack of studies employing a comprehensive conceptual framework, such as Total Data Quality Management (TDQM), that enables comprehensive data assessment across multiple data quality dimensions (Batini & Scannapieco, 2016). However, TDQM itself lacks an operational mechanism for assessing sensor performance in real time. In dynamic environments such as tropical rivers.

This gap is particularly pronounced in Indonesia, where ORWQM systems have strategic policy relevance, but published studies systematically assessing ORWQM data quality using a multidimensional framework are limited. What is needed is not only reporting descriptive sensor performance, but also identifying which data quality dimensions are structurally strong, which are operationally weak, and how these weaknesses should be considered in maintenance priorities and institutional governance.

To close this gap, this paper integrates TDQM with the Online Data Quality Evaluation Model (ODQEM), which was initially formulated for continuous IoT sensor networks. ODQEM provides parameter-specific temporal metrics such as the Stability Index (SI) and Range Validity (RV), which quantify sensor reliability under fluctuating field conditions. This hybrid approach enables multi-level evaluation, utilizing TDQM for strategic data governance and ODQEM for real-time operational assessment.

Therefore, the objectives of this study are to evaluate the quality of online river water quality monitoring (ORWQM) data from selected stations in the Ciliwung and Cisadane Rivers using an integrated TDQM–ODQEM framework to identify key weaknesses affecting data reliability and to support improvements in institutional data management and maintenance planning.

2. Methodology

2.1. Data Sources and Station Locations

The data used in this study were river water quality monitoring records collected between 2016 and 2020. The year 2016 marked the beginning of the construction of the four stations (Manggarai, Kelapa Dua, Empang Dam, and Pasar Baru), and 2020 marked the end of the database storage on the server of the BJ Habibie Science and Technology Area (KST BJ Habibie), Serpong, South Tangerang City, Banten Province. Subsequently, from 2021 onward, the database was transferred to a server at the Ministry of Environment and Forestry (KLHK). The Kelapa Dua and Manggarai stations selected represent the upstream and downstream segments of the Ciliwung River in the DKI Jakarta area, while Empang Dam and Pasar Baru stations represent the upstream and downstream segments of the Cisadane River in the Bogor and Tangerang areas. Figure 1 illustrates a map of the monitoring stations and data center facilities, while Table 1 provides the addresses and geographic coordinates of the four station locations.

An advanced data logger automatically manages the process of storing real-time, online water quality monitoring data. This device operates using two sampling intervals: the Early Warning System (EWS) interval and the periodic interval. Water quality measurements were carried out using a WQC-type multi-probe sensor with 11 parameters; in this study, only 10 were analyzed: temperature, DO, ammonia, conductivity, pH, TDS, turbidity, salinity, nitrate, and ORP. This sensor is equipped with a handheld to display the measurement results directly. Of the two data collection intervals, only data from the periodic measurements are stored in internal memory and transmitted to a central database via SMS or internet data services. The central server receives and stores monitoring data in a structured relational database system, which enables data retrieval through a web-based application interface designed for environmental data analysis and reporting.

In operational practice, calibration and maintenance are carried out following the technical guideline procedures from each vendor. This includes cleaning the probes (removing biofouling), calibrating standard solutions (such as pH buffer and conductivity standards), inspecting cables and connectors, checking telemetry, inspecting power subsystems, and recalibrating or resetting sensors after disturbances. Since complete records of calibration and historical maintenance are not always available in digital format for all stations and time periods, this data is not analyzed as a separate dataset. However, it very likely influence on data quality indicators is explicitly considered during result interpretation.

2.2. Integrated TDQM-ODQEM Framework

Data stored in a structured relational database system, then, raw data, including station identification, date, time, and multiparameter water quality observations, were downloaded from the database for analysis. Before conducting TDQM-ODQEM analysis, raw data is first processed through a transparent quality control workflow to ensure reproducibility. The preprocessing steps include: (i) extracting and organizing data based on station and parameter; (ii) standardizing the date-time field into a single timestamp format; (iii) sorting and identifying duplicate timestamps; (iv) detection of missing values (without imputation for completeness calculations); (v) detection of zero values using a formula check for all parameters; (vi) range-based outlier detection using established physical/sensor validity limits (Table 2); and (vii) harmonizing parameter names and units across stations.

To assess the quality of the data collected, this study utilized the Total Data Quality Management (TDQM) framework introduced by Wang (1998). TDQM is a proven methodology that offers a systematic way to manage and evaluate data quality through an ongoing improvement cycle (see Figure 2), consisting of four repeating steps: define, measure, analyze, and improve (Wang, 1998). The TDQM framework highlights that data should be regarded as a strategic asset, requiring proactive quality assurance measures throughout its lifecycle (Mohamed et al., 2009). The first step, Define, aims to identify user needs and establish relevant data quality dimensions such as accuracy, completeness, validity, and uniqueness. During this phase, quality standards and business rules are set as the basis for measurement. Next, the Measure stage is conducted to evaluate how well the available data meets the established standards using measurable metrics, such as the ratio of valid values, detection of extreme values, or comparison with reference data. After measurement, the analyze phase is used to find the root causes of poor data quality, including technical, procedural, and operational issues. This step helps identify where and why errors or discrepancies happen.

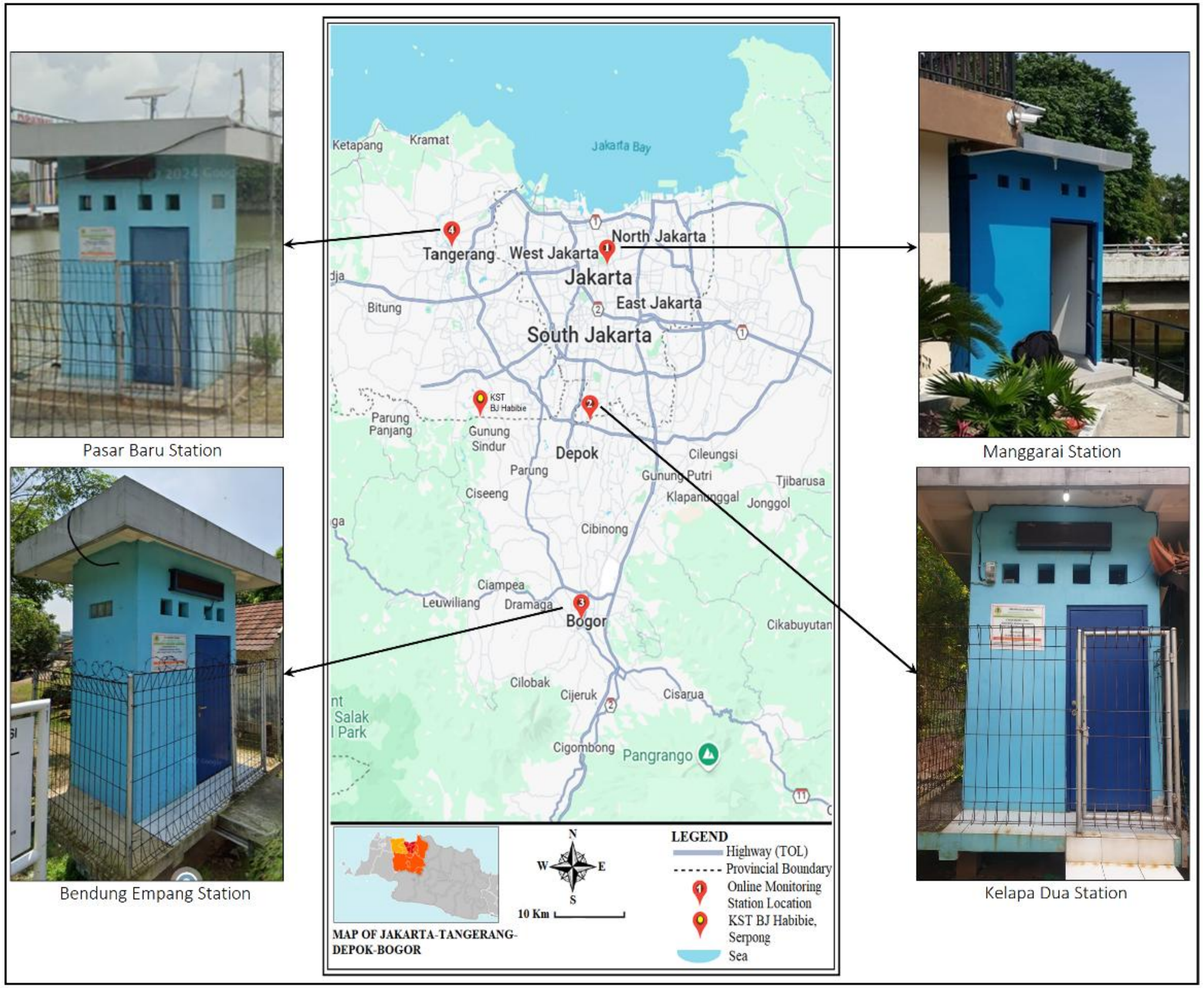

Figure 1. Location map of four monitoring stations and data centers at KST BJ Habibie, Serpong, South Tangerang City, Banten Province, Indonesia

Table 1. Location of four online monitoring stations on the Ciliwung River and Cisadane River

No.	Station Name	River	Address	Coordinate	
				Latitude	Longitudes
1.	Manggarai Station	Ciliwung River	Jl. Tambak, RT.07/ RW.04, Pegangsaan Subdistrict, Menteng District, Central Jakarta, DKI Jakarta	-6.20784	106.84850
2.	Kelapa Dua Station	Ciliwung River	Arus Hill, Srengseng Sawah, Jagakarsa, South Jakarta, DKI Jakarta.	-6.35264	106.83564
3	Empang Dam Station	Cisadane River	Cisadane Mountain, Paledang, Bogor City, West Java Province	-6.60777	106.79288
4	Pasar Baru Dam Station	Cisadane River	Jl. KS Tubun, Koang Jaya, Tangerang City, Banten Province	-6.16079	106.62772

Lastly, the Improve phase involves implementing corrective actions like sensor calibration, operator training, automated validation systems, or enhanced data integration. After improvements are implemented, the process loops back to the Define stage to update data quality standards and strategies based on previous lessons learned, creating a continuous cycle of data quality management. This framework has been widely adopted in Internet of Things (IoT)-based systems, including online water quality monitoring, due to its capacity to systematically and measurably uphold data integrity and reliability.

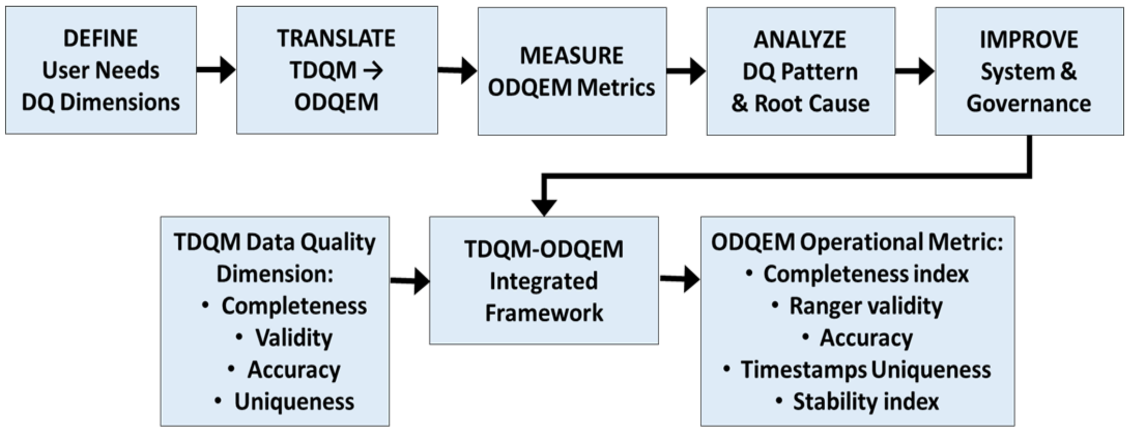

Figure 2. Four main components of the TDQM cycle
(Source: Wang, 1998; Klein and Lehner, 2009)

This diagram shows a hybrid data quality approach that combines the conceptual cycle of TDQM with the operational indicators of the Online Data Quality Evaluation Model (ODQEM). The process starts with defining user needs and the main data quality dimensions of completeness, validity, accuracy, and uniqueness (Cahyono et al., 2020), which form the basis for assessing real-time monitoring data. To evaluate the data quality collected from four stations, this study uses a combined method of Total Data Quality Management (TDQM) and the proposed Online Data Quality Evaluation Model (ODQEM). TDQM provides a conceptual framework based on common data quality dimensions, including completeness, validity, accuracy, and uniqueness (Wang & Strong, 1996; Wang, 1998). This framework is then tailored to the context of sensor-based online water quality monitoring, using ODQEM technical indicators such as the Completeness Index (CI), Range Validity (RV), Stability Index (SI), and Uniqueness (duplicate check). This integration allows data quality assessment not only from a technical standpoint (sensors) but also in line with global data quality management standards. Below is an explanation of each ODQEM indicator.

a. Completeness index (*CI*)

The Completeness Index (CI) measures the proportion of available data relative to the total observation period. In environmental datasets, completeness is vital because missing data can greatly affect analysis results (Batini & Scannapieco, 2016). In an online water quality monitoring system, completeness indicates whether data are recorded consistently without missing values caused by communication issues, power outages, or sensor failures. The CI formula is expressed as:

$$Completeness\ Index = \frac{Number\ of\ recorded\ (non-null)\ values}{Number\ of\ value\ that\ should\ have\ been\ recorded} \times 100\% \quad (1)$$

A high CI (>95%) indicates a well-functioning system, while a moderate CI (70–90%) indicates some incomplete data, possibly due to sensor failure or extreme river conditions. Completeness assessment is one of the first steps in data quality (Behmel et al., 2016; Li et al., 2018).

b. Range Validity (RV)

The range validity (RV) indicator evaluates whether sensor data is within acceptable limits, considering both the physical range and water quality standards (Table 2). This is crucial for filtering out unreasonable or out-of-specification readings that may result from sensor errors or data transmission issues (Teh et al., 2020). Sensor errors, drift, or extreme hydrological conditions often cause values to exceed these limits. The RV formula can be expressed as:

$$Range\ Validity = \frac{Number\ of\ data\ within\ standard\ range}{Number\ of\ data} \times 100\% \quad (2)$$

In this study, the Validity Range (VR) interpretation approach is operationally classified into three levels: high (≥90%), moderate (70–<90%), and low (<70%). These thresholds are used as research-based evaluation criteria to indicate the proportion of observations within a reasonable range. This interpretation aligns with quality control principles in sustainable water quality monitoring, where unreasonable values are typically addressed through rough range checks and further investigated for possible sensor damage, deviations, or calibration issues (Bushnell & Worthington; 2016 U.S. EPA, 2018).

Range validity is widely used to assess the quality of environmental sensor data (Ramirez et al., 2011). It is a basic and standard quality control (QC) technique that confirms sensor readings stay within specified upper and lower bounds.

c. Stability Index (SI)

The Stability Index quantifies the proportion of data that remains unchanged over time, specifically values that do not vary over ≥3 consecutive measurements (Qartod, 2018). Flatlining can occur if the sensor is covered with sediment (fouling), loses calibration, or sustains technical damage (Bushnell et al., 2015). This indicator has been used to evaluate the quality of environmental sensor data (Zhang et al., 2022).
Measuring the percentage of flatline data:

$$SI = \frac{Number\ of\ flatline\ segments}{Total\ segments} \times 100\% \quad (3)$$

In this study, the Stability Index (SI) is operationally defined as follows: high (>90%), moderate (70–90%), low (50–70%), and very low (<50%). These limits are used as research-based assessment criteria, not as universal standards. This interpretation aligns with the principles of continuous water quality QA/QC, where decreasing signal stability may indicate increasing anomalies and requires further verification through field inspection, comparison with discrete/manual samples, sensor maintenance, fouling checks, deviation correction, and recalibration (Qartod, 2018).

d. Accuracy

Accuracy was assessed as the percentage of sensor readings within a predetermined acceptable range based on standards or regulations (e.g., river water quality standards), and was calculated while the formula (Redman, 2005) used is as follows:

$$Accuracy = \frac{(Number\ of\ valid\ (in\ range)}{Total\ values} \times 100\% \quad (4)$$

Data accuracy was assessed by comparing sensor reading them with laboratory data (when available) or masked tests using imputation techniques. The metric used was the Mean Absolute Error (MAE) (Zhang & Thorburn, 2022). MAE measures the average absolute difference between sensor values and laboratory values, regardless of whether the difference is positive or negative.

$$MAE = \frac{1}{n}\sum |Y_{sensor} - Y_{lab}| \quad (5)$$

Where
n = number of data or observations
Y_{sensor} = actual value (sensor value)
Y_{lab} = predicted value or measurement result (lab value)
| ... | = absolute value

e. Uniqueness

Uniqueness indicators are used to ensure data uniqueness, thereby maintaining the integrity of time series analysis and forecasting models (Qin et al., 2016). In online river water quality monitoring systems, uniqueness primarily concerns timestamps since each measurement must be unique. If two datasets share the same timestamp, duplication occurs, which reduces data quality. The uniqueness index is calculated by dividing the number of unique data points by the total number of data points (Ehrlinger & Wöß, 2022).

$$UQ = \frac{Number\ of\ unique\ values}{Total\ values} \times 100\% \quad (6)$$

The dataset evaluated includes 15 key attributes listed in Table 3, which make up the main measurement records from four stations, totaling 66,505 entries. Each record has 15 columns, including the station ID, date, time, and 12 water quality parameters.

Table 2. Acceptable range for each water quality parameter

No.	Parameter	Range
1	BOD	0.1 - 60 mg/L
2	COD	0.1 - 500 mg/L
3	Temperature	0° C - 50° C
4	DO (Dissolved Oxygen)	0 - 15 mg/L; or 0 - 200%
5	pH	0 - 14 units
6	Nitrate	0 - 50 mg/L
7	Nitrite	0 – 10 mg/L*
8	TSS	0 - 500 mg/L
9	TDS / conductivity / salinity	0 - 100,000 µS/cm or 0 - 100 mS/cm
10	Turbidity	0 - 1000 NTU
11	Ammonia	0 - 100 mg/L as N /0 – 10 mg/L **
12	*Depth (pressure)*	0 - 10 m or more

(Source: Regulation of the Minister of Environment and Forestry No. 2 of 2022;
*Regulation of the Minister of Health of the Republic of Indonesia No. 492 of 2010;
** Regulation of the Minister of Environment and Forestry No. 68 of 2016)

Table 3. Description of river water quality monitoring data attributes

No.	Attribute Name	Data Type	Description
1	IDStation	Char(10)	Monitoring station identifier
2	Date	Date	Date of measurement
3	Time	Time	Time of measurement
4	Temperature	Double (5.2)	Water temperature (°C)
5	EC	Double (5.2)	Electrical Conductivity (μS/cm)
6	TDS	Double (5.2)	Total Dissolved Solids (mg/L)
7	Salinity	Double (5.2)	Water salinity (ppt)
8	DO	Double (5.2)	Dissolved Oxygen (mg/L)
9	pH	Double (5.2)	Acidity/alkalinity level
10	Turbidity	Double (5.2)	Water turbidity (NTU)
11	Depth	Double (5.2)	Water depth at the monitoring point (m)
12	SwSG	Double (5.2)	Seawater Specific Gravity
13	Ammonia	Double (5.2)	Ammonia concentration (mg/L)
14	Nitrate	Double (5.2)	Nitrate concentration (mg/L)
15	ORP	Double (5.2)	Oxidation-Reduction Potential (mV)

(Source: BPPT, 2015)

3. Results and Discussion

3.1. Validity Range Assessment

The results show that most physical parameters, including temperature, conductivity, TDS, salinity, and turbidity, remained within the valid range across all monitoring stations (90%-100%). This pattern suggests that the sensors operated within their specified limits and did not produce values outside the expected environmental ranges. This performance supports the ODQEM assumption about parameter robustness. It also aligns with the observations of Karkouch et al. (2016), who emphasized the inherent stability of physical IoT water quality sensors compared to biochemical probes.

In contrast, nutrient parameters exhibited significant variability. Nitrate validity was very low at Kelapa Dua (61.67%) and moderately low at Manggarai (84.30%) and Pasar Baru (82.56%), indicating possible ion-selective electrode drift, environmental interference, or intermittent fouling. Low nitrate validity is often caused by long-term sensor drift, measurement surface fouling, and ionic cross-interference, especially in nutrient sensors used in dynamic water environments (Snyder et al., 2018; Daniel et al., 2020).

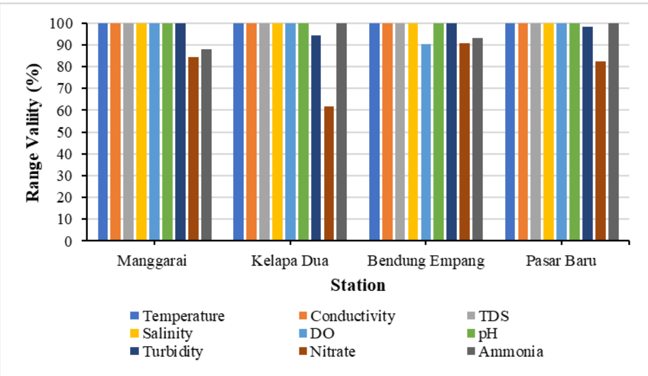

Figure 3. Validation range at four stations

Dissolved oxygen (DO) validity at the Empang Dam also decreased (90.40%), suggesting membrane fouling, disrupted optical paths, or unstable power supply (Samuelsson et al., 2018). These differences underscore the importance of the TDQM definition and analysis phase in ensuring that acceptable ranges are based on realistic environmental conditions and maintained through regular recalibration. Figure 3 displays the average validation ranges from the four stations.

3.2. Accuracy Evaluation

Accuracy values for physical parameters remained consistently high across all stations, approaching 100%. This confirms that raw measurements are generally consistent with reference values or expected ranges, and that the calibration procedures used at the installation remain effective. These findings align with U.S. EPA (2018) guidelines, which indicate that physical sensors typically exhibit low intrinsic error when supported by reliable power and continuous communication links. Inaccuracies can occur due to sensor calibration deviations, contamination, or configuration errors (Wei et al., 2019).

Chemical parameters showed notably more variability. The accuracy of nitrate in Kelapa Dua (61.67%) and ammonia in Manggarai (74.67%) may be due to measurement noise, signal drift, or occasional sensor interference (Luna et al., 2020). ORP accuracy was the lowest, especially in Manggarai (27.26%), suggesting possible biofilm buildup on the probe surface, which is known to affect oxidation-reduction potentials (Saboe, 2022). Within the TDQM framework, these deviations suggest that the Measure and analyze phases should include more frequent assessment cycles for high-risk parameters, especially those related to nutrients and the redox sensor (ORP). Figure 4 displays the average accuracy across the four stations.

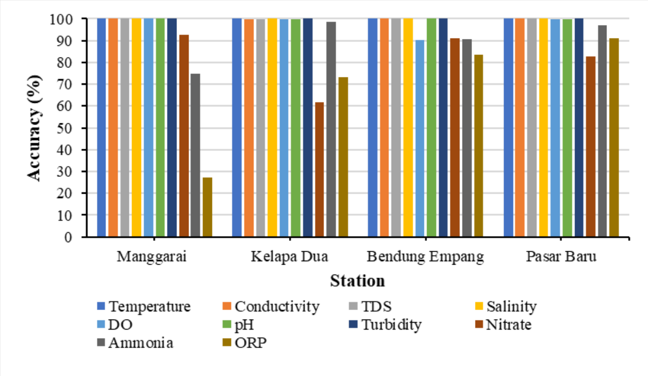

Figure 4. Average dimension accuracy at the four stations

Since manual data were unavailable for the 2016-2020 period, monitoring data from the Manggarai and Kelapa Dua stations for 2023-2024 were used and compared with manual measurement data. The manual data were obtained from the Ministry of Environment and Forestry's report, which included six sampling data points taken in December 2023, June 2024, July 2024, and August 2024. The following are the results of the analysis using the Mean Absolute Error (MAE) statistical approach for the Manggarai and Kelapa Dua stations. As shown in Tables 4 and 5.

MAE analysis at Manggarai Station revealed that the DO and ammonia parameters showed large and unstable deviations, along with extremely negative MAE, indicating sensor drift, power failure, or masking failure under poor conditions. This matches the findings of Lucas et al. (2025), who explained that chemical sensors in online monitoring systems are highly prone to fouling, membrane degradation, and fluctuations in river current, which can cause significant errors. In contrast, pH, temperature, and TDS exhibited high and consistent MAE (>88%), indicating better measurement stability, which agrees with the literature stating that physical sensors are more stable than electrochemical sensors (Karkouch et al., 2016). The TDS parameter experienced moderate fluctuations due to variability in sediment loads, which is common in urban river flows.

At Kelapa Dua Station, MAE instability was more noticeable, with almost all DO and ammonia parameters showing large negative MAEs. These could indicate periods of complete measurement failure or system shutdown, consistent with what is observed when dissolved oxygen and ammonium sensors completely fail or produce constant errors (Tena et al., 2020; Yang et al., 2021). Although pH and temperature remained relatively stable, a single instance of negative MAE values at both stations suggests a systemic failure affecting the entire sensor array. Overall, the MAE patterns at these two stations demonstrate that chemical sensor accuracy heavily relies on environmental stability, routine maintenance, and power conditions, as highlighted by Mahmud et al. (2020).

Table 4. MAE Analysis at Manggarai Station

Sample data date	Test	DO	pH	Ammonia	Temperature	TDS
28-Dec-23	MAE (%)	-17.31	97.56	-76.72	99.86	71.86
3-Jun-24	MAE (%)	-20.69	98.40	-61.81	91.97	46.96
June 13, 2024	MAE (%)	-65.81	89.18	94.29	94.89	91.46
30-Jul-24	MAE (%)	-75.40	97.90	-94.03	95.93	99.21
5-Aug-24	MAE (%)	68.61	98.58	-91.30	92.38	44.92
15-Aug-24	MAE (%)	32.84	98.87	-87.10	88.36	85.93

Table 5. MAE Analysis at Kelapa Dua Station

Sample Data Date	Test	DO	pH	Ammonia	Temperature	TDS
December 31, 2023	MAE (%)	-51.64	92.96	74.35	98.84	66.21
June 3, 2024	MAE (%)	-93.57	97.02	-93.06	92.83	95.69
June 11, 2024	MAE (%)	-92.58	90.00	17.65	98.58	60.52
5-Aug-2024	MAE (%)	-100.00	-100.00	-100.00	-100.00	-100.00
August 15, 2024	MAE (%)	-90.96	90.01	-91.90	82.49	46.23

3.3. Completeness and Zero-Value Analysis

The Completeness Index values, as shown in Figure 5, across all stations and parameters stayed at 100%, indicating no missing timestamps and dependable data transmission. From a TDQM perspective, this indicates effective data collection, communication, and storage, especially during the Measurement phase. However, the zero-completeness analysis tells a different story when evaluating the validity of the collected data. While all timestamps were recorded, some parameters exhibited a high rate of unreasonable zeros, which ODQEM classifies as a systematic anomaly.

The Empang Dam in Table 6 showed complete data with extremely low values for DO (1.02%) and nitrate (0.19%), while Pasar Baru experienced a sharp decline in TDS (9.57%), despite having perfect timestamp completeness. Such zero or constant readings are common indicators of sensor failure, including constant output mode or total failure, which is often linked to a dead sensor or malfunctioning probe (Yang et al., 2021). Kelapa Dua exhibited significant missing data for ammonia, with only 60.02% of values being non-zero, further highlighting the vulnerability of the ion sensors.

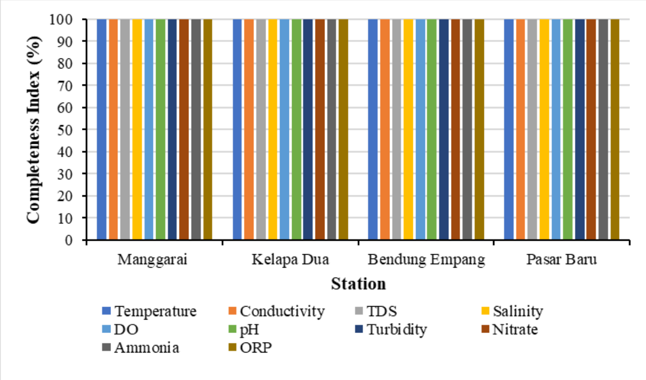

Figure 5. Completeness index at the four stations.

Table 6. Completeness with non-zero values at the four stations

Station	Temp.	EC	TDS	Sali-nity	DO	pH	Turbi-dity	Nitrate	NH_3	ORP
Manggarai	99.30	99.27	75.22	47.86	83.46	99.31	96.77	76.38	85.14	98.95
Kelapa Dua	99.14	98.83	89.27	42.82	83.50	99.16	79.31	92.34	60.02	99.14
Empang Dam	95.92	99.76	98.36	30.93	1.02	97.8	99.74	0.19	46.15	39.38
Pasar Baru	99.06	50.66	9.57	90.58	99.23	59.1	63.15	53.70	98.23	90.4
Average	**98.35**	**87.13**	**68.11**	**53.05**	**66.8**	**88.84**	**84.75**	**55.65**	**72.38**	**81.97**

The results show that although the time-stamp-based Completeness Index reached 100% at all stations, the proportion of non-zero values indicates that some parameters are only nominally complete and not analytically informative. The average non-zero completeness was very low for salinity (53.05%), nitrate (55.65%), DO (66.80%), and TDS (68.11%), with the most severe degradation observed at Dam Empang for DO (1.02%) and nitrate (0.19%) and at Pasar Baru for TDS (9.57%). These results imply that a dataset may appear structurally complete but still contain prolonged periods of inactivity and repeated zero values, which weaken the interpretation of river water dynamics. Therefore, non-zero completeness should be treated not merely as a supplementary descriptive metric, but as a key operational indicator for assessing whether the transmitted data remains meaningful for environmental analysis and decision-making support.

These findings demonstrate that completeness alone does not fully capture true data quality because incomplete data sets can significantly weaken the integrity of analytical results and the reliability of decision-making (Batini et al., 2009). Therefore, ensuring a high level of completeness is critical in supporting robust and actionable environmental intelligence (Pipino et al., 2002). TDQM analysis requires a more in-depth investigation beyond surface-level completeness, while the zero-valued ODQEM metric offers an important measure for identifying subtle sensor errors. This multi-layered assessment highlights that nominal completeness does not guarantee the completeness of information, emphasizing the need for a TDQM Improvement phase to address sensor maintenance, recalibration schedules, and system-level corrections.

3.4. Timestamp Uniqueness and Temporal Integrity

The uniqueness score exceeded 99.9% across all stations, as seen in Figure 6, indicating that the system does not produce duplicate timestamps and that the data stream maintains its temporal integrity. This demonstrates effective interaction between the sensor's Real-Time Clock (RTC), the gateway, and the server synchronization protocol. The high level of uniqueness meets the ODQEM requirements for structural data reliability and aligns with the ISO/IEC 30141 standard (IEC, 2024), which ensures that various IoT devices, systems, and platforms can connect, communicate, and exchange data.

In TDQM, temporal uniqueness relates to the Definition phase (where rules for timestamp integrity are established) and the Measurement phase (which tracks timestamp anomalies as a quality indicator). The observed stability demonstrates effective time synchronization governance and a low risk of redundant or duplicated data that could distort analysis results. Therefore, implementing systematic deduplication strategies, such as constraint-based validation, timestamp hashing, and version control, is crucial for maintaining high-quality data streams (Kaur et al., 2018).

3.5. Stability Index

At all four stations, as shown in Table 7, the Stability Index values show that Manggarai has strong stability in physical parameters, including temperature (89.81%), conductivity (86.53%), pH (91.69%), turbidity (91.24%), and ORP (92.44%). However, nutrient and ion parameters display different patterns. The stability of nitrate (65.97%) and ammonia (71.40%) indicates moderate temporal consistency, while TDS and salinity are very unstable (5.55% and 5.56%). The instability of these ion parameters may be caused by variable sedimentation, sudden changes in dissolved solids from rainwater runoff, or sensor problems such as scaling or membrane clogging.

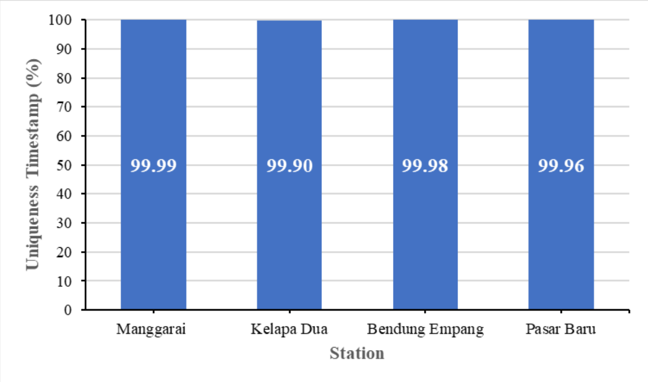

Figure 6. Average timestamp uniqueness across the four stations.

Kelapa Dua showed a strong stability pattern in physical parameters, especially conductivity (88.80%), DO (80.99%), pH (92.74%), turbidity (78.78%), and ORP (90.26%). Nitrate stability was relatively high (89.43%), but ammonia was volatile at 35.21%, possibly due to equipment malfunction or significant temporal variability in ammonia levels at this site. This aligns with urban runoff patterns, where ammonia spikes caused by domestic wastewater input can occur irregularly.

The Empang Dam showed the highest physical stability among the stations, with temperature at 95.23%, pH at 97.21%, turbidity at 93.09%, and ORP at 90.59%. DO stability was also quite high at 94.34%. However, nutrient stability remained low, with nitrate at 49.91% and ammonia at 41.78%. This pattern of high physical stability but low chemical stability matches previous findings in the Completeness and Validity of Range section, where the Empang Dam posed significant challenges for nutrient sensors. The high sediment load and periodic turbulence in this river segment probably worsen these challenges.

Pasar Baru showed the lowest overall stability, especially for conductivity (76.32%), turbidity (47.05%), nitrate (80.47%), and ammonia (30.20%). TDS (5.49%) and salinity (1.47%) also had very low stability, indicating significant environmental variability and possible sensor degradation. Despite these issues, the DO stability of 84.42% and pH stability of 89.01% suggest some core sensors remain operational.

The Stability Index results clearly show differences in sensor performance across stations, and provide a clearer basis for empirical prioritization across parameters and indicators. These results indicate that the most critical weaknesses were found in salinity (4.30%) and TDS (7.89%), followed by ammonia (44.65%), while pH (92.66%), ORP (90.06%), temperature (88.58%), and conductivity (85.02%) remained relatively stable across all stations. This emphasizing that physical parameters generally stay reliable. In contrast, chemical parameters tend to be unstable over time. This aligns with global patterns seen in WQMS systems, where ion sensors like nitrate and ammonia consistently perform no better than physical sensors (Rinn et al., 2025). During the TDQM cycle, such variability requires targeted improvements, especially recalibration, anti-fouling measures, and redundancy for critical parameters. Physical parameters generally remain more stable across stations, while chemical and ionic parameters are more susceptible to drift, fouling, scale formation, and flat-line behavior. Unstable variables dominated by zero values require more stringent recalibration intervals and targeted cleaning and anti-fouling measures. Meanwhile, variables with high validity and strong temporal stability can remain under routine quality control procedures. This analysis provides a more robust empirical foundation for the discussion by going beyond generalizations about sensor instability. The findings specifically identify parameters experiencing consistent failures and their impact on data reliability across multiple indicators.

Table 7. Average stability index at the four stations

Station	Temp.	EC	TDS	Sali-nity	DO	pH	Turbi-dity	Nitrate	NH_3	ORP
Manggarai	89.81	86.53	5.55	5.56	59.36	91.69	91.24	65.97	71.40	92.44
Kelapa Dua	88.38	88.80	9.61	6.62	80.99	92.74	78.78	89.43	35.21	90.26
Empang Dam	95.23	88.43	10.92	3.55	94.34	97.21	93.09	49.91	41.78	90.59
Pasar Baru	80.91	76.32	5.49	1.47	84.42	89.01	47.05	80.47	30.2	86.94
AVG	**88.58**	**85.02**	**7.89**	**4.3**	**79.78**	**92.66**	**77.54**	**71.44**	**44.65**	**90.06**

3.6. Integrated Interpretation of TDQM–ODQEM Results

The hybrid evaluation showed that while structural data quality indicators such as uniqueness and validity were generally strong, operational indicators, particularly stability and zero-value analysis, revealed chronic weaknesses in chemical sensor performance. Average non-zero completeness values were low for salinity (53.05%), nitrate (55.65%), DO (66.80%), and TDS (68.11%), while stability was very poor for salinity (4.30%), TDS (7.89%), and ammonia (44.65%). In other words, a dataset may appear good because its records are complete and well-organized, but it is generally unreliable for practical use if many readings are unstable or repeatedly recorded as zero. This finding demonstrates why data quality should be assessed not only by whether the data are present and well-structured, but also by whether the measurements are consistent and can be used to interpret actual river conditions and support timely decision-making. A further implication is that indicator integration should be interpreted hierarchically rather than descriptively. In this dataset, uniqueness is a fundamental structural requirement that has largely been met, while completeness and non-zero stability serve as decisive filtering criteria for operational utility. Range validity and accuracy remain important, but both should only be interpreted after ensuring that the time series is not dominated by persistent zeros, flat lines, or repetitive data. Practically, a parameter should be classified as high risk when it performs poorly in at least two operational dimensions, particularly completeness and non-zero stability. This explains why salinity, TDS, and ammonia require more urgent corrective action than temperature, pH, or ORP. Therefore, completeness and non-zero stability are treated as primary operational indicators, as they more directly determine whether the time series still contains meaningful environmental information over time. Validity and accuracy are then interpreted as supporting indicators, but only after the data has undergone operational filtering. Within this hierarchy, parameters exhibiting poor performance across more than one operational dimension are classified as higher-priority targets for corrective action, while parameters with strong structural quality and stable temporal behavior are considered more reliable for routine interpretation. This approach does not assign formal mathematical weights but rather provides a clear, practical prioritization framework for assessing the usefulness of data.

The integrated TDQM–ODQEM model suggests that improving ORWQM data quality requires not only sensor-level interventions but also institutional-level improvements in calibration governance, maintenance standardization, and interagency coordination. In Indonesia, water pollution control efforts are carried out through prevention, mitigation, and restoration of water quality, as stipulated in Government Regulation (GP) of the Republic of Indonesia No. 22 of 2021 concerning the implementation of environmental protection and management. GR Number 82 of 2001 concerning Water Quality Management and Water Pollution Control. This GR specifically regulates water quality management and water pollution control, including river water quality monitoring. Although it does not explicitly mention online river water quality monitoring, this regulation serves as the basis for developing monitoring systems, including technology-based systems that monitor water quality parameters in real time. Further institutional responsibilities are regulated by the Ministry of Environment and Forestry's organizational regulations, including Regulation of the Minister of Environment and Forestry No. 15 of 2021 concerning organizational structure and functions. These legal instruments imply that data quality assurance, validation procedures, and monitoring follow-up should not be treated as ad hoc technical tasks but as formal institutional responsibilities related to environmental monitoring, reporting, and control functions. Therefore, the governance recommendations proposed in this study, such as calibration standardization, maintenance scheduling, and cross-agency quality control, are based not only on observed sensor anomalies but also on the existing regulatory architecture of the Indonesian environmental monitoring system. Furthermore, the identified instability and frequent zero values should not only be interpreted as limitations requiring further research, but also as a basis for practical handling of existing datasets. A practical workflow is to first flag prolonged zero values and flatlines, second apply robust filtering to isolate short anomalous spikes, third incorporate unstable chemical parameters into daily or weekly summaries, and fourth retain only validated segments for interpretation against regulatory thresholds. This is particularly relevant for stations such as Empang Dam and Pasar Baru, where certain parameters exhibit significant degradation while others remain operationally informative. Thus, this study not only identifies quality issues but also highlights the potential for further analysis. Therefore, future research should explore multidimensional strategies for improving data quality, encompassing not only technological improvements but also institutional coordination, stakeholder engagement, socioeconomic factors, and environmental dynamics (Zhang & Thorburn, 2022).

4. Conclusion

This study approach uses an integrated framework of Total Data Quality Management (TDQM) and the Online Data Quality Evaluation Model (ODQEM) to identify priority issues for improvement to enhance the reliability of real-time monitoring data. The aims of this study were to evaluate the quality of data generated by four online river water quality monitoring (ORWQM) stations located in the downstream and upstream areas of the Ciliwung and Cisadane Rivers. Results indicate that the ORWQM system performs strongly on the structural data quality dimension, with validity, accuracy, and uniqueness generally ranging from 99% to 100%, and timestamp uniqueness consistently exceeding 99.9%. However, the evaluation also revealed persistent operational weaknesses. Completeness levels based on non-zero values varied significantly across parameters, with average scores of 68.11% for TDS, 53.05% for salinity, 66.80% for DO, and 55.65% for nitrate. The Stability Index indicates that physical parameters remain relatively stable, including temperature (88.58%), EC (85.02%), pH (92.66%), and ORP (90.06%), while chemical and ionic parameters are much less stable, particularly TDS (7.89%), salinity (4.30%), and ammonia (44.65%). These findings suggest that nominally complete and structurally valid data streams remain operationally unreliable for decision-making when impacted by flat data, anomalous zero values, or unstable chemical sensor behavior.

The novelty of this study lies in integrating TDQM with ODQEM, a strategic data governance framework with a real-time operational diagnostic framework, to evaluate data quality in Indonesia's ORWQM system. This integration is a substantive contribution because it goes beyond conventional assessments that focus solely on missing data or basic sensor outputs, and instead provides a multidimensional interpretation that links data structure, sensor behavior, and governance implications. Practically, this study provides a replicable evaluation model to identify which parameters can be managed institutionally and which require targeted technical interventions, particularly for ion sensors. More broadly, these findings suggest that the success of an ORWQM program should not be judged solely by the presence of real-time infrastructure or high-volume data streams, but rather by whether the resulting data are reliable enough to support pollution control, regulatory enforcement, and timely environmental responses. In this regard, this study has broader relevance for watershed management in Indonesia and other developing country contexts, where digital monitoring systems are evolving faster than the institutional capacity needed to ensure data quality.

At the same time, this study has several limitations: it only analyzed four stations, relied primarily on historical records from 2016–2020, and had limited laboratory reference data for direct accuracy validation, which was only available for specific parameters and periods at two stations in Manggarai and Kelapa Dua in 2023–2024. Therefore, future research should pursue three priority areas. First, we tested this framework across a range of stations, river types, and seasonal conditions to evaluate its robustness and applicability. Second, we integrated more systematic laboratory validation and maintenance logs to better distinguish between sensor drift, fouling, and true environmental variation. Third, we extended this framework to predictive and decision-support applications, including an anomaly alert system and a management dashboard that translates data quality diagnostics into operational actions. Overall, this study addresses a key analytical gap by demonstrating that the reliability of ORWQM depends not only on sensor placement but also on integrated data quality governance that supports transparent, evidence-based river pollution control. Without robust data quality governance, real-time monitoring remains merely an infrastructure promise, and with such governance, ORWQM can serve as a credible foundation for adaptive and accountable river management.

5. Data Availability Statement

The data used in this study are primary data (2016-2020), consisting of online water quality monitoring of the Ciliwung and Cisadane Rivers at four stations. The data is stored in the BPPT's Onlimo data center database. The data are available upon request and subject to approval by the relevant agencies.

6. Funding Institutions

This research was conducted without external funding. The authors confirm that no financial support was received from any funding agency for this research.

7. Conflict of Interest

Each author has declared that there is no conflict of interest in the writing or submission of this manuscript.

8. Acknowledgement

The authors would like to thank the Degree by Research (DBR) Program of the National Research and Innovation Agency (BRIN) for providing a scholarship for the Doctoral Program at the Graduate School of Sustainable Development (GSSD), University of Indonesia.

9. Author Contribution

SY conducted investigations, formal analysis, literature review, and manuscript preparation. **DNM**, **DS**, and **BK** were involved in the conceptualization and review of the manuscript. All authors read and approved the final manuscript.

References

Adu-Manu, K. S., Katsriku, F. A., Abdulai, J. D., & Engmann, F. (2020). Smart River Monitoring Using Wireless Sensor Networks. *Wireless Communications and Mobile Computing*, *2020*. https://doi.org/10.1155/2020/8897126

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3). https://doi.org/10.1145/1541880.1541883

Batini, C., & Scannapieco, M. (2016). *Data and Information Quality*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24106-7

Behmel, S., Damour, M., Ludwig, R., & Rodríguez, M. J. (2016). Water quality monitoring strategies—A review and future perspectives. *Science of the Total Environment*, *571*, 1312–1329. http://dx.doi.org/10.1016/j.scitotenv.2016.06.235

Bowo, W. A., Ma'mun, S., Suhanto, A., Hidayanto, A. N., Naisuty, M., & Habsari, I. C. (2019). *Data Quality Assessment: A Case Study of PT JAS Using TDQM Framework*. https://doi.org/doi:10.1109/ICIC47613.2019.8985896.

Bushnell, M., Toll, R., & Worthington, H. (2015). Manual for real-time quality control of dissolved oxygen observations: a guide to quality control and quality assurance for dissolved oxygen observations in coastal oceans. https://doi.org/10.7289/V5ZW1J4J

Bushnell, M., & Worthington, H. (2016). Manual for real-time quality control of in-situ temperature and salinity data: a guide to quality control and quality assurance for in-situ temperature and salinity observations. https://doi.org/10.7289/V5V40SD4

BPPT (2015). Final Report: Installation of Online System Monitoring Devices at Ciliwung (2 Points) and Citarum (1 Point). Technology Service Center / BPPT Engineering, Agency for the Assessment and Application of Technology, Jakarta.

Cahyono, S. H., Sucahyo, Y. G., Raya, J. S., & Pusat, J. (2020). Pengukuran Kualitas Data Menggunakan Framework Total Data Quality Management (TDQM): Studi Kasus Sistem Informasi Beasiswa Universitas Indonesia. *Jurnal Ilmu Pengetahuan Dan Teknologi Komunikasi*, *22*(2), 193–206. https://doi.org/10.33164/iptekkom.22.2.2020.193-206

Chapman, D. V. (1996). *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring* (No. 17872). E & FN Spon. https://iris.who.int/bitstream/handle/10665/41850/0419216006_eng.pdf

Daniel, A., Laës-Huon, A., Barus, C., Beaton, A. D., Blandfort, D., Guigues, N., ... & Achterberg, E. P. (2020). Toward a harmonization for using in situ nutrient sensors in the marine environment. *Frontiers in Marine Science*, *6*, 773. https://doi.org/10.3389/fmars.2019.00773

Ehrlinger, L., & Wöß, W. (2022). A Survey of Data Quality Measurement and Monitoring Tools. In *Frontiers in Big Data* (Vol. 5). Frontiers Media S.A. https://doi.org/10.3389/fdata.2022.850611

IEC. (2024). ISO/IEC 30141:2024. *Internet of things (IoT) – Reference architecture*. https://standards.iteh.ai/catalog/standards/iso/d571f944-575e-49b2-a919-39a721ff7209/iso-iec-30141-2024

Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. In *Journal of Network and Computer Applications* (Vol. 73, pp. 57–81). Academic Press. https://doi.org/10.1016/j.jnca.2016.08.002

Kaur, R., Chana, I., & Bhattacharya, J. (2018). Data deduplication techniques for efficient cloud storage management: a systematic review. *Journal of Supercomputing*, *74*(5), 2035–2085. https://doi.org/10.1007/s11227-017-2210-8

KLHK. (2024). *Laporan Kinerja Direktorat Pengendalian Pencemaran Air*. https://www.scribd.com/document/891351611/Ditjen-PPKL-2024-L-Rev1-compressed

Klein, A. and Lehner, W. (2009). Representing data quality in sensor data streaming environments. ACMJ. Data Inform. Quality 1, 2, Article 10 (September 2009), 28 pages. DOI =10.1145/1577840.1577845.

Li, X. T., Zhai, J., Zheng, G. F., & Yuan, C. F. (2018). Quality assessment for open government data in China. In *Proceedings of the 2018 10th International Conference on Information Management and Engineering* (pp. 110–114). https://doi.org/10.1145/3285957.3285962

Lin, L., Yang, H. & Xu, X. (2022). Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. Front. Environ. Sci. 10:880246. https://doi.org/10.3389/fenvs.2022.880246

Lucas, L. V., Brown, C. J., Robertson, D. M., Baker, N. T., Johnson, Z. C., Green, C. T., Cho, S. J., Erickson, M. L., Gellis, A. C., Jasmann, J. R., Knowles, N., Prein, A. F., & Stackelberg, P. E. (2025). Gaps in Water Quality Modeling of Hydrologic Systems. *Water*, *17*(8), 1200. https://doi.org/10.3390/w17081200

Luna Juncal, M. J., Skinner, T., Bertone, E., & Stewart, R. A. (2020). Development of a real-time, mobile nitrate monitoring station for high-frequency data collection. *Sustainability*, *12*(14), 5780. https://doi.org/10.3390/su12145780

Mahmud, M. P., Ejeian, F., Azadi, S., Myers, M., Pejcic, B., Abbassi, R., ... & Asadnia, M. (2020). Recent progress in sensing nitrate, nitrite, phosphate, and ammonium in aquatic environment. *Chemosphere*, *259*, 127492.https://doi.org/10.1016/j.chemosphere.2020.127492

Mansouri, T., Reza, M., Moghadam, S., & Zareravasan, A. (2023). *IoT Data Quality Issues and Potential Solutions: A Literature Review*. *The Computer Journal*, *66*(3), 615–625. DOI: 10.1093/comjnl/bxab183

Meyer, A. M., Klein, C., Fünfrocken, E., Kautenburger, R., & Beck, H. P. (2019). Real-time monitoring of water quality to identify pollution pathways in small and middle scale rivers. *Science of the Total Environment*, *651*, 2323–2333. https://doi.org/10.1016/j.scitotenv.2018.10.069

Mohamed, A., Nik, A., Kadir, A., Lin, M.-, Rahman, S. A., & Arshad, N. H. (2009). Data completeness analysis in the Malaysian Educational Management Information System. In *International Journal of Education and Development using Information and Communication Technology* (Vol. 5, Issue 2). https://www.learntechlib.org/p/188080/

Pezoulas, V. C., Kourou, K. D., Kalatzis, F., Exarchos, T. P., Venetsanopoulou, A., Zampeli, E., Gandolfo, S., Skopouli, F., De Vita, S., Tzioufas, A. G., & Fotiadis, D. I. (2019). Medical data quality assessment: On the development of an automated framework for medical data curation. *Computers in Biology and Medicine*, *107*, 270–283. https://doi.org/10.1016/j.compbiomed.2019.03.001

Pipino, L. L., Lee, Y. W., Wang, R. Y., & Yang, R. Y. (2002). *Data Quality Assessment* (*Communications of the ACM*, Vol. 45, Issue 4 pages 211-218). https://doi.org/10.1145/505248.506010

Qartod.(2018). Manual for Real-Time Quality Control of Stream Flow Observations, U.S. Integrated Ocean Observing System (IOOS). DOI : https://doi.org/10.25923/gszc-ha43

Qin, Y., Sheng, Q. Z., Falkner, N. J. G., Dustdar, S., Wang, H., & Vasilakos, A. V. (2016). When things matter: A survey on data-centric internet of things. *Journal of Network and Computer Applications*, *64*, 137–153. https://doi.org/10.1016/j.jnca.2015.12.016

Ramirez, G., Fuentes, O., & Tweedie, C. E. (2011, March). Assessing data quality in a sensor network for environmental monitoring. In *2011 Annual Meeting of the North American Fuzzy Information Processing Society* (pp. 1-6). IEEE. DOI: 10.1109/NAFIPS.2011.5752010

Redman, T. C. (2005). *Measuring data accuracy a framework and review*. https://www.researchgate.net/publication/292332376_Measuring_data_accuracy_A_framework_and_review

Rinn, P., Boruta, F., Czermak, P., & Ebrahimi, M. (2025). Fouling Control of Ion-Selective Electrodes (ISEs) in Aquatic and Aquacultural Environments: A Comprehensive Review. *Sensors*, *25*(24), 7515. https://doi.org/10.3390/s25247515

Saboe, D. (2022). *Monitoring Key Water Quality Parameters in Water Resources Systems Using Bioactive Electrodes* (Doctoral dissertation, Arizona State University). https://www.proquest.com/openview/c74b150c093c032a1ca5408edfc69f45/1?pqorigsite=gscholar&cbl=18750&diss=y

Samuelsson, O., Björk, A., Zambrano, J., & Carlsson, B. (2018). Fault signatures and bias progression in dissolved oxygen sensors. *Water Science and Technology*, *78*(5), 1034–1044. https://doi.org/10.2166/wst.2018.350

Soares, A. L. C., Pinto, C. C., & Oliveira, S. C. (2020). Impacts of anthropogenic activities and calculation of the relative risk of violating surface water quality standards established by environmental legislation: a case study from the Piracicaba and Paraopeba river basins, Brazil. *Environmental Science and Pollution Research*, *27*(12), 14085–14099. https://doi.org/10.1007/s11356-020-07647-1

Sulthonuddin, I., Hartono, D. M., Abidin Said, C. A., & Utomo, S. W. (2019). River Water Pollution Indication in the Cimanuk River Downstream, Indramayu District. *IOP Conference Series: Earth and Environmental Science*, *239*(1). https://doi.org/10.1088/1755-1315/239/1/012010

Snyder, L., Potter, J. D., & McDowell, W. H. (2018). An evaluation of nitrate, fDOM, and turbidity sensors in New Hampshire streams. *Water Resources Research*, *54*(3), 2466–2479. https://doi.org/10.1002/2017WR020678Digital Object Identifier Talagala,

Teh, H. Y., Kempa-Liehr, A. W., & Wang, K. I. K. (2020). Sensor data quality: a systematic review. In *Journal of Big Data* (Vol. 7, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40537-020-0285-1

Tena, D., Penarrocha-Alós, I., Sanchis, R., & Moliner-Heredia, R. (2020). Ammonium Sensor Fault Detection in Wastewater Treatment Plants. In *ICINCO* (pp. 681-688). DOI: 10.5220/0009875406810688

U.S. EPA (2018). *Online Water Quality Monitoring in Distribution Systems*. https://share.google/SzANXckRwU2f6KrDq

Wang, R. Y. (1998). A Product Perspective on Total Data. In *COMMUNICATIONS OF THE ACM* (Vol. 41, Issue 2). https://doi.org/10.1145/269012.269022

Wang Richard Y & Strong Diane M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–34. https://doi.org/10.1080/07421222.1996.11518099

Wei, Y., Jiao, Y., An, D., Li, D., Li, W., & Wei, Q. (2019). Review of dissolved oxygen detection technology: From laboratory analysis to online intelligent detection. In *Sensors (Switzerland)* (Vol. 19, Issue 18). MDPI AG. https://doi.org/10.3390/s19183995

Wijayanti, W., Hidayanto, A. N., Wilantika, N., Adawati, I. R., & Yudhoatmojo, S. B. (2018). *Data Quality Assessment on Higher Education: A Case Study of Institute of Statistics*. DOI: 10.1109/ISRITI.2018.8864476

Yang, P., Li, Z., Yu, Y., Shi, J., & Sun, M. (2021). Studies on fault diagnosis of dissolved oxygen sensor based on GA-SVM. *Mathematical Biosciences and Engineering*, *18*(1), 386–399. doi: 10.3934/mbe.2021021

Zhang, L., Jeong, D., & Lee, S. (2021). Data quality management in the internet of things. In *Sensors* (Vol. 21, Issue 17). MDPI AG. https://doi.org/10.3390/s21175834

Zhang, Y., & Thorburn, P. J. (2022). Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. *Future Generation Computer Systems*, *128*, 63–72. https://doi.org/10.1016/j.future.2021.09.033

Zhang, X., Ma, L., Zhu, Y., Lou, W., Xie, B., Sheng, L., ... & Gu, Q. (2022). Temporal stability analysis for the evaluation of spatial and temporal patterns of surface water quality. *Water Resources Management*, *36*(4), 1413–1429. https://doi.org/10.1007/s11269-022-03090-8