# Data Quality Assessment: A Practical Application

Eliana Costa e Silva(✉), Teresa Peixoto, Óscar Oliveira, and Bruno Oliveira

CIICESI, Escola Superior de Tecnologia e Gestão, Instituto Politécnico do Porto, Porto, Portugal
{eos,tmop,oao,bmo}@estg.ipp.pt

**Abstract.** This paper presents a novel data quality score designed to address the challenges of ensuring high-quality data in Internet of Things (IoT) deployments. Given the growing reliance on IoT systems and the volume of data they generate, maintaining data quality is essential for reliable decision-making and effective analytics. The proposed score synthesizes key data quality dimensions, providing a comprehensive measure of data quality that can be applied across various IoT contexts. The results obtained for a public dataset on a water pumping system show the applicability and flexibility of the proposed data quality score. This work contributes to the ongoing efforts to improve data management in IoT environments, ultimately supporting the development of robust, data-driven solutions.

**Keywords:** Data Quality Metrics · Data Quality Indices · Quality Measurement · IoT

## 1 Introduction

In today's data-driven world, the importance of high-quality data cannot be overstated, especially within the context of the Internet of Things (IoT), where data is the heart of intelligent decision-making [3]. Data quality directly influences the reliability of insights derived from it, impacting business operations, system performance and the effectiveness of real-time applications [3]. However, IoT data is frequently compromised by factors such as sensor inaccuracies, network latency, and heterogeneous data sources [10]. This highlights the need for effective methods of measuring, assessing, and managing data quality, and also emphasises the need for an aggregated metric of overall data quality [3]. Data quality dimensions, which are attributes or characteristics that define data's fitness for use, are used to quantify and manage data quality [7]. While a large number of data quality dimensions have been described in the literature [5], some of the most pertinent for IoT systems are generally considered to be accuracy, completeness, timeliness, and consistency [11]. Although these dimensions are individually important, their interdependencies and trade-offs highlight the


J. Machado et al. (Eds.): icieng 2025, LNME, pp. 512–523, 2025.
https://doi.org/10.1007/978-3-031-94484-0_42

need for a unified approach to data quality assessment [8]. For example, optimising for accuracy may impact timeliness. Given this, there is a need for an aggregated measure of data quality that considers all the key dimensions in an appropriate and balanced way. Such a metric would assist in making data quality management more tangible, comparable, and manageable [7].

The study in [13] examines data quality in an extrusion industrial setting, developing quality dimensions from a decision-maker's perspective. It emphasizes aligning data with expected utility and relevance for decision-making. Expanding on this work, the present study generalizes these metrics for broader industrial applications, reducing reliance on subjective data interpretation. By using well-defined data quality metrics, this approach provides clear, actionable insights, enhancing data management and decision-making across diverse industrial environments. The present paper presents a data quality score that integrates key dimensions of data quality, offering a comprehensive and actionable measure to evaluate data quality in IoT deployments. The proposed score contributes to the ongoing effort to improve data quality in IoT systems, fostering the development of robust, reliable, and trustworthy data-driven solutions.

The remainder of this paper is structured as follows: Sect. 2 provides a review of the existing literature on data quality. Section 3 introduces the case study in which the data quality assessment metrics will be applied. Section 4 introduces the data quality metrics employed, followed by the results obtained for the case study and the analyses of the effectiveness and limitations of the proposed metrics. Finally, Sect. 5 summarizes the key findings and proposes directions for future research.

## 2 Related Work

The concept of data quality is central to the effective use of data in diverse fields. In essence, data quality refers to the fitness of data for its intended use [1]. This multi-dimensional construct goes beyond the absence of errors and incorporates numerous factors that impact the value and reliability of data. Data quality dimensions are attributes or characteristics of the data that define its suitability for use [7]. They are used to quantify, assess and manage the quality of data [7]. For many IoT applications, the most relevant data quality dimensions-accuracy, completeness, timeliness, and consistency-are generally considered interdependent, with potential overlaps between them [8,9].

**Accuracy** refers to how well data values represent real-world entities, measured by their agreement with a reliable source [7][?]. It indicates closeness to the true value [7] and adherence to the most plausible corresponding value [9]. Accuracy comprises structural accuracy-covering syntactic and semantic aspects-and time-related accuracy, which includes currency, volatility, and timeliness [9]. Syntactic accuracy concerns conformity to a definition domain, while semantic accuracy ensures adherence to natural language rules [9]. Sources of correct information vary and may include databases of record, corroborative datasets, computed values, or manual processes, though no definitive source may exist [4]. In IoT, accuracy is critical, as sensor data informs real-world decision-making.

**Completeness** measures whether all necessary information is present, ensuring no missing values [7]. It reflects the sufficiency of data in breadth, depth, and scope for a given task [5] and the ability of a system to represent all meaningful real-world states [9]. Essentially, completeness requires certain attributes to have values [4], either independently or based on other attributes [4]. Types of completeness include schema, column, population, linkability, and property completeness [9]. It can be quantified as the ratio of filled attributes to the total attributes [9]. In IoT, completeness is crucial, as missing data can lead to inaccurate analysis and decisions.

**Timeliness** refers to how up-to-date data is and its availability when needed [5]. It relates to data currency and recency [7] and is defined by the delay between a real-world change and its update in the information system [9]. Timeliness is quantified from 0 (poor) to 1 (optimal) and depends on data volatility-highly volatile data requires greater currency, while low volatility data is less affected [9]. In IoT, timeliness is critical for real-time applications that rely on up-to-date data to react to environmental changes. However, it may be compromised when prioritizing accuracy or completeness.

**Consistency** ensures data is uniformly formatted and compatible across sources and systems [5]. It synchronizes data definitions, structures, meanings, and presentation across business lines [4] and reflects equivalence across tables, sources, or systems [7]. Inconsistencies may result from varied formats or differing data entry interpretations. Types of consistency include semantic and structural consistency [7], presentation completeness, and temporal consistency [4]. Semantic consistency ensures similar data objects have uniform names and meanings, while structural consistency maintains uniform representation of attribute values [7]. In IoT, consistency is crucial as data from heterogeneous sources must be harmonized to prevent misinterpretation and inaccurate usage [7].

These four dimensions of accuracy, completeness, timeliness, and consistency are critical for data quality in IoT systems. These data quality dimensions also have trade-offs, for instance, optimising for accuracy might affect timeliness [9]. Therefore, it is necessary to continuously measure and manage data quality to ensure that data meet the requirements of business needs.

Despite the importance of data quality, several challenges arise in its assessment. The heterogeneity of data sources in IoT environments is a major obstacle, as data streams from various sensors, devices, and networks each have unique characteristics. This diversity introduces inconsistencies and complexities in data processing and evaluation [1]. Additionally, sensor inaccuracies, such as electromagnetic interference, signal processing errors, and packet loss, compromise data reliability. Data may also be affected by noise, outliers, and missing values. The high volume and velocity of IoT data further complicate real-time quality monitoring and management [2]. Furthermore, data may experience drift, discontinuity, and imprecision over time. Finally, concerns about data security and privacy also impact data quality [1]. Various frameworks, methodologies, and techniques have been developed to evaluate data quality, typically focusing on defining relevant data quality dimensions, creating metrics to measure them, and establishing processes for improvement [3]. Data profiling is commonly used to analyze datasets and identify inconsistencies or anomalies. Other methods involve statistical techniques, machine learning algorithms, and data mining to detect data quality issues [4]. Additionally, formal data quality management frameworks, such as the POSMAD (Planning, Obtaining, Storing, Sharing, Maintaining, Applying, and Disposing of Data) approach, support decision-making by considering the data life cycle. Some frameworks adopt a strategic approach to system complexity, emphasizing dependencies [5]. Data quality assessment is also guided by quality standards, with the ISO 8000 standard being frequently referenced [1].

While existing literature provides a comprehensive understanding of data quality, several gaps remain. Many studies focus on structured or semi-structured data, with less attention given to the unstructured data generated in the IoT [8]. Moreover, there is a need for more robust and scalable data quality assessment methods that can effectively handle the volume, velocity, and variety of IoT data streams [2]. Specifically, few studies have addressed data quality management in edge-based architecture levels for IoT. Furthermore, many existing data quality frameworks lack practical implementation details for IoT environments [1]. The study in [2] analyzes data quality in IoT, emphasizing its dimensions, relationships, and improvement strategies. It explores maintaining data quality during acquisition and transformation stages and examines data cleaning techniques to address uncertainty and ensure purified data for IoT. In [14], key data quality challenges in smart manufacturing are identified, with Accuracy, Completeness, Consistency, and Timeliness highlighted as the most critical factors for reliable data analysis. The study underscores the need for a structured approach to data quality assessment, particularly in data-intensive environments such as IoT and industrial systems.

## 3 Case Study

To demonstrate the applicability of the data quality scores proposed here, a case study from a field distinct from the extrusion scenario in [13], where the metrics were originally developed, was selected in the present work. The current case study presents a range of differing characteristics, including a higher number of sensors, enabling us to highlight the generality of the proposed scores and their adaptability to various environments and use cases.

Water pumping systems play a crucial role in supplying water to communities, industries, and agricultural areas [12], ensuring a continuous and reliable supply. However, in remote regions far from major urban centers, infrastructure often lacks proper maintenance, resulting in frequent failures that directly impact local populations and economies. This study analyzes a real dataset[1] from a water pumping system in a small community facing recurrent failures due to technical and structural issues, leading to water supply interruptions that impact residents' daily lives and the region's agricultural and industrial

[1] https://www.kaggle.com/datasets/nphantawee/pump-sensor-data.

productivity. In some cases, water unavailability creates critical situations for families relying on regular supplies. The dataset includes measurements from 52 sensors monitoring variables such as pressure, speed, temperature, and humidity from April to August 2018 [12]. These sensors generate large amounts of raw data every minute, including redundant measurements for enhanced security and reliability. Additionally, the dataset records the pump's operating status, enabling precise identification of failure moments. For this analysis, only the data from April 2018 (43200 records) were considered. Figure 1 shows the operation of four sensors during this month[2], illustrating the measurement distribution and redundant data. In April, two system failures occurred: the first from 00:30 on the 12th to 13:39 on the 13th, and the second from 00:30 on the 18th to 4:20 on the 20th. In Fig. 1, these failures are marked with red lines, followed by a recovery period highlighted in yellow.

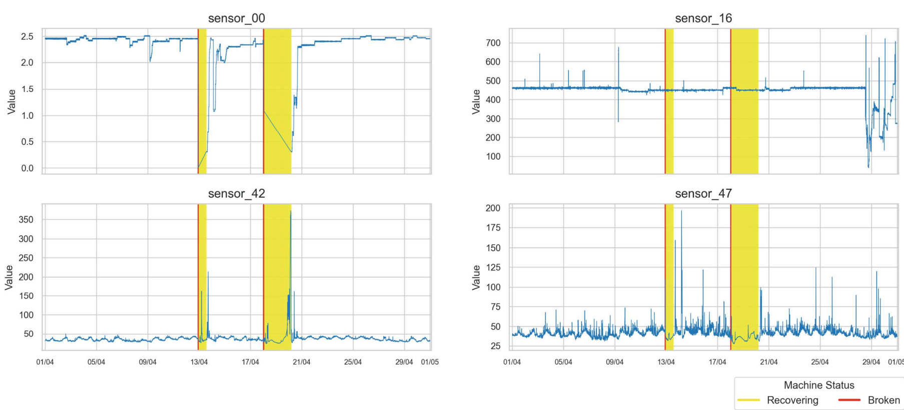


**Fig. 1.** Raw data on four of the 52 sensors, throughout April 2018.

## 4 Data Quality Assesment

This section begins with an overview of the quality dimensions used in the proposed data quality score. Subsection 4.1 details the computation of each dimension, while Subsect. 4.2 introduces the data quality score, formulated as a linear combination of these dimensions, and is designed to address the challenges of ensuring high-quality data in IoT deployments. Subsection 4.3 presents the case study results.

[2] All sensors plots at https://github.com/TeresaPeixoto/Data-Quality-Assessment.

### 4.1 Computation of Each Quality Dimension

For the analysis of each data quality dimension, specific metrics were selected based on [14]. These metrics are calculated in 5-minute blocks, enabling continuous assessment of data quality over time. Some metrics rely exclusively on data within each block, while others require historical data for calculation. Historical data is crucial when considering prior correlations or reference values to ensure a more robust evaluation. Additionally, each metric can be calculated either for the entire block, evaluating all records simultaneously, or individually by column, providing a more detailed analysis of each variable.

The accuracy dimension is evaluated for each observation $x$ based on the M4 metric, proposed in [2]. This metric normalizes values within a range of 0 to 1, where values close to 0 indicate the observation is near the minimum value, while values close to 1 indicate it is near the maximum value. The equation for the accuracy metric is defined in Eq. (1) where $X$ is the set of all observations up to observation $x$. The accuracy for each data block $j$ is computed as the average of the accuracy of all observations on that block.

$$Accuracy(x) = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{1}$$

To mitigate the impact of outliers, the metric proposed in [2] is extended by defining $X$ as the set of acceptable values, where the lower and upper limits are established using the 10th and 90th percentiles of the historical data up to the current block $j$. Figure 2 shows, for sensor_00, the variation of the minimum and maximum values for each block $j$ (green and red dashed lines), as well as the 10th and 90th percentiles (blue and orange lines). The minimum and maximum values fluctuate strongly over time, risking distortion if used directly for accuracy calculation. Percentiles provide a more stable alternative, less affected by outliers, ensuring consistent normalization bounds and capturing real data trends, leading to a more reliable accuracy assessment. When the accuracy of a given data block $j$ falls within the [0, 1] range, it can be inferred that the data follows an acceptable pattern. However, if the accuracy falls outside this range, it indicates imprecision in the data or potential measurement errors.

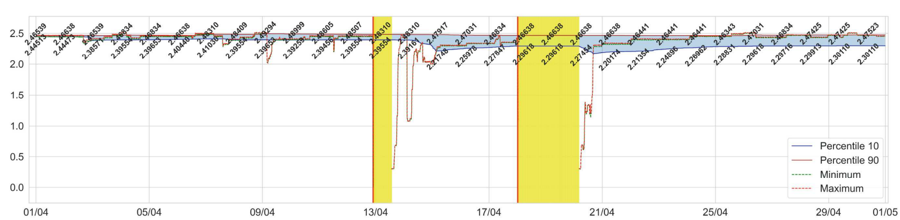


**Fig. 2.** Variation of the minimum, maximum, 10th, and 90th percentiles for sensor_00.

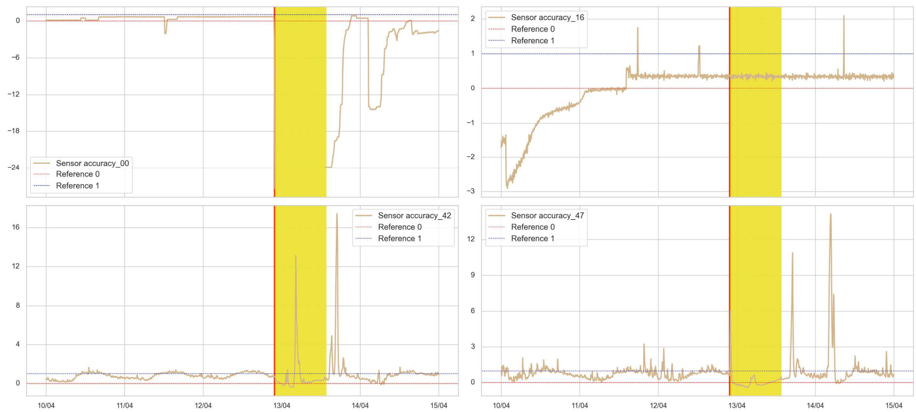


**Fig. 3.** Accuracy results on four of the 52 sensors, from April 10th to 15th.

The results shown in Fig. 3 indicate that in some cases, the accuracy is outside the expected range. For clarity, the analysis focuses on the period from April 10th to 15th, which includes the first observed system failure on the 12th. Although the accuracy generally remains within the [0, 1] range, which is ideal, several points stand out and warrant further attention. For example, sensor_00 shows a sharp drop at the end of the 12th, with accuracy below -24. Sensor_16 has variations above 1 between the 11th and 15th, and negative fluctuations between the 10th and 11th. Sensor_42 and sensor_47 show values above 13 from the 13th to 15th, with sensor_47 showing the most variation. These anomalies may indicate unexpected events.

The completeness dimension is assessed using a common metric, with variations in terminology across studies, such as [2] and [7]. The completeness of data block $j$ is calculated using Eq. (2), where $N_{miss}$ represents missing values (e.g., nulls or spaces), and $N_{total}$ is the expected total of filled data in block $j$. This metric can be applied at both the record (i.e., observations) and attribute (i.e., sensors) levels, enabling the identification of data completeness gaps across different layers. This approach helps uncover the underlying causes of completeness issues, thereby improving overall data quality.

$$Completeness_j = \frac{N_{total} - N_{miss}}{N_{total}} \tag{2}$$

The results applied at the attribute level are shown in Fig. 4. Sensor_00 shows a completeness value of 0 between the blocks on April 12th at 22:00 and April 13th until 13:10. This period aligns with one of the times when the machine was in "broken" and "recovering" states. No issues were observed with the other sensors during this interval[3]. Overall, completeness remains at 1, as expected.

[3] However, as shown in the plots available at https://github.com/TeresaPeixoto/Data-Quality-Assessment, other sensors experience failures at different times.

At the record level (Fig. 5), a significant drop in completeness is seen on April 27th, indicating a global failure. Completeness never reaches 1 due to sensor_15 not transmitting data. Smaller fluctuations in mid-April suggest brief data loss without major impact on overall quality.

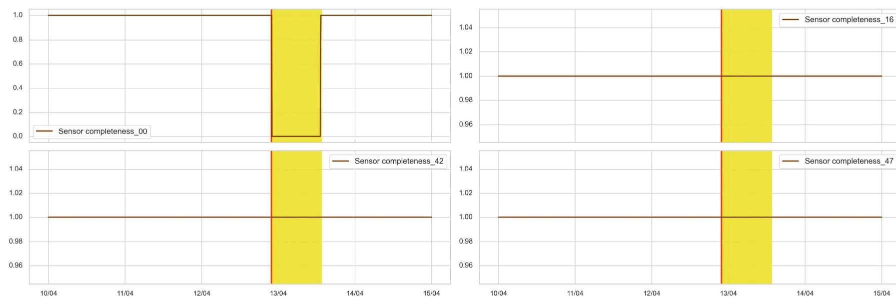


**Fig. 4.** Completeness results on four of the 52 sensors, from April 12th to 15th.

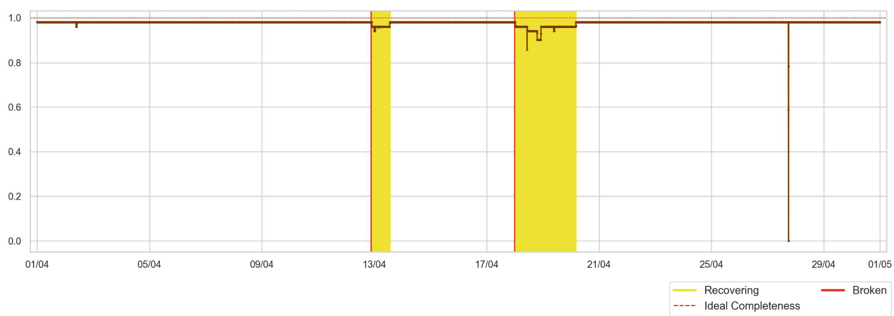


**Fig. 5.** Completeness results at record level, throughout April 2018.

In this study, accuracy and completeness were prioritized over consistency and timeliness due to their direct impact on the reliability of data-driven decisions in streaming scenarios. Accuracy was selected since sensor data is inherently prone to noise and measurement errors, which can significantly affect the detection of anomalies and failures in the water pumping system. Completeness was chosen since missing values compromise the ability to generate continuous and reliable insights, which is critical for real-time monitoring and system performance assessment. Consistency was not included as a primary dimension since we assume that the dataset originates from a single sensor network, where structural and semantic inconsistencies are less prevalent. Since the data is collected in a controlled environment, discrepancies between different sources or formats are minimal, reducing the need for consistency as a key assessment metric. Timeliness, while important in some streaming applications, was deprioritized in this study because the focus was on evaluating the intrinsic quality of the data rather than the delay in its arrival. Given that the data ingestion process does not introduce significant latencies and that analysis is performed over short time windows, ensuring the correctness and completeness of the data was deemed more impactful for assessing overall data quality.

### 4.2 Proposed Quality Scores

In this section, a Quality Score Delta ($QSD$) is introduced, integrating both accuracy and completeness dimensions. This metric, inspired by previous work [13], reflects the dynamic nature of data quality trends, supporting informed decision-making for continuous data management refinement. In [13], data from an extrusion industrial scenario was analyzed, with metrics designed from a decision-maker's perspective, focusing on the data's utility and relevance. This study aims to generalize these metrics, extending their applicability beyond extrusion to broader industrial contexts. By defining clear data quality metrics, the decision-making process is simplified, removing the need for subjective interpretation of data meaning. For each data block $j$, $QSD_j$ is calculated as shown in Eq. (3), where $WQS_j$ and $LWQS_j$ represent the Weighted Quality Score and Longitudinal WQS, respectively. These two quality scores integrate accuracy and completeness, with weights $w_a$ and $w_c$, respectively, and that are such that $w_a + w_c = 1$ and $w_a, w_c \geq 0$. In each dimension, the individual information on each sensor is also weighted.

$$QSD_j = \begin{cases} WQS_j - LWQS_j & \text{if } j > 1 \\ 1 & \text{if } j = 1 \end{cases} \tag{3}$$

While $WQS_j$ provides a snapshot evaluation of the quality of data at a particular block $j$, $LWQS_j$ is tailored for historical data analysis as a metric that scores quality over time while giving more relevance to newer data compared to older one. The WQS and LWQS are calculated as shown in Eqs. (4) and (5), respectively, where $s \in S = \{1, ..., n_s\}$, $n_j$ is the number of row of block $j$, $a_j^s$ is the number of rows with accuracy of the data sensor $s$ between the allowed range, $c_j^s$ is the number of rows with no missing values, $w_s$ are the weights of each sensor $s$ and are such that $w_1 + ... + w_{n_s} = 1$ and $w_1, ..., w_{n_s} \geq 0$, $K = \{j - m, ..., j - 1\}$ is the set of integers containing indexes of the last $m$ data blocks before block $j$ and $f_k$ is a function of $k \in K$ such that the weight attributed to the $k^{th}$ data block before $j$ decreases as $k$ is further away from $j$, and is defined as $f_k \equiv f(k) = \exp\left(-\frac{j-k-1}{\beta}\right)$, where $\beta > 0$ and $m > 0$ control the decay rate of the weight attributed to the $k^{th}$ block before $j$ and the temporal range, respectively.

$$WQS_j = w_a \left( \sum_{s \in S} w_s \frac{a_j^s}{n_j} \right) + w_c \left( \sum_{s \in S} w_s \frac{c_j^s}{n_j} \right) \tag{4}$$

$$LWQS_j = w_a \left[ \sum_{s \in S} w_s \left( \frac{\sum_{k \in K} f_k \, a_k^s}{\sum_{k \in K} f_k \, n_k} \right) \right] + w_c \left[ \sum_{s \in S} w_s \left( \frac{\sum_{k \in K} f_k \, c_k^s}{\sum_{k \in K} f_k \, n_k} \right) \right] \tag{5}$$

Alternatively, WQS and LWQS can also be computed by first aggregating the accuracy and completeness of all the sensors, as shown in Eqs. (6) and (7), respectively, where $N_j = n_s \times n_j$, $A_j = a_j^1 + ... + a_j^{n_s}$ and $C_j = c_j^1 + ... + c_j^{n_s}$.

$$WQS_j = \frac{1}{N_j}\left(w_a A_j + w_c C_j\right) \tag{6}$$

$$LWQS_j = w_a \left(\frac{\sum_{k \in K} f_k\, A_k^r}{\sum_{k \in K} f_k\, N_k}\right) + w_c \left(\frac{\sum_{k \in K} f_k\, C_k^r}{\sum_{k \in K} f_k\, N_k}\right) \tag{7}$$

### 4.3 Results

In this study, the decision-maker assigns weights to the dimensions (accuracy and completeness) and to each sensor based on its importance. The weights used are $w_a = 0.4$ for accuracy and $w_c = 0.6$ for completeness, emphasizing the greater impact of completeness on data quality. Data is processed in 5-minute blocks for the 52 sensors. The LWQS is calculated using the previous 12 blocks (1 h), and sensor weights are set to $w_s = 0.019$ for sensors $s = 1, ..., n_s - 1$, with the last sensor having $w_{n_s} = 1 - 0.019(n_s - 1)$. This setup ensures balanced contributions from all sensors, while allowing for some to be prioritized based on their criticality.

Figure 6 shows the quality scores computed using Eqs. (3)–(5). The *DQS* indicates overall stability, while effectively highlighting both positive and negative peaks in the quality of the received blocks. Notably, immediately before the failures (marked by red lines), both WQS and LWQS show an abrupt decline, and DQS dips into negative values. This pattern suggests that the system was nearing a breakdown. Increasing the value of $m$ could enhance the early detection of such quality score declines, providing more advanced notice. The proposed accuracy computation efficiently handles outliers and integrates the accuracy dimension by counting observations within ideal values, ensuring efficient data quality score computation.

When aggregating sensor data to compute quality scores (as shown in Eqs. (6)–(7)), the overall quality behavior is similar to that in Fig. 6, since all sensor weights are equal. While this approach offers greater computational efficiency, crucial in IoT, it treats all sensors equally, limiting flexibility. In contrast, using Eqs. (4)–(5) allows decision-makers to prioritize sensors based on their relevance. Given that no specific sensor information was available, equal sensor weighting was assumed, leading to similar results. Regardless of the method, both approaches offer flexibility in prioritizing dimensions based on the decision-maker's needs.

The proposed data quality scores is designed as a flexible metric for assessing data quality in IoT environments. Unlike traditional techniques that rely on fixed criteria, the DQS adapts to varying contexts and data needs. Data quality is multifaceted, influenced by both its creation and context of use, impacting business performance [7,9]. By incorporating key dimensions like accuracy and completeness, the DQS offers a structured approach to real-time quality assessment. Its flexibility allows for easy integration of additional dimensions, such as

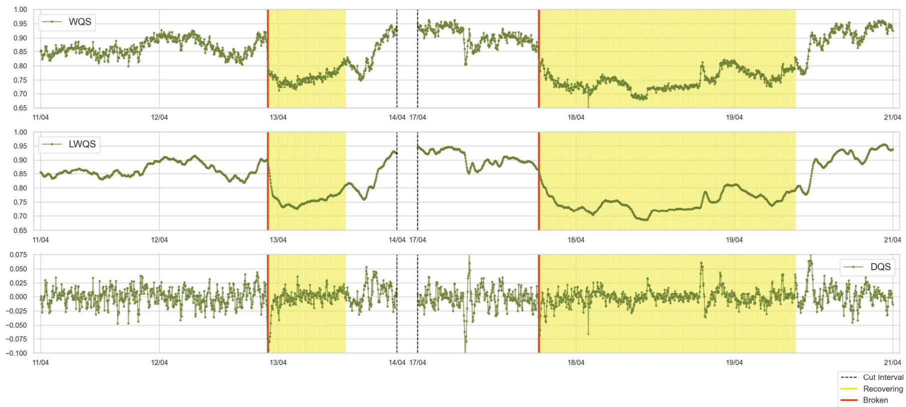


**Fig. 6.** WQS (Top), LWQS (Middle), and DQS (Bottom) computed using Eqs. (3)–(5).

timeliness or consistency, making it adaptable to diverse IoT scenarios where the importance of each dimension may differ.

## 5 Conclusion

This paper addresses the challenge of ensuring high data quality in IoT deployments. Low-quality data can compromise decision-making and analytics reliability. To address this, we introduce a data quality score that combines key dimensions into a comprehensive measure, providing a practical tool for assessing data quality in IoT environments.

The results of this study show that the proposed data quality scores can effectively integrate various quality metrics and offer a holistic assessment of the data in IoT systems. By applying these scores, organizations can quickly identify data quality issues and make informed decisions regarding data management and improvement strategies. Moreover, the proposed scores can be adapted to different IoT applications, enhancing its utility across a wide range of deployment contexts.

Future work will refine and extend the proposed score for real-time data quality monitoring, integrating machine learning for automated profiling and repair to enable dynamic, context-aware adjustments. Additionally, research will explore incorporating the score into IoT frameworks and platforms to enhance data management. Future efforts will also focus on tuning the score for specific use cases, prioritizing relevant dimensions based on system needs, ultimately supporting more robust and reliable data-driven solutions.

**Acknowledgments.** This work has been supported by national funds through FCT - Fundação para a Ciência e Tecnologia through projects UIDB/04728/2020 and UIDP/04728/2020.

## References


1. Sicari, S., Cappiello, C., De Pellegrini, F., Miorandi, D., Coen-Porisini, A.: A security-and quality-aware system architecture for Internet of Things. Inf. Syst. Front. **18**(4), 665–677 (2016)
2. Goknil, A., et al.: A systematic review of data quality in CPS and IoT for industry 4.0. ACM Comput. Surv. **55**(14S) (2023)
3. Karkouch, A., Mousannif, H., Moatassime, H.A., Noel, T.: Data quality in Internet of Things: a state-of-the-art survey. J. Netw. Comput. Appl. **73**, 57–81 (2016)
4. Loshin, D.: The practitioner's guide to data quality improvement. Morgan Kaufmann (2011). ISBN: 9780123737175
5. Cichy, C., Rass, S.: An overview of data quality frameworks. IEEE Access **7**, 24634–24648 (2019)
6. Zhang, L., Jeong, D., Lee, S.: Data quality management in the internet of things. Sensors **21**(17) (2021)
7. Mahanti, R.: Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. ASQ Quality Press, Milwaukee (2019)
8. Zhang, L., Jeong, D., Lee, S.: Data quality management in the internet of things. Sensors **21**(17), 5834 (2021)
9. Batini, C., Scannapieco, M.: Data-Centric Systems and Applications: Data and Information Quality. Springer, Heidelberg (2008). http://www.springer.com/series/5258
10. Tverdal, S., et al.: Edge-based data profiling and repair as a service for IoT. In: ACM International Conference on Proceeding Series, pp. 17–24 (2023)
11. Byabazaire, J., O'Hare, G., Delaney, D.: Using trust as a measure to derive data quality in data shared IoT deployments. In: 2020 29th International Conference on Computer Communications and Network (ICCCN), pp. 1–9 (2020)
12. Roy, A.K., Jena, K.K., Mohapatra, D.: Water pump health prediction in industrial wireless sensor network using supervised learning. In: 2023 OITS International Conference on Information Technology (OCIT), pp. 178–183. IEEE (2023)
13. Costa e Silva, E., Oliveira, Ó., Oliveira, B.: Enhancing real-time analytics: streaming data quality metrics for continuous monitoring. In: Proceedings of the 2024 7th International Conference on Mathematics and Statistics, pp. 97–101 (2024)
14. Liu, C., Peng, G., Kong, Y., Li, S., Chen, S.: Data quality affecting big data analytics in smart factories: research themes, issues, and methods. Symmetry *13*(8), 1440 (2021)