Temperature Sensor Data Quality Assessment in Manufacturing Environment Using Hampel Filter and QSD

Bambang Purnomosidi D.P [1]
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia
Yogyakarta, Yogyakarta, Indonesia
email: bpdp@utdi.ac.id

Widyastuti Andriyani
Master in Information Technology
Faculty of Information Technology
Universitas Teknologi Digital Indonesia
Yogyakarta, Yogyakarta, Indonesia
email: widya@utdi.ac.id

Akhmad Dahlan
Faculty of Computer Science
Informatics Management
emial: alland@amikom.ac.id

In the Industry 4.0 era, integrated temperature sensors in system production become source main data for taking decisions. However, the quality of the data produced often influenced by noise, missing values, and disturbing anomalies accuracy of analytical processes. Research This proposes a monitoring pipeline designed data quality For environment manufacturing based on the Internet of Things (IoT), with focus on usage Hampel Filter and Quality Score Delta (QSD) methods. Hampel Filter is used for detecting and handling outliers in temperature data in a way adaptive, while QSD is used for measure dynamics change data quality from time to time. Architecture system built with using Apache Kafka for data ingestion, InfluxDB For time-series storage, and Grafana for real-time visualization. Case study performed on temperature sensor data from the conveyor motor, and the results show that method. This capable detect degradation data quality in general proactive. Findings show potential big in increase reliability industrial monitoring system as well as support maintenance predictive data- based. Research This give contribution significant in developing modular and adaptive approach for management data quality in the manufacturing sector.

KeyWords: *Data Quality, Temperature Sensor, Industry 4.0, Hampel Filter, QSD.*

This Article was:
submited: 25-06-25
accepted: 08-07-25
publish on: 20-07-25

How to Cite:
B. Purnomosidi. D. P, et al, "Temperature Sensor Data Quality Assessment in Manufacturing Environment Using Hampel Filter and QSD", Journal of Intelligent Software Systems, Vol.4, No.1, 2025, pp.11–15, 10.26798/jiss.v4i1.2003

1 Introduction

Digital transformation in the industrial world has pushed the adoption of Internet of Things (IoT) technology to increase efficiency and operational accuracy. In the context of modern manufacturing, sensors become component crucial for collecting real-time data used in taking decision related process control, maintenance predictive, and early detection failure. One of the types of sensors that are often used is a temperature sensor, which plays an important role in monitor performance engine, detect overheating, and maintain stability system production [1].

Although the data from the temperature sensor own a big potential for increasing efficiency operational, poor data quality can become obstacle significant. Challenges that are often faced include missing data (missing values), format inconsistency, outdated data (untimeliness), and the presence of outliers due to disturbance environment or sensor failure. If not handled with right, the data is not accurate. This can lead to wrong decisions, decline productivity, up to loss financially [2].

For overcome problem mentioned, it is necessary system evaluation data quality capable of Work real -time, modular and adaptive to condition dynamic in the environment manufacturing. One of a promising approach is use Hampel Filter as method statistics For outlier detection and handling based on median absolute deviation (MAD), which has been proven more robust against fluctuations in sensor data. In addition, metrics are also used Quality Score Delta (QSD) which calculate difference mark current data quality to trend historical, so that can detect degradation data quality in general proactive [3], [4].

With this approach, it is expected system can support taking more data based reliable decisions, improve the readiness system production to disturbances, as well as open opportunity implementation continued on other sensor variables like vibration, pressure, and current electricity.

1.1 Literature Review.

Data quality is element key in system data -based, especially in the environment an increasingly growing industry depend on Internet of Things (IoT) technology and smart manufacturing [5]. In the context of Industry 4.0, IoT sensors generate data continuous in time-series form, with frequency high and deep amount large [6]. However, the high volume and speed of this data not always ensure utility value when data quality is not fulfill standard For analysis that can reliable.

1.1.1 Data Quality Dimensions.

Data quality can measure through a number of dimensions, such as accuracy, completeness, consistency, and precision time. Strengthens that dimensions this is very relevant in environment manufacturing intelligent Because related direct with effectiveness taking decision, efficiency operational and implementation maintenance predictive [4], [7], [8], [9]. For example, data that is not accurate or not complete can interfere with the detection process early to failure machine, while the data is not consistent will make it difficult integration cross system [10].

1.1.2 Profiling and Mapping Quality.

Data profiling is the process of analyze data characteristics, such as distribution value , pattern anomaly, or frequency of missing values. This profiling underlying evaluation metric data quality [11]. Proposed integrated

[1] Corresponding Author.

profiling method direct with Kafka and InfluxDB based streaming architecture, so that capable handle data in real-time [12], [13]. Data quality evaluation is carried out in a way gradually through calculation metric specific for each dimension, and the results visualized using the Grafana dashboard [14].

1.1.3 Outlier Detection with Hampel Filter. One of problem common in sensor data is the existence of outliers that are not reflect actual condition in the field. For overcome matter This, Hampel Filter is used method statistics based on the median and median absolute deviation (MAD) which is proven effective in detect and correcting outliers without affected by distribution extreme. Compared with approach based on percentile fixed percentile, Hampel Filter is more adaptive to dynamics of sensor data in the environmentally volatile industry [15].

1.1.4 Quality Score Delta (QSD). For evaluate dynamics change data quality from time to time, introduced the Quality Score Delta (QSD). QSD is calculated as a difference between Weighted Quality Score (WQS) and Longitudinal Weighted Quality Score (LWQS), so capable represent trend decline or improvement data quality in a window time. Implementation of QSD is important in monitoring context because signal early to degradation sensor performance and potential disturbance system [16].

1.1.5 Architecture Industry Data Ingestion. Various study has develop industrial data ingestion architecture capable of handling complexity of sensor data. Highlight synchronization, segmentation time, and conversion protocol in ingestion system. However, some big architecture the Not yet integrate evaluation data quality in real-time [16].

2 Method

2.1 Experimental Design. This study uses public datasets from the Intel Berkeley Research Lab, which contain data from 54 wireless sensors installed inside Intel labs. These sensors measure temperature, humidity, and lighting in time-series format. Each sensor records data simultaneously and periodically with a resolution of approximately 31 seconds per observation, making the dataset highly suitable for experiments related to the Internet of Things (IoT) and for evaluating the quality of environmental sensor data.

In the study Here, a subset of data is selected from 3 temperature sensors (Temp1, Temp2, Temp3) are active and consistent. The data used covers range time around 5 minutes, with an average total of approx. 10 observations per sensor in that interval. For simulate condition environment non - ideal industry, added outlier (value extreme) and missing values (missing data) in a way under control in part data point. The use of the Intel Berkeley dataset provides profit Because:

(1) It is public and can be reproduced , improved transparency study
(2) Own sufficient scale and complexity for represent real -world IoT conditions.
(3) Allows data quality pipeline testing realistically with sensor and time variations.

2.2 Application of Method. This method used for detect outliers statistics. The Hampel Filter calculates the median and median absolute deviation (MAD) on a sliding window for every data point. If $|X_i - \text{median}(X)| > T$, so X_i considered an outlier. The k value is generally between 2.5 to 3. Temp1 sensor shows fluctuation significant success identified

2.3 Evaluation of Metrics of Data Quality. Evaluation was carried out every minute to four dimensions data quality: accuracy, completeness, consistency, and precision time use settlement as following :

2.4 Interpretation Calculation. The accuracy of the Temp1 sensor shows fluctuating values, which indicates the presence of outliers or anomaly local in the recorded data. This condition shows that sensor reading not always stable and able influenced by environmental factors or disturbance technical. In addition, the level of Complete data is below the figure 0.95 indicates existence data loss, which is a big possibility caused by interference such as noise in the transmission process or buffer overrun occurs in the system While that, consistency inter sensor tends to low, even in a number of cases show mark negative. This reflects the existence of inconsistency measurements that can originate from difference sensor configuration or variation condition environment around the sensor. On the other hand, the timeliness metrics in simulation show mark perfect, but in its application in the real world, performance This can decrease consequence existence latency on IoT devices or delay network communication like MQTT.

2.5 Calculation Index Combined Quality: WQS, LWQS, and QSD.

2.5.1 Weighted Quality Score (WQS).

$$\text{WQSj}_j = w_a\left(\sum_s \frac{A_{sj}}{3} + w_c \sum_s \frac{C_{sj}}{3}\right) \quad (1)$$

example with $w_a = 0.7$, $w_c = 0.3$
Minute 1: $WQS_j = 0.7\frac{0.47+0.49+0.56}{3} + 0.3\frac{0.95+0.93+0.83}{3} = 0.6257$

2.5.2 Longitudinal Weighted Quality Score (LWQS). Using the average WQS of 10 blocks previously, with weight exponential

$$fk = exp\left(-\frac{j-k-i}{\beta}\right) \quad (2)$$

Example: If the
historical average the heaviest produces 0.628 → then: $LWQS_1 = 0.628$

2.5.3 Quality Score Delta (QSD).

$$QSD_1 = WQS_1 - LWQS_1 = 0.6257 - 0.628 = -0.0023$$

3 Results and Discussion

3.1 Problems with Data Quality in Industrial Environments. In the sensor monitoring system, data loss (missing values) often occurs as a consequence of transmission disturbances, the presence of noise, or even damage to the sensor device itself. The existence of extreme outliers in the data reflects significant anomalies, which can be caused by power fluctuations or abnormal temperature conditions. Low-level consistency between sensors is also important, as it indicates a lack of synchronization between devices or possibly non-uniform configurations. Potential delays in data delivery can disrupt the effectiveness of the real-time monitoring system. Although simulations assume ideal conditions, real-world challenges such as network latency or communication obstacles must still be anticipated with appropriate mitigation strategies.

3.2 Impact to Industrial Systems. The problem of sensor data quality—such as missing values, outliers, inconsistencies between sensors, and time delays—can have a serious impact on industrial systems as a whole. One of the most significant consequences is the emergence of errors in predictive maintenance decision-making. When sensor data does not accurately reflect the actual condition, the system may produce false alarms or fail to detect early signs of equipment degradation. Unexpected downtime becomes a real risk when incorrect temperature readings trigger system shutdowns or reduce production speed.

Table 1 Metrics Data Quality

Metric	Formula	Information
Completeness	$C_{sj} = \frac{N^{sj}_{valid}}{N_{total}}$	Proportion of available data from a total of 60 data per minute
Accuracy	$A_{sj} = \frac{1}{N} = \sum_{i=1}^{N} \frac{x_i - min(x)}{max(x)-min(x)}$	Average value min-max normalization in 1 minute
Consistency	$A_{sj} = \frac{1}{N} = \sum_{i=1}^{N} \frac{x_i - min(x)}{max(x)-min(x)}$	Average Pearson correlation between sensor pair
Timeliness	Timeliness_j = 1.0$_{(ideal)}$	In simulation considered No there is a time delay delivery

For example, cooling systems or lubrication mechanisms that rely on temperature data may operate suboptimally, leading to machinery damage or reduced product quality. The performance of automatic control systems also heavily depends on the accuracy and reliability of sensor data. When the system receives invalid or inconsistent input, the resulting responses may not match real-world conditions, causing deviations from production targets or even posing operational hazards.

Inaccurate trend analysis of production or misestimation of machine workload may occur when relying on poor-quality data. This, in turn, impacts resource planning, production capacity settings, and inefficient rescheduling. Low sensor data quality not only hampers the technical function of industrial systems but also disrupts strategic and operational decision-making in data-based environments.

3.2.1 Calculation Metric Data Quality. The table below display results measurement metric quality from Temp1, Temp2, and Temp3 sensors per minute:

3.3 Calculations Index Combination: WQS, LWQS, and QSD. For evaluate data quality longitudinally and aggregately, used three index:

(1) WQS (Weighted Quality Score) :
$WQS_j = 0.7.A_j + 0.3\dot{C}_j$
(2) LWQS (Longitudinal WQS) :
$LWQS_j = \frac{\sum_{k-1}^{j-1} fk\dot{W}QS_k}{\sum_{k-1}^{j-1} fk}, fk = \epsilon^{(j-k-1)/\beta}$
(3) QSD (Quality Score Delta) :
$QSD_j = WQS_j - LWQS_j$

3.4 Visualization of Results. Visualization This serve four dimensions main sensor data quality (accuracy, completeness, consistency) and dynamics change quality through Quality Score Delta (QSD) index. This visual analysis give deep temporal overview to performance sensor systems in the environment manufacturing:

a. Sensor Accuracy per Minute show difference level reliability between sensors. From the graph , it can be seen that the Temp3 sensor shows mark the most stable accuracy compared to Temp1 and Temp2, which experienced fluctuation Enough significant. Temp3 stability indicates that the sensor stands to disturb environment or local noise , and tends to produce measurement higher temperature can reliable as seen on figure 1.
b. Data Completeness per Minute show fluctuation data occupancy, especially on the Temp3 sensor. Some minute show mark completeness below 0.9, which indicates the possibility there is missing data due to disturbance transmission or decline performance device. This condition demands more attention on integrity channel communication or buffer management in system data collection as seen on figure 2.
c. Consistency between Sensors displayed in form mark Pearson correlation between combination sensor pair. Interestingly, the graph shows that coefficient correlation at several time intervals nature negative. This reflects the existence of discrepancy measurement between sensors, which can be caused by differences position, orientation, or sensor calibration. Correlation value negative indicates that these sensors read trend opposite temperatures direction, so that can lower trust to uniformity system sensing as seen on figure 3.
d. QSD per Minute (Quality Score Delta) visualizes difference between current data quality This with trend historical. The QSD value is close to zero and show stability data quality from time to time. However, the graph also shows a number of significant QSD spike indicator this is very important Because reflect moments when happen change big to condition data quality. QSD is soaring to signify possibility existence anomaly new or change in dynamics system production that is needed quick followed up as seen on figure 4.

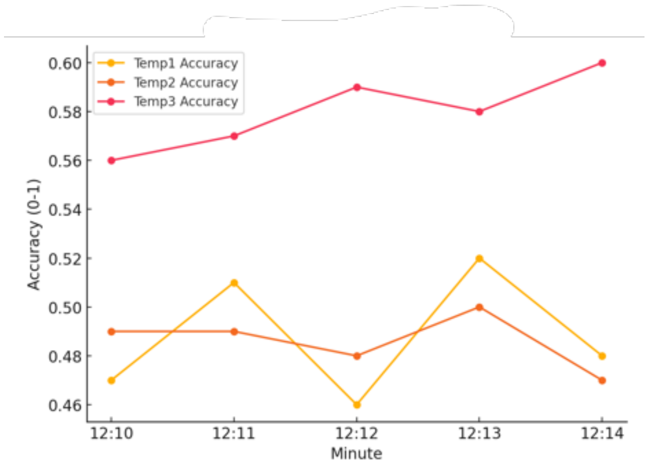

Fig. 1 Sensor Accuracy Per- minute

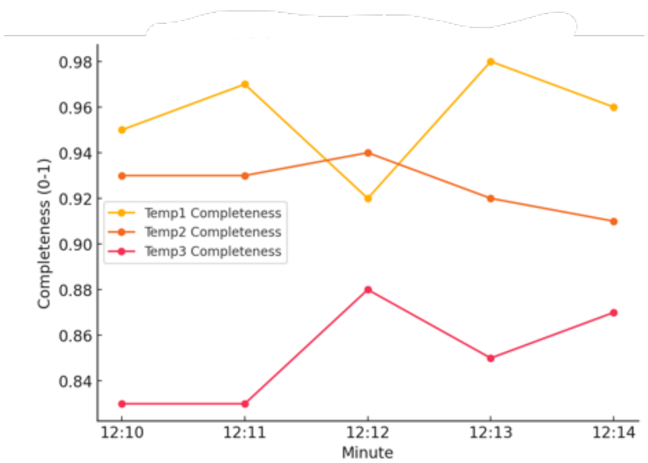

Fig. 2 Completeness of Data Per- minute

Visualization This confirms effectiveness method evaluation proposed data quality in research. Every dimension quality capa-

Table 2 Sensor Metrics

Temp1 Accuracy	Temp1 Completeness	Temp2 Accuracy	Temp2 Completeness	Temp3 Accuracy	Temp3 Completeness	Consistency	Timeliness
0.47	0.95	0.49	0.93	0.56	0.83	-0.12	1
0.51	0.97	0.49	0.93	0.57	0.83	-0.08	1
0.46	0.92	0.48	0.94	0.59	0.88	0.01	1
0.52	0.98	0.50	0.92	0.58	0.85	0.05	1
0.48	0.96	0.47	0.91	0.60	0.87	-0.03	1

Table 3 Measurement Results Table Metric

Minute	WQS	LWQS	QSD
12:10:00 PM	0.6257	0.6257	0
12:11:00 PM	0.6364	0.631	0.0054
12:12:00 PM	0.6366	0.6329	0.0037
12:13:00 PM	0.6504	0.6373	0.0131
12:14:00 PM	0.6495	0.6397	0.0098

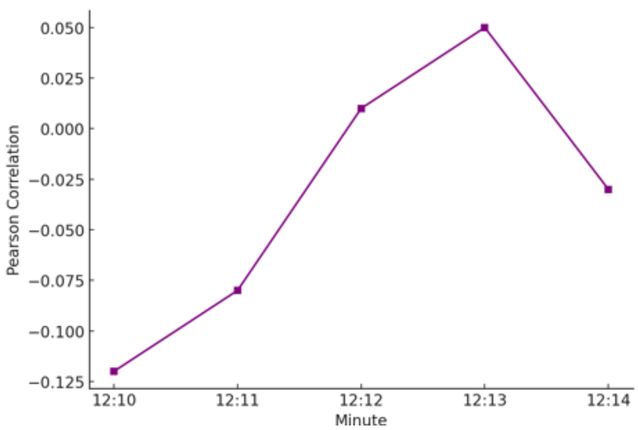

Fig. 3 Consistency Between Sensors Per- minute

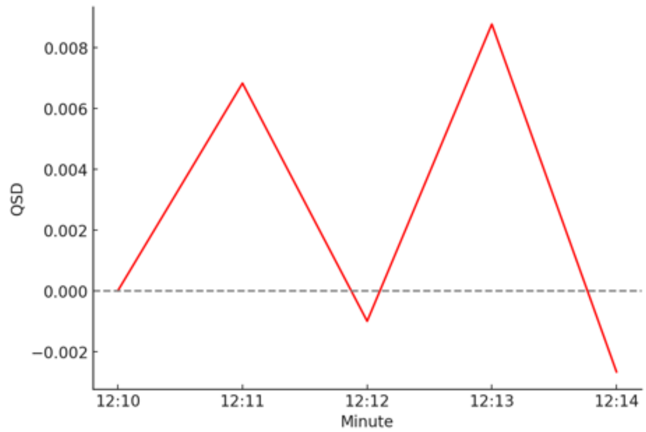

Fig. 4 Changes Stability Quality (QSD)

ble catch dynamics different in sensor system, while QSD provides mechanism adaptive and sensitive longitudinal monitoring to change contextual.

3.5 Implications towards Industry. Implementation system evaluation Hampel Filter and Quality Score Delta (QSD) based sensor data quality in environment manufacturing show a number of implications important for industry.

(1) The system has the capability to detect real-time data quality degradation, which is crucial in production environments that demand high accuracy and continuity. Early identification of anomalies, missing data, or reduced accuracy enables quick responses before these issues impact the production process.
(2) The system enhances the reliability of data-driven decision-making. With measurable data quality information, managers and control systems can filter out invalid data before using it in performance evaluation, production schedule optimization, or operational condition monitoring. This directly contributes to operational efficiency and reduces the risk of errors caused by poor data quality.
(3) The integration of such systems provides a strong foundation for the development of automatic control and predictive maintenance systems. With automated and longitudinal data quality monitoring, the system signals the reliability of sensors and production processes, forming a key component in the transformation toward smart manufacturing. More advanced implementations can include automatic notifications, QSD-based alarm triggers, and adaptive actions powered by machine learning algorithms.

System This not only functioning as tool evaluation passive, but also as mechanism active for ensure data integrity, support taking decision based on facts, and strengthen foundation system automation industry based on intelligence artificial and IoT.

4 Conclusion

This study successfully developed and implemented an evaluation of temperature sensor data quality in a manufacturing environment using the Hampel Filter method for outlier detection and Quality Score Delta (QSD) as a longitudinal indicator. The data used came from Intel Berkeley Research Lab, which reflects the real conditions of the IoT system in recording temperature data periodically.

Through this approach, the study measured four dimensions of data quality, accuracy, completeness, consistency between sensors, and timeliness calculated per minute. The results of the metric and QSD calculations showed that the method used was able to identify fluctuations in data quality, detect sensor performance degradation, and inform critical moments of data quality changes.

The application of this method has significant practical implications in supporting the reliability of production monitoring systems, predictive maintenance, and data-based automatic control system integration. Overall, the developed model makes a concrete contribution to increasing the trust and quality of sensor information in IoT-based manufacturing systems.

Based on the findings in this study, several suggestions for further development can be proposed. First, the integration of machine learning algorithms, such as Random Forest or LSTM, is recommended to enhance anomaly detection and support the prediction of long-term sensor quality trends. Additionally, testing the system on a broader and more complex network involving multiple sensors and locations is essential to assess its scalability and robustness in dynamic industrial environments. Furthermore, incorporating contextual validity aspects—beyond statistical metrics—can improve data quality evaluation by considering factors such as the type of industrial process or specific machine characteristics. Another development opportunity lies in the automation of follow-up responses, where the system could be equipped with mechanisms for automatic notifications or actions, such as recalibration or early warnings, based on QSD threshold values and observed longitudinal trends. With a more integrative and data-driven approach, the sensor quality system proposed in this study has the potential to serve as a solid foundation for advancing digital transformation efforts within the framework of Smart and Sustainable Manufacturing.

References

[1] Javaid, M., Haleem, A., Pratap, R., and Kumar, A., 2024, "Green Technologies and Sustainability Digital economy to improve the culture of industry 4.0: A study on features, implementation and challenges," Green Technology and Sustainability, **2**(2), p. 100083.

[2] Morewood, J., 2023, "Building energy performance monitoring through the lens of data quality: A review," Energy and Buildings, **279**, p. 112701.

[3] Berendrecht, W., Vliet, M. V., and Griffioen, J., 2023, "Combining statistical methods for detecting potential outliers in groundwater quality time series," Environmental Monitoring and Assessment.

[4] Bezerra, A., Júnior, B., and Sérgio, P., 2014, "An Approach to Outlier Detection and Smoothing Applied to a Trajectography Radar Data," **6**, pp. 237–248.

[5] Luque, A. et al., 2017, "State of the Industry 4.0 in the Andalusian food sector," Procedia Manufacturing, **13**, pp. 1199–1205.

[6] Javaid, M., Haleem, A., Pratap, R., Rab, S., and Suman, R., 2021, "Significance of sensors for industry 4.0: Roles, capabilities, and applications," Sensors International, **2**(May), p. 100110.

[7] Alabduljabbar, A. and Alshammari, A., 2024, "The Effect of Data Quality on Decision-Making: A Quasi Experimental Study," **3**, pp. 1552–1566.

[8] Barroso, P., Miguel, B., Bernardo, V., and Henrique, S., 2024, "Journal of Innovation," Journal of Innovation, **9**.

[9] Miller, R., Hin, S., Chan, M., and Whelan, H., 2025, "A Comparison of Data Quality Frameworks: A Review," (Ml), pp. 4–13.

[10] Jayawardene, V., Huggins, T. J., Prasanna, R., and Fakhruddin, B., 2021, "The role of data and information quality during disaster response decision-making," Progress in Disaster Science, **12**, p. 100202.

[11] Xie, J., Sun, L., and Fiona, Y., 2025, "On the Data Quality and Imbalance in Machine Learning-based Design and Manufacturing — A Systematic Review," Engineering, **45**, pp. 105–131.

[12] Kafka, E. and Lekkala, C., 2024, "Integration of Real-Time Data Streaming Technologies in Hybrid Cloud Environments: Kafka, Spark, and Kubernetes," October 2022.

[13] Oladele, S., 2025, "Design and Implementation of a Real-Time Data Processing Framework for High-Throughput Applications," January.

[14] Iuhasz, G., Fortiş, T.-F., and Panica, S., 2025, "Exploring machine learning methods for the identification of production cycles and anomaly detection," Internet of Things, **30**(November 2024), p. 101508.

[15] Toribio, M. R., García, I. A., Menéndez, A. G., and Toribio, S. R., 2025, "A Novel Approach to Speed Up Hampel Filter for Outlier Detection," .

[16] Peixoto, T., Oliveira, O., Costa, E., and Oliveira, B., 2025, "A Data Quality Pipeline for Industrial Environments: Architecture and Implementation," **C**, pp. 1–23.