International Standard

ISO 8000-210

First edition
2024-12

Data quality —

Part 210: Sensor data: Data quality characteristics

Qualité des données —

Partie 210: Données des capteurs: Caractéristiques de qualité des données

Reference number
ISO 8000-210:2024(en)

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 184, *Automation systems and integration*, Subcommittee SC 4, *Industrial data*.

A list of all parts in the ISO 8000 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

0.1 Foundations of the ISO 8000 series

Digital data deliver value by enhancing all aspects of organizational performance including:

— operational effectiveness and efficiency;

— safety and security;

— reputation with customers and the wider public;

— compliance with statutory regulations;

— innovation;

— consumer costs, revenues and stock prices.

In addition, many organizations are now addressing these considerations with reference to the United Nations Sustainable Development Goals[1].

The influence on performance originates from data being the formalized representation of information[2]. This information enables organizations to make reliable decisions. This decision making can be performed by human beings and also automated data processing, including artificial intelligence systems.

Organizations become dependent on digital data through widespread adoption of digital computing and associated communication technologies. This dependency amplifies the negative consequences of the lack of quality in these data, leading to the decrease of organizational performance.

The biggest impact of digital data comes from two key factors:

— the data having a structure that reflects the nature of the subject matter;

EXAMPLE 1 A research scientist writes a report using a software application for word processing. This report includes a table that uses a clear, logical layout to show results from an experiment. These results indicate how material properties vary with temperature. The report is read by a designer, who uses the results to create a product that works in a range of different operating temperatures.

— the data being computer processable (machine readable) rather than just being for a person to read and understand.

EXAMPLE 2 A research scientist uses a database system to store the results of experiments on a material. This system controls the format of different values in the data set. The system generates an output file of digital data. This file is processed by a software application for engineering analysis. The application determines the optimum geometry when using the material to make a product.

ISO 9000 explains that quality is not an abstract concept of absolute perfection. Rather, quality is the conformance of characteristics to requirements. This means that any item of data can be of high quality for one purpose but not for a different purpose. The quality is different because the requirements are different between the two purposes.

EXAMPLE 3 Time data are processed by calendar applications and also by control systems for propulsion units on spacecraft. These data include start times for meetings in a calendar application and activation times in a control system. These start times require less precision than the activation times.

1) https://sdgs.un.org/goals

2) ISO 8000-2 defines information as “knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning”.

The nature of digital data is fundamental to establishing requirements that are relevant to the specific decisions made by each organization.

EXAMPLE 4 ISO 8000-1 identifies that data have syntactic (format), semantic (meaning) and pragmatic (usefulness) characteristics.

To support the delivery of high-quality data, the ISO 8000 series addresses:

— data governance, data quality management and maturity assessment;

EXAMPLE 5 ISO 8000-61 specifies a process reference model for data quality management.

— creating and applying requirements for data and information;

EXAMPLE 6 ISO 8000-110 specifies how to exchange characteristic data that are master data.

— monitoring and measuring information and data quality;

EXAMPLE 7 ISO 8000-8 specifies approaches to measuring information and data quality.

— improving data and, consequently, information quality;

EXAMPLE 8 ISO/TS 8000-81 specifies an approach to data profiling, which identifies opportunities to improve data quality.

— issues that are specific to the type of content in a data set.

EXAMPLE 9 ISO/TS 8000-311 specifies how to address quality considerations for product shape data.

Data quality management covers all aspects of data processing, including creating, collecting, storing, maintaining, transferring, exploiting and presenting data to deliver information.

Effective data quality management is systemic and systematic, requiring an understanding of the root causes of data quality issues. This understanding is the basis for both correcting existing inconsistencies and implementing solutions that prevent future reoccurrence of those nonconformities.

EXAMPLE 10 If a data set includes dates in multiple formats including “yyyy-mm-dd”, “mm-dd-yy” and “dd-mm-yy”, then data cleansing can correct the consistency of the values. Such cleansing requires additional information, however, to resolve ambiguous entries (such as, “04-05-20”). The cleansing also does not address any process issues and people issues, including training, that have caused the inconsistency.

0.2 Understanding more about the ISO 8000 series

ISO 8000-1 provides a detailed explanation of the structure and scope of the whole ISO 8000 series.

ISO 8000-2 specifies the single, common vocabulary for the ISO 8000 series. This vocabulary supports understanding the overall subject matter of data quality. ISO 8000-2 presents the vocabulary structured by a series of topic areas (for example, terms relating to quality and terms relating to data and information).

ISO has identified ISO 8000-1, ISO 8000-2 and ISO 8000-8 as horizontal deliverables[3)].

0.3 Role of this document

As a contribution to the overall capability of the ISO 8000 series, this document describes quality characteristics and related data anomalies of data produced by sensors. This document focuses, in particular, on data that are a stream of individual, discrete digital values. The quality characteristics and data anomalies can serve as the basis for quality criteria to measure and improve the quality of sensor data. Such criteria are suitable when preparing data prior to data analysis or exploitation of the data.

Sensors are a fundamental enabler of digital transformation, which has resulted in the proliferation of sensor networks and sensing devices connected to the Internet-of-Things (see ISO/IEC 30141). Such sensors capture data about a wide range of aspects of the physical world. These data have significant volume, velocity and

3) Deliverable dealing with a subject relevant to a number of committees or sectors or of crucial importance to ensure coherence across standardization deliverables.

variety, making them an essential asset serving as the basis for insight and foresight that improves decision making across organizations of all types.

While offering this potential, sensor data are also vulnerable to disruption from a wide range of sources, including limited capacity of hardware (such as processors, memory and batteries), software, congested wireless communications and impact from harsh operating environments. The data are, therefore, likely to include data anomalies, which require detection and handling in order to improve the quality of sensor data. The anomalies can affect various characteristics of the data, including the time-variant aspects of sensors.

The quality characteristics are appropriate for use when preparing sensor data for subsequent data analysis and exploitation. This use involves measuring and improving the quality of the data. These characteristics are relevant to any type of industrial application, including smart manufacturing, social infrastructure and healthcare.

This document supports activities that affect:

— one or more information systems;

— data flows within the organization and with external organizations;

— any phase of the data life cycle.

Organizations can use this document individually or in conjunction with other parts in the ISO 8000 series.

Annex A contains an identifier that conforms to ISO/IEC 8824-1. The identifier unambiguously identifies this document in an open information system.

0.4 Benefits of the ISO 8000 series

By implementing parts of the ISO 8000 series to improve organizational performance, an organization can achieve the following benefits:

— objective validation of the foundations for digital transformation of the organization;

— a sustainable basis for data in digital form becoming a fundamental asset class that the organization relies on to deliver value;

— securing evidence-based trust from other parties (including supply chain partners and regulators) about the repeatability and reliability of data and information processing in the organization;

— portability of data with resulting protection against loss of intellectual property and reusability across the organization and applications;

— effective and efficient interoperability between all parties in a supply chain to achieve traceability of data back to original sources;

— readiness to acquire or supply services where the other party expects to work with common understanding of explicit data requirements.

Data quality —

Part 210:
Sensor data: Data quality characteristics

1 Scope

This document specifies quality characteristics of data that are recorded by sensors as a stream of single, discrete digital values.

The following are within the scope of this document:

— quality characteristics of sensor data;

— types of anomalies in sensor data;

— relationships between quality characteristics of sensor data and anomalies in sensor data;

— application of quality characteristics of sensor data.

The following are outside the scope of this document:

— analogue, image, video and audio data that are captured by sensors;

— signal processing that converts or modifies analogue data to create digital data;

— methods to measure and improve data quality.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8000-2, *Data quality — Part 2: Vocabulary*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000-2 apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

4 Quality characteristics of sensor data

4.1 General

Quality characteristics are the foundation for being able to measure and improve data quality. Many different quality characteristics are specified by documents such as ISO 8000-8, ISO/IEC 25012, ISO 19157-1 and ISO/TR 21707. These documents specify relevant characteristics according to the subject of interest or the nature of the particular domain.

This document specifies quality characteristics for sensor data. These characteristics address the key features of such data. These features are the time-ordered sequence of the values in the sensor data stream and the variations in the patterns of those values over time. The characteristics are:

— accuracy (see 4.2);

— completeness (see 4.3);

— consistency (see 4.4);

— precision (see 4.5);

— timeliness (see 4.6).

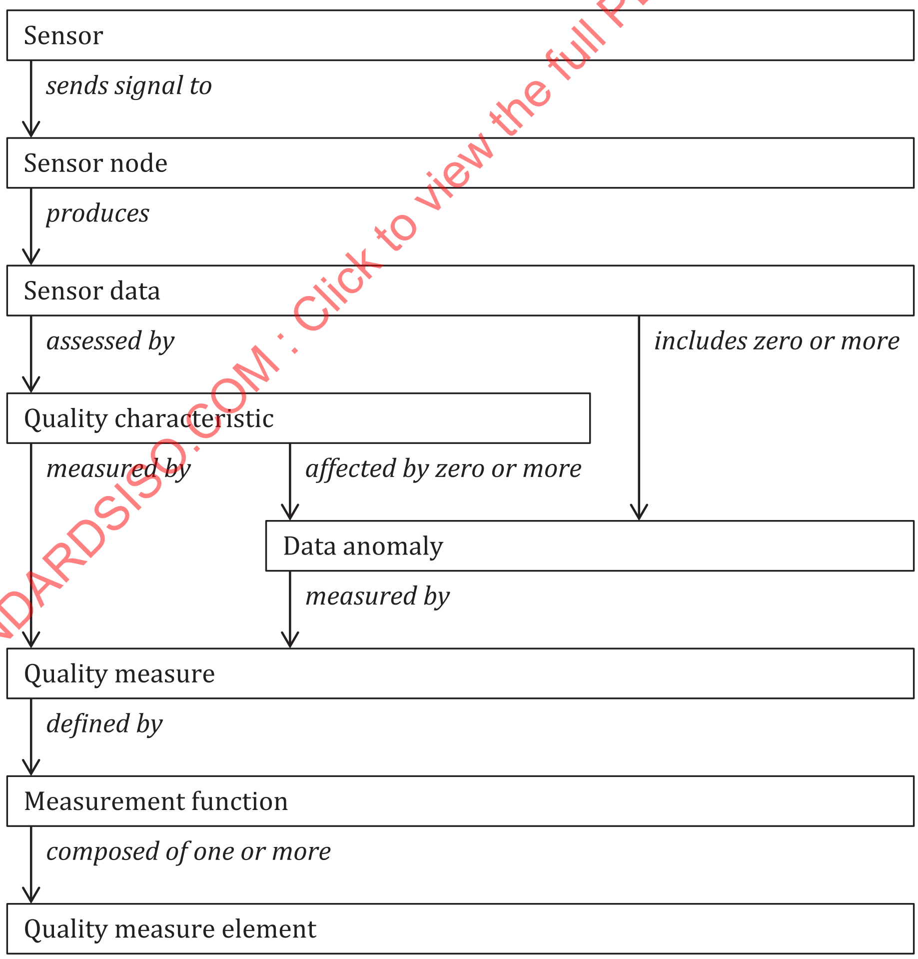
NOTE See Annex B for a comparison of these characteristics with those specified by other documents.

Such quality characteristics are (see Figure 1):

— typically affected by anomalies in the sensor data (see Clause 5 and Clause 6);

— measured by appropriate quality measures (with examples provided by this clause).

By providing the above capabilities, the quality characteristics are relevant to the data generated by different applications of sensors (see Clause 7).



NOTE See ISO/IEC 19505-1 for details on the notation in this diagram.

Figure 1 — Overview of sensor data quality

4.2 Accuracy

For sensor data, the data quality characteristic accuracy is the degree to which the sensor data, in a specific context of use, correctly represent the true value of the intended property of a concept or event. See Table 1 for an example of a data quality measure to provide quantification of the data quality characteristic accuracy.

NOTE 1 Accuracy is the basis for quality criteria that include tolerances, which determine whether sensor data values are accurate or inaccurate.

NOTE 2 ISO 8000-2 defines "data accuracy" as "quality of data in respect of the represented value agreeing with the corresponding true value to a degree necessary for an intended purpose".

NOTE 3 ISO/IEC Guide 99 defines "measurement accuracy" as "closeness of agreement between a measured quantity values and a true quantity value of a measurand". According to the guide, the concept "measurement accuracy" is not a quantity and is not given a numerical quantity value.

Table 1 — Example data quality measure for the data quality characteristic accuracy

Data quality measure	Measurement function	Quality measure elements within the function
Accuracy of a data set	A / B	A = count of those evaluated data values being accurate B = count of all evaluated data values

4.3 Completeness

For sensor data, the data quality characteristic completeness is the degree to which the sensor data, in a specific context of use, consists of all values expected to be captured. See Table 2 for an example of a data quality measure to provide quantification of the data quality characteristic completeness.

Table 2 — Example data quality measure for the data quality characteristic completeness

Data quality measure	Measurement function	Quality measure elements within the function
Completeness of a data set	A / B	A = count of those data values actually captured B = count of all data values expected to be captured

4.4 Consistency

For sensor data, the data quality characteristic consistency is the degree to which sensor data comply with rules imposed on the time-variant patterns or values produced by sensors. Such rules are derived from not only data patterns of a single sensor but also relationships between multiple sensors. See Table 3 for an example of a data quality measure to provide quantification of the data quality characteristic consistency.

Table 3 — Example data quality measure for the data quality characteristic consistency

Data quality measure	Measurement function	Quality measure elements within the function
Consistency of a data set	A / B	A = count of those measured data values conforming with applicable rules B = count of all measured data values

4.5 Precision

For sensor data, there are two kinds of the data quality characteristic precision.

a) Representational precision is the degree to which sensor data have a value that is exact or that is sufficient to provide discrimination in a specific context of use.

EXAMPLE A precision of five decimal places allows different functionalities rather than a precision of two decimal places.

b) Measurement precision is the degree to which sensor data are within a specified confidence bound of the random distribution based on the variance of a set of data values.

NOTE ISO/IEC Guide 99 and ISO 5725-1 do not recognize representational precision as being a kind of precision. In ISO/IEC Guide 99, precision is defined as the closeness of agreement between measured quantity values obtained by repeated measuring of the same or similar objects under specified conditions. This characteristic is usually expressed numerically by measures of imprecision, such as standard deviation, variance or coefficient of variation under the specified conditions of measurement.

See Table 4 for examples of data quality measures to provide quantification of the data quality characteristic precision.

Table 4 — Example data quality measure for the data quality characteristic precision

Data quality measure	Measurement function	Quality measure elements within the function
Representational precision of a data set	A / B	A = count of those measured data values having the required representational precision B = count of all measured data values
Measurement precision of a data set	A / B	A = count of those measured data values within a specified confidence bound of the random distribution B = count of all measured data values

4.6 Timeliness

For sensor data, the data quality characteristic timeliness is the degree to which sensor data are delivered within acceptable time limits for a specific time point or for a time period. This characteristic involves the time information for an individual sensor and the temporal relationship to related sensors. See Table 5 for an example of a data quality measure to provide quantification of the data quality characteristic timeliness.

Table 5 — Example data quality measure for the data quality characteristic timeliness

Data quality measure	Measurement function	Quality measure elements within the function
Timeliness of a data set	A / B	A = count of those measured data values having timestamps within acceptable time limits B = count of all measured data values

5 Types of anomaly in sensor data

5.1 General

Sensor data are the output from a sensor capturing changes in the physical world for a specific time point or for a time period. Sensors often operate in harsh environments, capturing large volumes of data, which are likely to include data anomalies. These anomalies are irregular or abnormal data patterns that include cases which reflect either real circumstances in the physical world or failures such as sensor faults and system errors.

NOTE Data anomalies occur due to reasons such as sensor faults, software errors, poor field conditions and hardware or network failures, where:

— a fault is any kind of defect that is the cause of an error;

— an error is an incorrect system state, where, without suitable corrective intervention, the state leads to a failure;

— a failure is the deviation of a system from the applicable specification, in other words, the deviation is the delivered service that is not the correct service.

EXAMPLE A sensor node periodically sends measured sensor data to a monitoring system. This node, however, has a battery voltage level that is low. The node, as a consequence, is unable to execute its embedded software at various points in time, while rebooting. The node does not send data to the monitoring system and, thus, does not provide the expected service. This situation arises from a fault, which is the low battery voltage, causing an error, which is the state of the software when not functioning, causing a failure, which is when the sensor node does not deliver the required services to the sensor network. The situation results in the sensor data suffering from a data anomaly called “stuck”, which is where sensor readings show zero (or near zero) difference over a period longer than when everything is functioning correctly.

Finding and handling such anomalies are important because they affect the quality of sensor data. Finding an anomaly is also typically a trigger to investigate root causes. Anomalies will, otherwise, adversely affect subsequent analysis or other exploitation of the data.

The following subclauses identify a set of typical data anomalies that appear in either the patterns of data captured by individual sensors (see 5.2) or in the correspondence between data sets captured by multiple related sensors (see 5.3).

5.2 Data anomalies for individual sensors

5.2.1 General

The following are typical data anomalies appearing in the patterns of data captured by individual sensors:

— offset (see 5.2.2);

— drift (see 5.2.3);

— trim (see 5.2.4);

— spike (see 5.2.5);

— noise (see 5.2.6);

— data loss (see 5.2.7);

— lack of amount (see 5.2.8);

— shift (see 5.2.9);

— drop or rise (see 5.2.10);

— stuck (see 5.2.11);

— bound oscillation (see 5.2.12);

— inconsistent frequency (see 5.2.13);

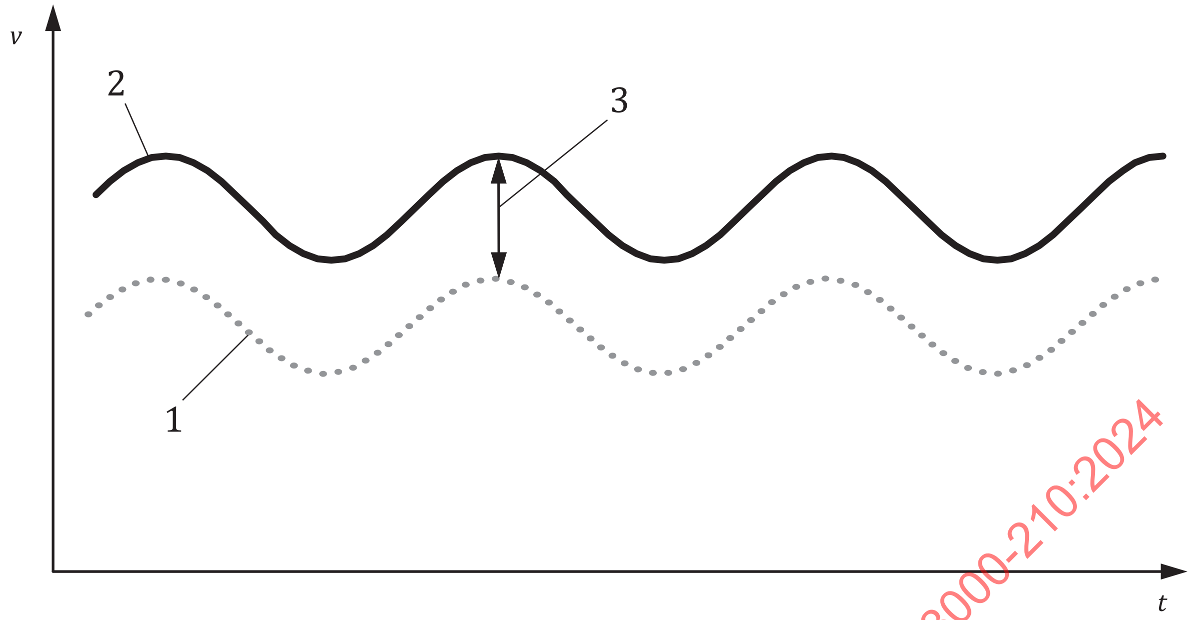
— different resolution (see 5.2.14);

— incorrect timestamp (see 5.2.15);

— latency (see 5.2.16).

5.2.2 Offset

Offset is an anomaly in which captured data values always deviate from the expected values by a consistent gap (see Figure 2).



Key

1 expected data values

2 data values captured by the sensor

3 constant offset between measured and expected values

t time at which the sensor captures the value of interest (*v*)

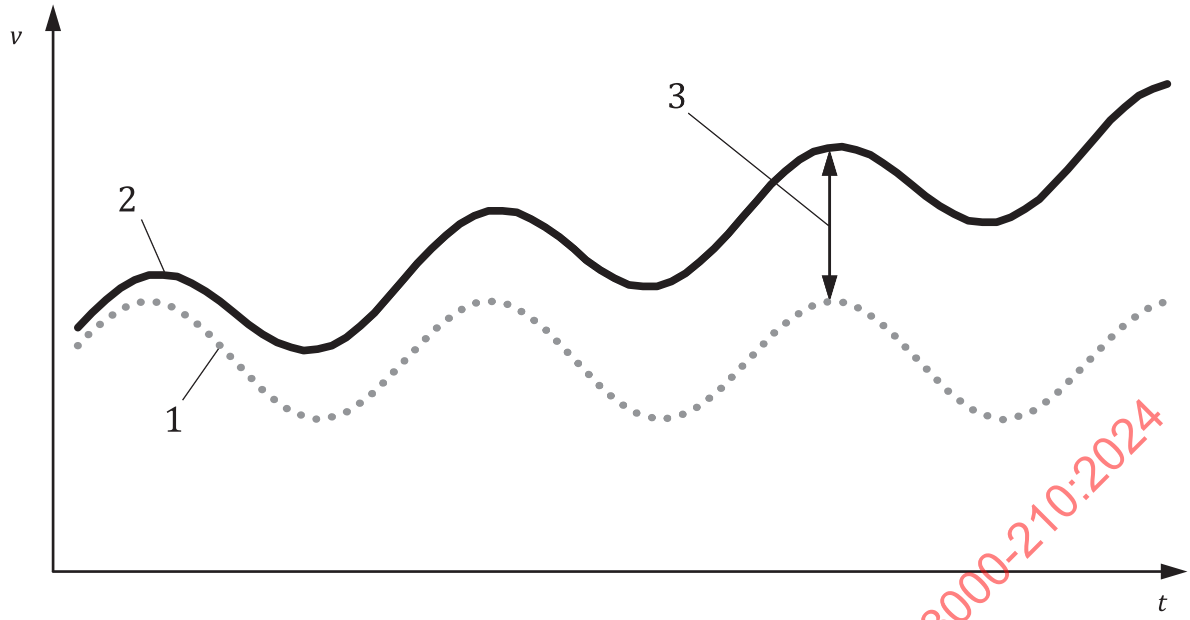
v value of interest the sensor is designed to capture

Figure 2 — Offset

EXAMPLE Although an expected range of sensor data values is between 1 and 3, the actual range measured by a sensor is between 2 and 4. Each value in this actual range has a constant difference from the expected value.

5.2.3 Drift

Drift is an anomaly in which there is a deviation between the captured data values and the expected data values. This deviation continuously either increases or decreases (see Figure 3).



Key

1 expected data values

2 data values captured by the sensor

3 growing offset between measured and expected values

t time at which the sensor captures the value of interest (*v*)

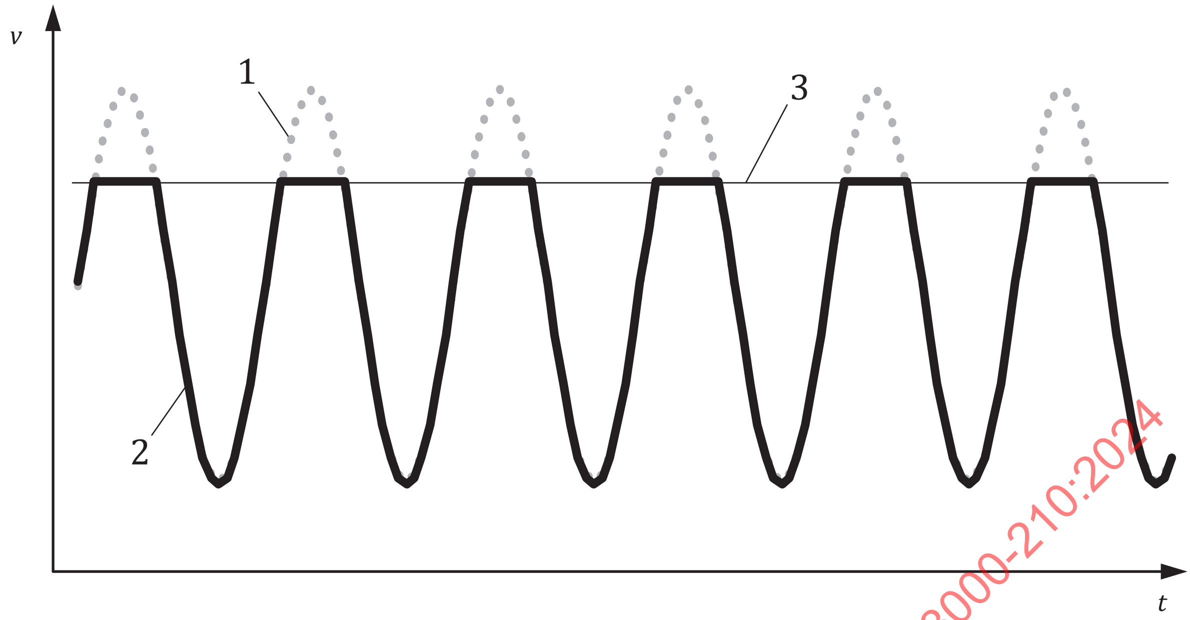
v value of interest the sensor is designed to capture

Figure 3 — Drift

EXAMPLE Sensor data start from a normal state but then differ from that state with a deviation that increases linearly over time.

5.2.4 Trim

Trim is an anomaly in which a sensor only has the capability to capture data values within a range that is too narrow to cover all the true values that occur (see Figure 4).



Key

1 expected data values

2 data values captured by the sensor

3 threshold beyond which the sensor cannot capture expected value

t time at which the sensor captures the value of interest (*v*)

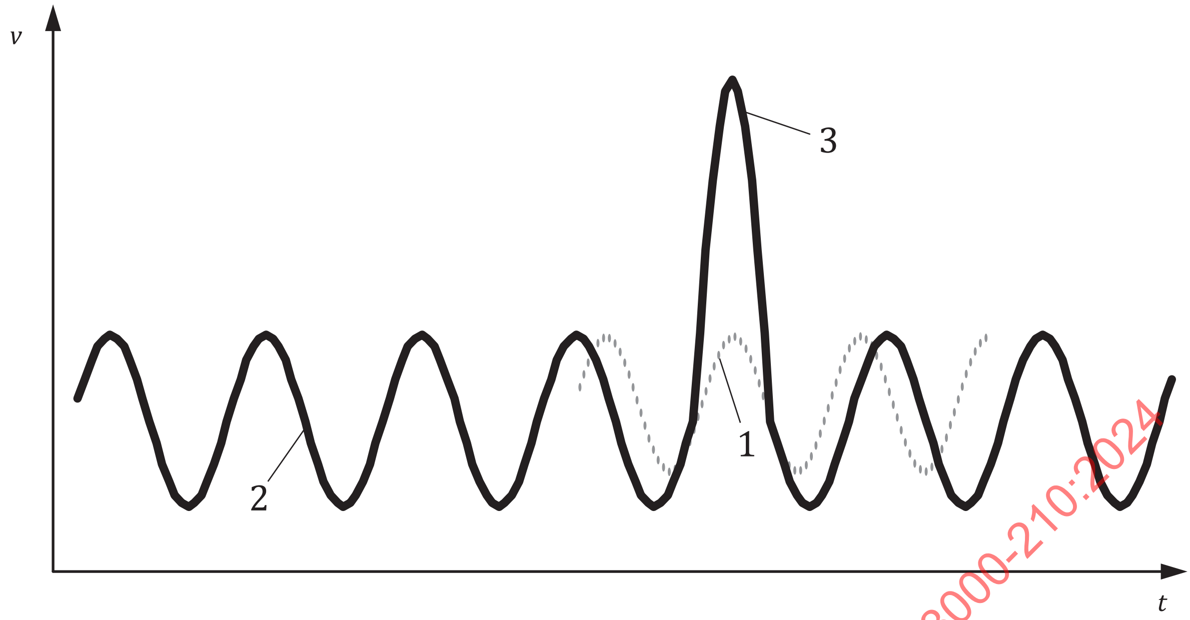
v value of interest the sensor is designed to capture

Figure 4 — Trim

EXAMPLE A sensor captures data for the surface temperature of a molten metal. These data include multiple values of "1 000 °C", which is actually the measurement limit of the sensor rather than the true value of the temperature, which is over 1 000 °C.

5.2.5 Spike

Spike is an anomaly in which sensor data do not match the pattern of normal values at a point of time. This anomaly, generally, shows an abnormal pattern over a period of time. The data are sometimes only abnormal for a single value before the data return to normal. Such single values are an outlier (see Figure 5).



Key

1 expected data values

2 data values captured by the sensor

3 spike that deviates excessively from corresponding one or more expected values

t time at which the sensor captures the value of interest (*v*)

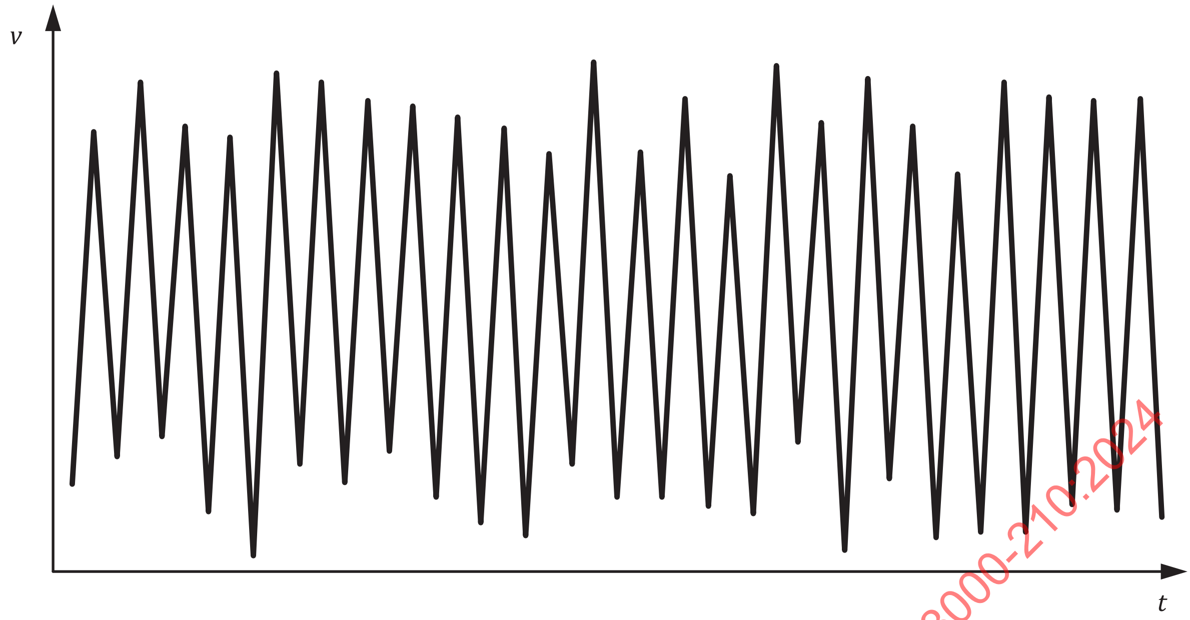
v value of interest the sensor is designed to capture

Figure 5 — Spike

EXAMPLE A humidity sensor captures some data values that are dramatically lower than the true value because of a network communication issue.

5.2.6 Noise

Noise is an anomaly in which captured data values fluctuate within the specified range for the sensor because meaningless content appears in the data set (see Figure 6).



Key

t time at which the sensor captures the value of interest (*v*)

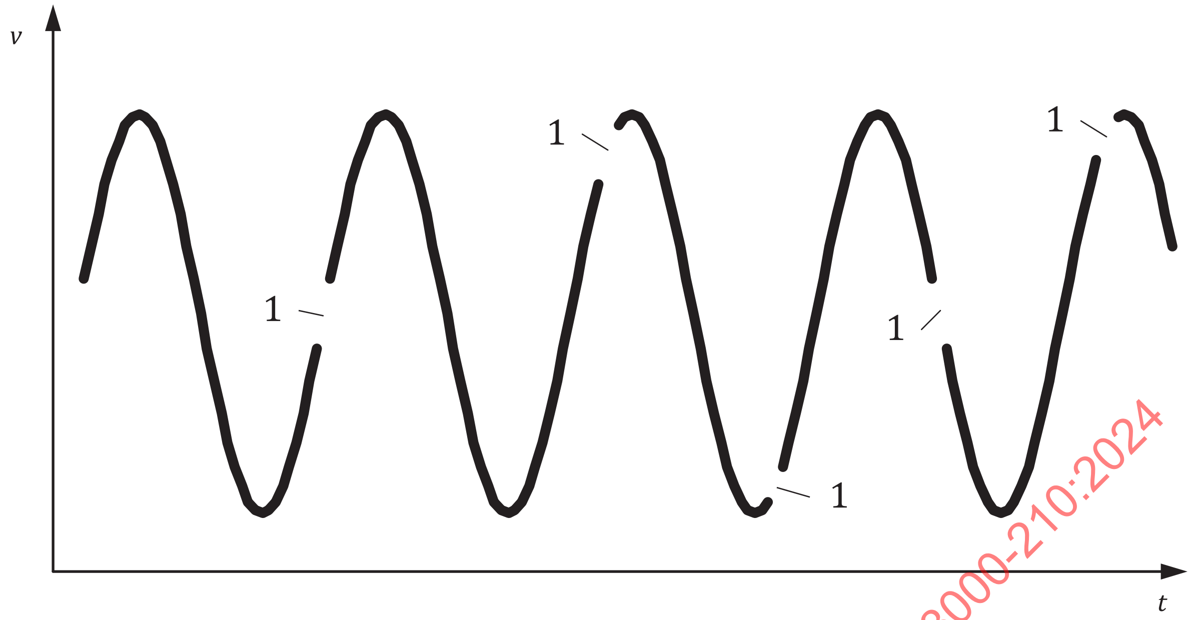
v value of interest the sensor is designed to capture

Figure 6 — Noise

EXAMPLE A sensor captures data about a cold air drainage system. The data exhibit an unexpectedly high amount of variation because of a hardware failure.

5.2.7 Data loss

Data loss is an anomaly in which sensor data includes periods where the data values are missing or have a default value (for example, zero) that is generally indicative of a failure (see Figure 7).



Key

1 gap where sensor fails to capture expected data values

v value of interest the sensor is designed to capture

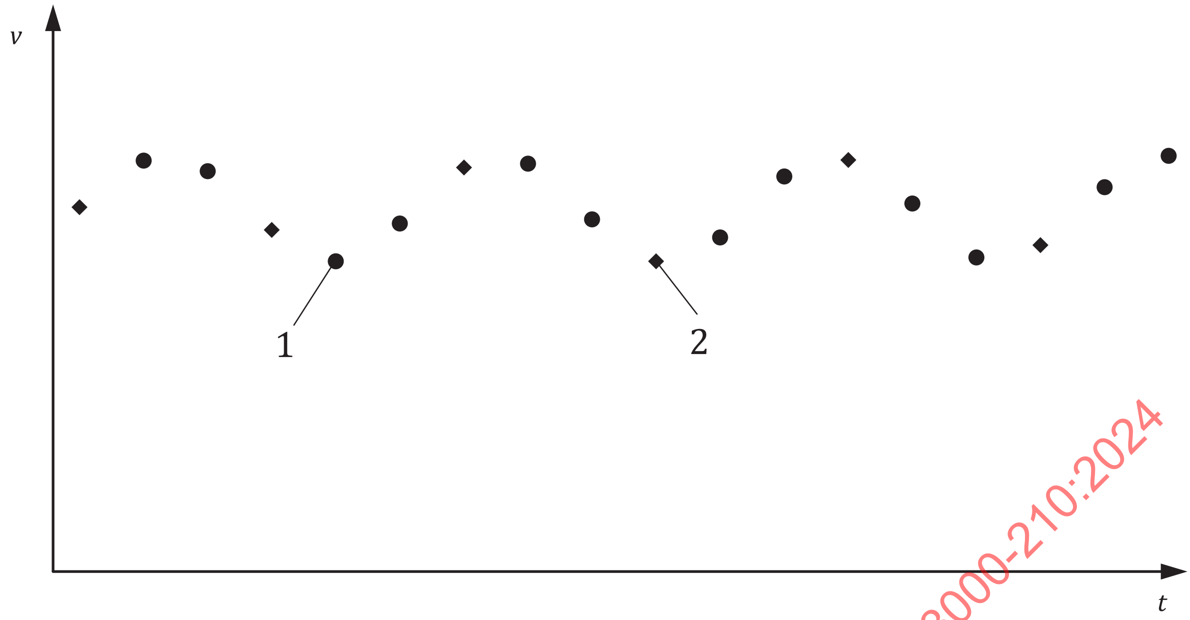
t time at which the sensor captures the value of interest (*v*)

Figure 7 — Data loss

EXAMPLE A sensor normally captures data values for water level every 10 min but the set is missing some values because of connectivity problems in the sensor network.

5.2.8 Lack of amount

Lack of amount is an anomaly in which, for a period of time, a sensor does not capture a sufficient quantity of data to meet the applicable requirement (see Figure 8).



Key

1 expected data value not captured by the sensor and, thus, total data volume is insufficient

2 data values captured by the sensor

t time at which the sensor captures the value of interest (*v*)

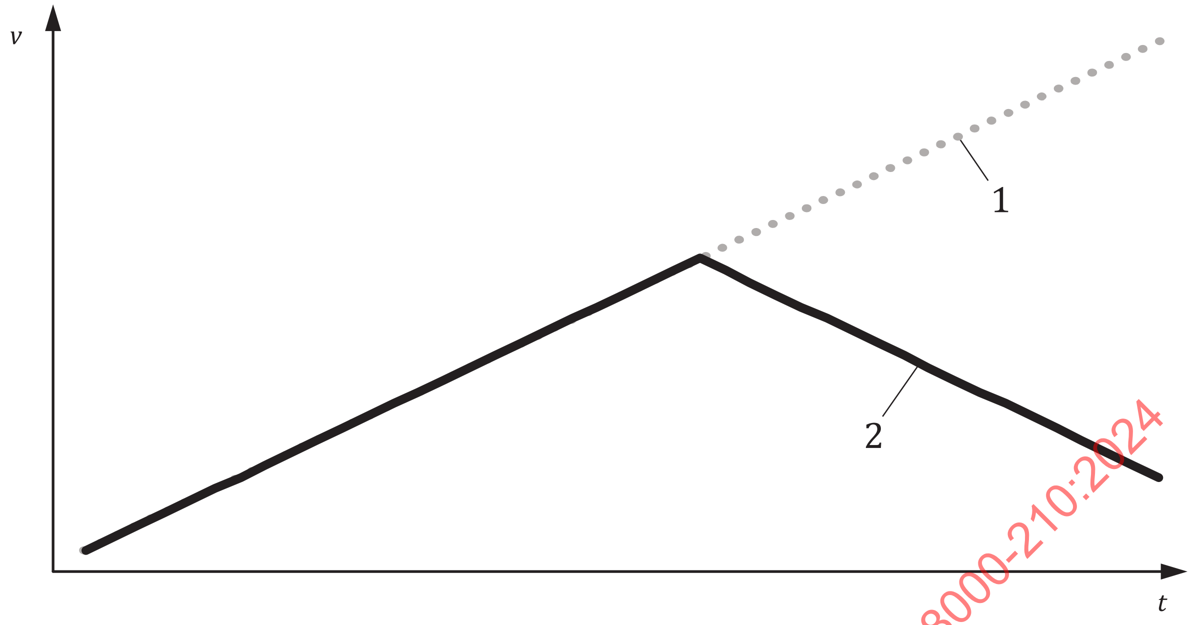
v value of interest the sensor is designed to capture

Figure 8 — Lack of amount

EXAMPLE To enable prediction of defects in products, the appropriate sensors capture data (including vibrations, speeds and pressure) from a facility machine. These predictions require each sensor to capture an appropriate minimum quantity of data.

5.2.9 Shift

Shift is an anomaly in which sensor data have a trend or rate of change that violate applicable requirements (see Figure 9).



Key

1 expected trend of the captured data values

2 actual trend of the captured data values

t time at which the sensor captures the value of interest (*v*)

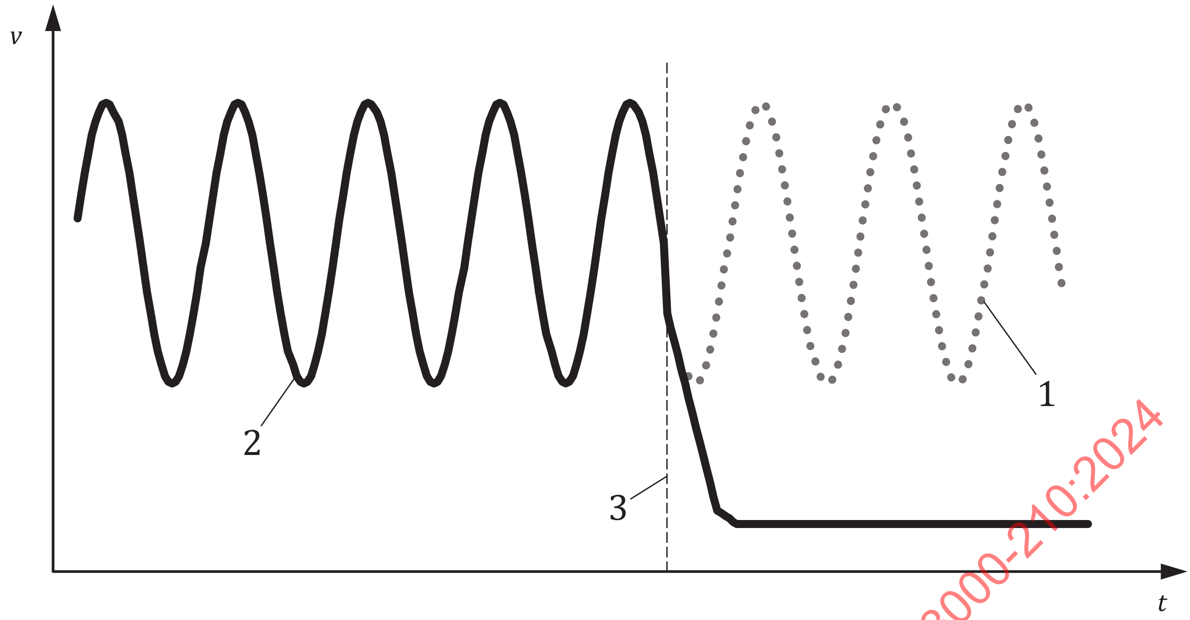
v value of interest the sensor is designed to capture

Figure 9 — Shift

EXAMPLE A geothermal sensor functions correctly by capturing data values that show a consistent pattern of temperature change from morning to afternoon on sunny days.

5.2.10 Drop or rise

Drop or rise is an anomaly in which captured sensor data experience a drop or rise before or after having been in a normal pattern (see Figure 10).



Key

1 expected data values

2 data values captured by the sensor

3 the sensor fails and only captures values as a drop from expected data values

t time at which the sensor captures the value of interest (*v*)

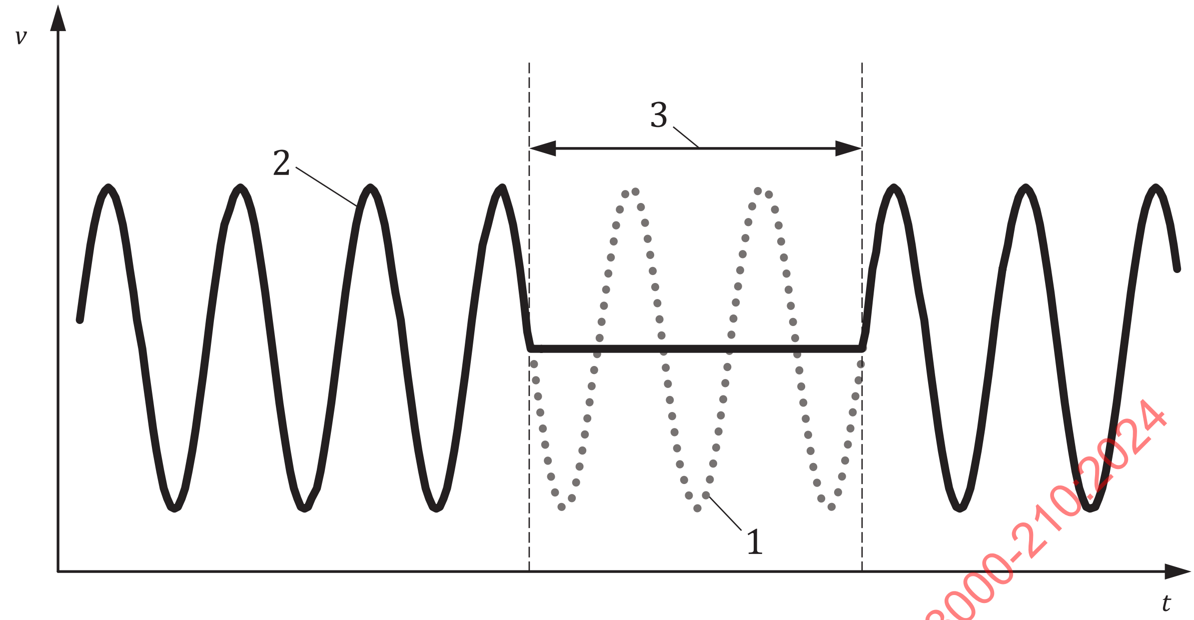
v value of interest the sensor is designed to capture

Figure 10 — Drop or rise

EXAMPLE For a wireless sensor, the battery requires periodic maintenance to prevent the sensor from malfunctioning and capturing data values that are inappropriately high or low.

5.2.11 Stuck

Stuck is an anomaly in which a series of data values show zero or almost zero variation for a longer period of time than is the reality for the true values being captured by a sensor (see Figure 11).



Key

1 expected data values

2 data values captured by the sensor

3 period during which there is continuous zero variation in the captured data values

t time at which the sensor captures the value of interest (*v*)

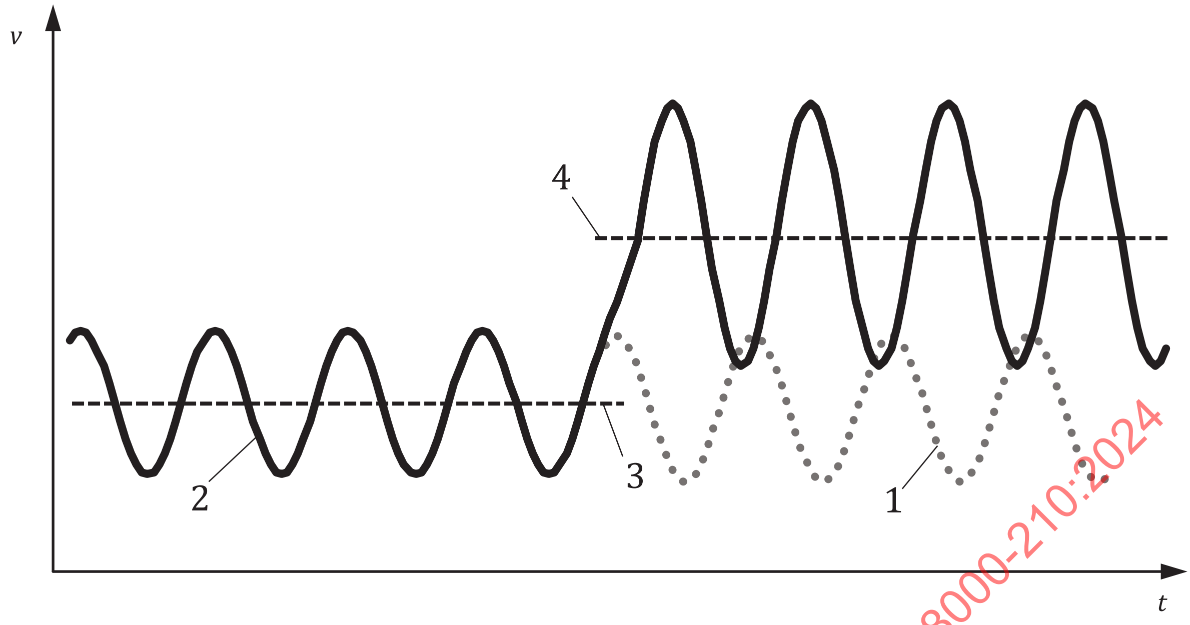
v value of interest the sensor is designed to capture

Figure 11 — Stuck

EXAMPLE Snow prevents a sensor from capturing wind speed data for a period of time. The data, thus, include values having no variation during that period.

5.2.12 Bound oscillation

Bound oscillation is an anomaly in which a set of sensor data has a range that changes for the whole set (see Figure 12).



Key

1 expected data values

2 data values captured by the sensor

3 mid-line of the captured data before bound oscillation occurs

4 mid-line of the captured data after bound oscillation occurs

t time at which the sensor captures the value of interest (*v*)

v value of interest the sensor is designed to capture

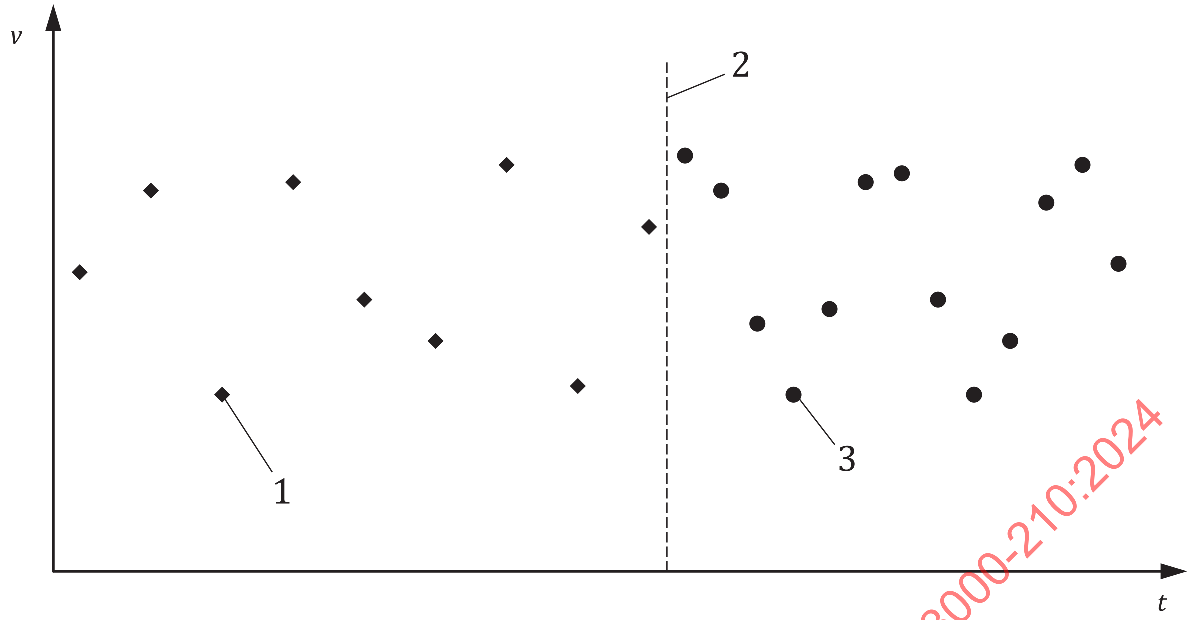
Figure 12 — Bound oscillation

EXAMPLE An acoustic sensor has captured data within the range 1 to 3. The sensor subsequently captures data in the range of 4 to 8 after a change to the position of the sensor.

5.2.13 Inconsistent frequency

Inconsistent frequency is an anomaly in which the frequency of measurement changes, fails to meet applicable requirements or varies continuously (see Figure 13).

NOTE This anomaly typically occurs when an update to the system configuration changes measurement frequency. This frequency affects how much data is captured by a sensor over a period of time.



Key

1 data captured at original frequency by the sensor

2 data capture frequency change due to configuration change

3 data captured at new frequency by the sensor

t time at which the sensor captures the value of interest (*v*)

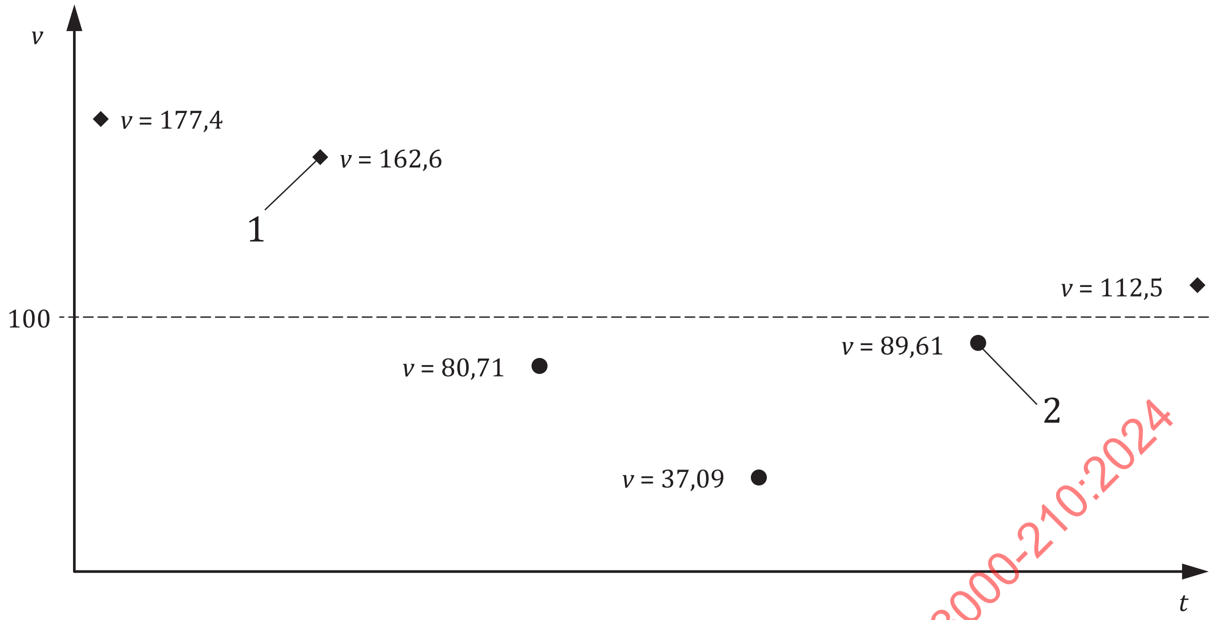
v value of interest the sensor is designed to capture

Figure 13 — Inconsistent frequency

EXAMPLE Originally, a weathervane captures data values at six times per minute. The vane captures data at 60 times per minute after a system configuration change.

5.2.14 Different resolution

Different resolution is an anomaly in which sensor data includes values that have differing numbers of decimal places because the sensor can only capture a particular number of significant figures (see Figure 14).



Key

1 data captured by sensor at resolution of 4 significant figures with 1 decimal places

2 data captured by sensor at resolution of 4 significant figures with 2 decimal places

t time at which the sensor captures the value of interest (*v*)

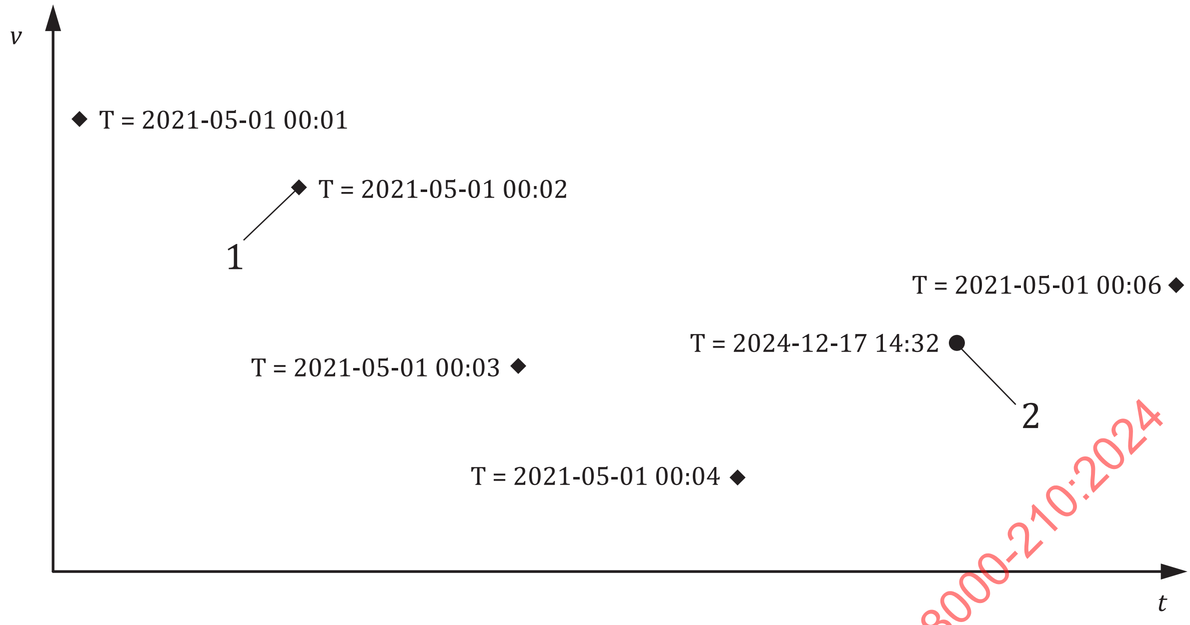
v value of interest the sensor is designed to capture

Figure 14 — Different resolution

EXAMPLE For an atmospheric pressure sensor having a 5-digit limit on significant figures, the sensor captures the true values “10,123 4”, “89,665 4” and “100,123 4” as the data values “10,123”, “89,665” and “100,12”, respectively. These data values have an anomaly where those values less than 100 have three decimal places but values greater than or equal to 100 have only two decimal places.

5.2.15 Incorrect timestamp

Incorrect timestamp is an anomaly in which sensor data include a timestamp that does not represent the true value of the time at which the sensor captured the data (see Figure 15).



Key

1 data captured with correct timestamp recorded by sensor

2 data captured with incorrect timestamp recorded by sensor

t time at which the sensor captures the value of interest (*v*)

v value of interest the sensor is designed to capture

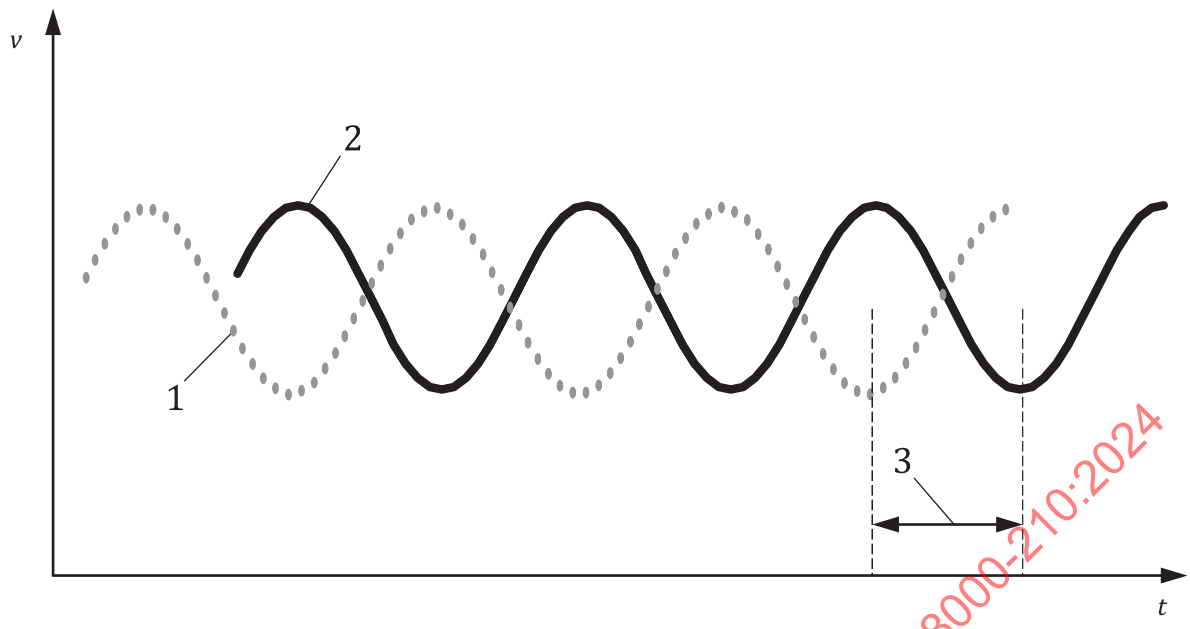
T timestamp recorded by sensor

Figure 15 — Incorrect timestamp

EXAMPLE A sensor captures data every minute. The data include a sequence of timestamps “2021-05-01 00:00”, “2021-05-01 00:01”, “2021-05-01 00:02”, “2021-05-01 00:03”, “2224-12-01 12:30” and “2021-05-01 00:05”. The timestamp “2224-12-01 12:30” is an anomaly.

5.2.16 Latency

Latency is an anomaly in which a sensor delivers data more slowly than specified by the applicable requirement (see Figure 16).



Key

1 expected data values

2 data values captured by the sensor

3 delay between value of interest occurring and the timestamp recorded by the sensor

t time at which the sensor captures the value of interest (*v*)

v value of interest the sensor is designed to capture

Figure 16 — Latency

EXAMPLE A heartbeat sensor delivers data with a delay of 1 s, when the requirement is for a delay of only 10 ms.

5.3 Data anomalies for collections of multiple sensors

5.3.1 General

The following are typical data anomalies appearing in the correspondence between data sets captured by multiple related sensors:

— dissimilarity (see 5.3.2);

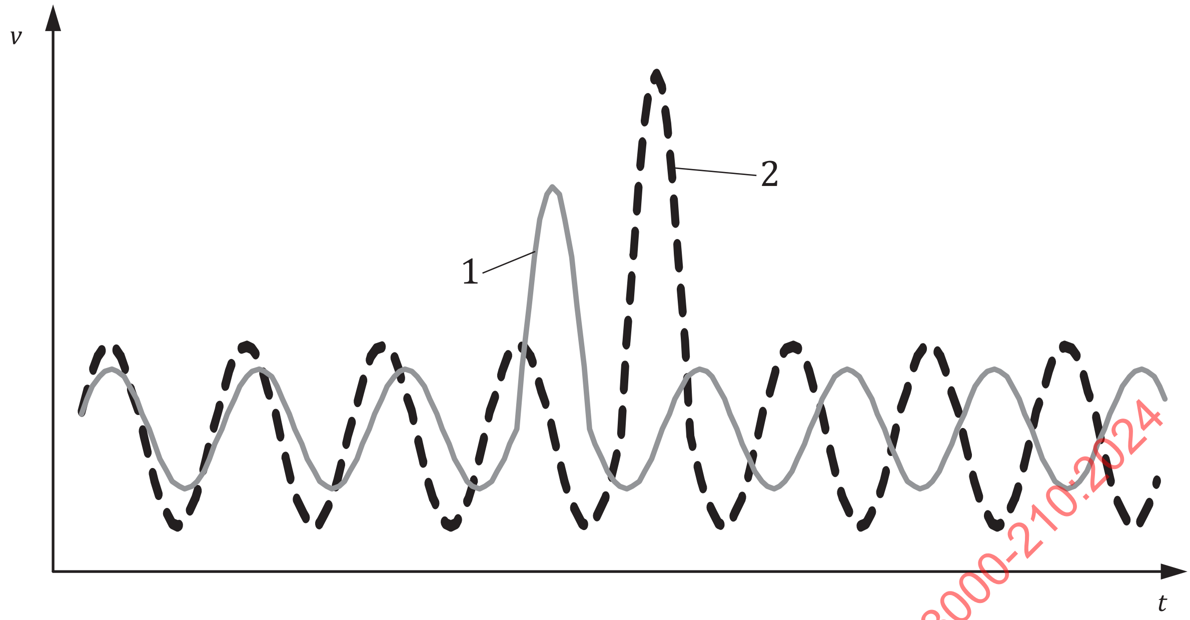
— rule violation (see 5.3.3);

— inconsistent timestamp (see 5.3.4).

5.3.2 Dissimilarity

Dissimilarity is an anomaly in which sets of sensor data contain values having significant differences when those values are captured by two or more sensors and the values are supposedly representing the same true value (see Figure 17).

NOTE This anomaly occurs due to reasons including differences in units of measurement, conditions and scales between two data sets.



Key

1 data values captured by sensor A

2 data values captured by sensor B

t time at which the sensor captures the value of interest (*v*)

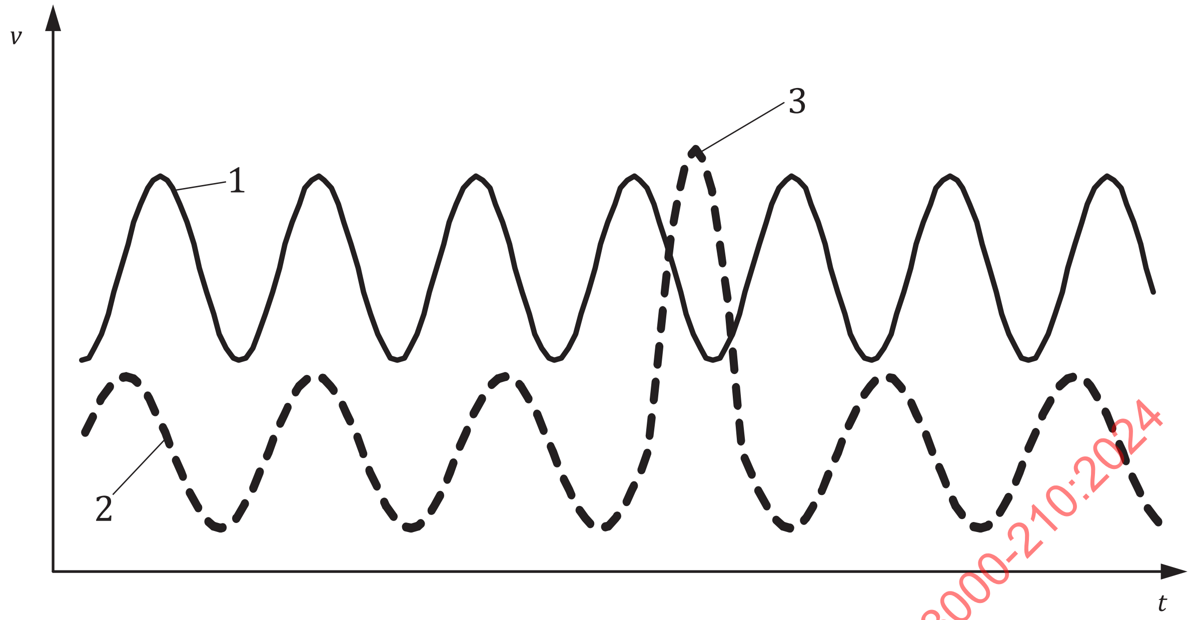
v value of interest the sensor is designed to capture

Figure 17 — Dissimilarity

EXAMPLE Two sensors capture the temperature of the same object. If the two sensors do not generate data values that are within a tolerance specified by the applicable requirement, then an anomaly has occurred.

5.3.3 Rule violation

Rule violation is an anomaly in which sensor data sets violate predetermined relationships or rules between those data sets (see Figure 18).



Key

1 data values captured by sensor A

2 data values captured by sensor B

3 violation of rules between the two sensors

t time at which the sensor captures the value of interest (*v*)

v value of interest the sensor is designed to capture

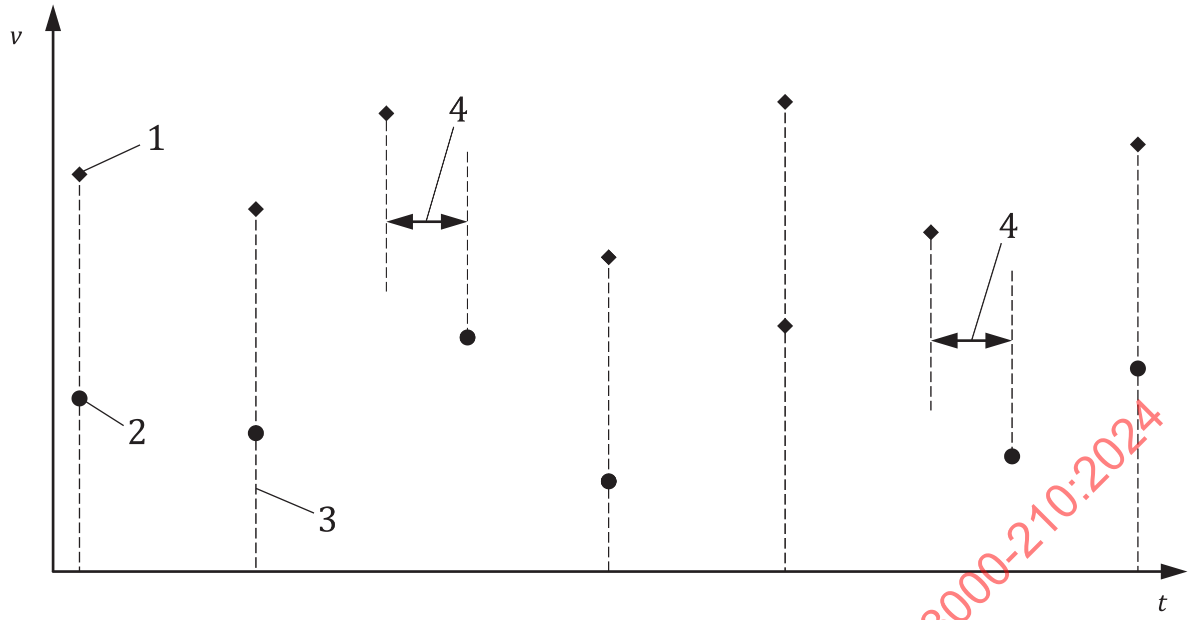
Figure 18 — Rule violation

EXAMPLE In a cooling system with two sensors, one measuring high temperature and the other low, the value captured by the low-temperature sensor is an anomaly if this value is higher than the one captured by the high temperature sensor at the same moment in time.

5.3.4 Inconsistent timestamp

Inconsistent timestamp is an anomaly in which the time information differs between data sets captured by sensors that are supposedly working in synchronization with each other (see Figure 19).

NOTE This anomaly occurs due to various reasons including instability of communication between sensors and inconsistency of the time zones of different sensors.



Key

1 data values captured by sensor A

2 data values captured by sensor B

3 same timestamps for both sensors

4 inconsistent timestamps between the two sensors

t time at which the sensor captures the value of interest (*v*)

v value of interest the sensor is designed to capture

Figure 19 — Inconsistent timestamp

EXAMPLE Two temperature sensors are capturing data at different altitudes. There are times when the timestamps of the two sensors do not coincide, reducing the consistency of the relationship between the two data sets.

6 Relationship between quality characteristics and data anomalies

Different data anomalies affect different data quality characteristics (see Table 6), although in some circumstances the effect can vary from that indicated by the table.

Table 6 — Relationship between quality characteristics and data anomalies

Data anomaly		Quality characteristic of sensor data				
		Accuracy	**Complete-ness**	**Consistency**	**Precision**	**Timeliness**
Individual sensors	Offset	A	-	-	-	-
	Drift	A	-	-	-	-
	Trim	A	-	-	-	-
	Spike	A	-	-	A	-
	Noise	A	-	-	A	-
	Data loss	-	A	-	-	-
	Lack of amount	-	A	-	-	-
	Shift	-	-	A	-	-
	Drop or rise	-	-	A	-	-
	Stuck	-	-	A	-	-
	Bound oscillation	-	-	A	-	-
	Inconsistent frequency	-	-	A	-	-
	Different resolution	-	-	-	A	-
	Incorrect timestamp	-	-	-	-	A
	Latency	-	-	-	-	A
Collections of multiple sensors	Dissimilarity	-	-	A	-	-
	Rule violation	-	-	A	-	-
	Inconsistent timestamp	-	-	-	-	A

Key

A data anomaly affects the indicated data quality characteristic

\- data anomaly does not affect the indicated data quality characteristic

7 Application of quality characteristics of sensor data

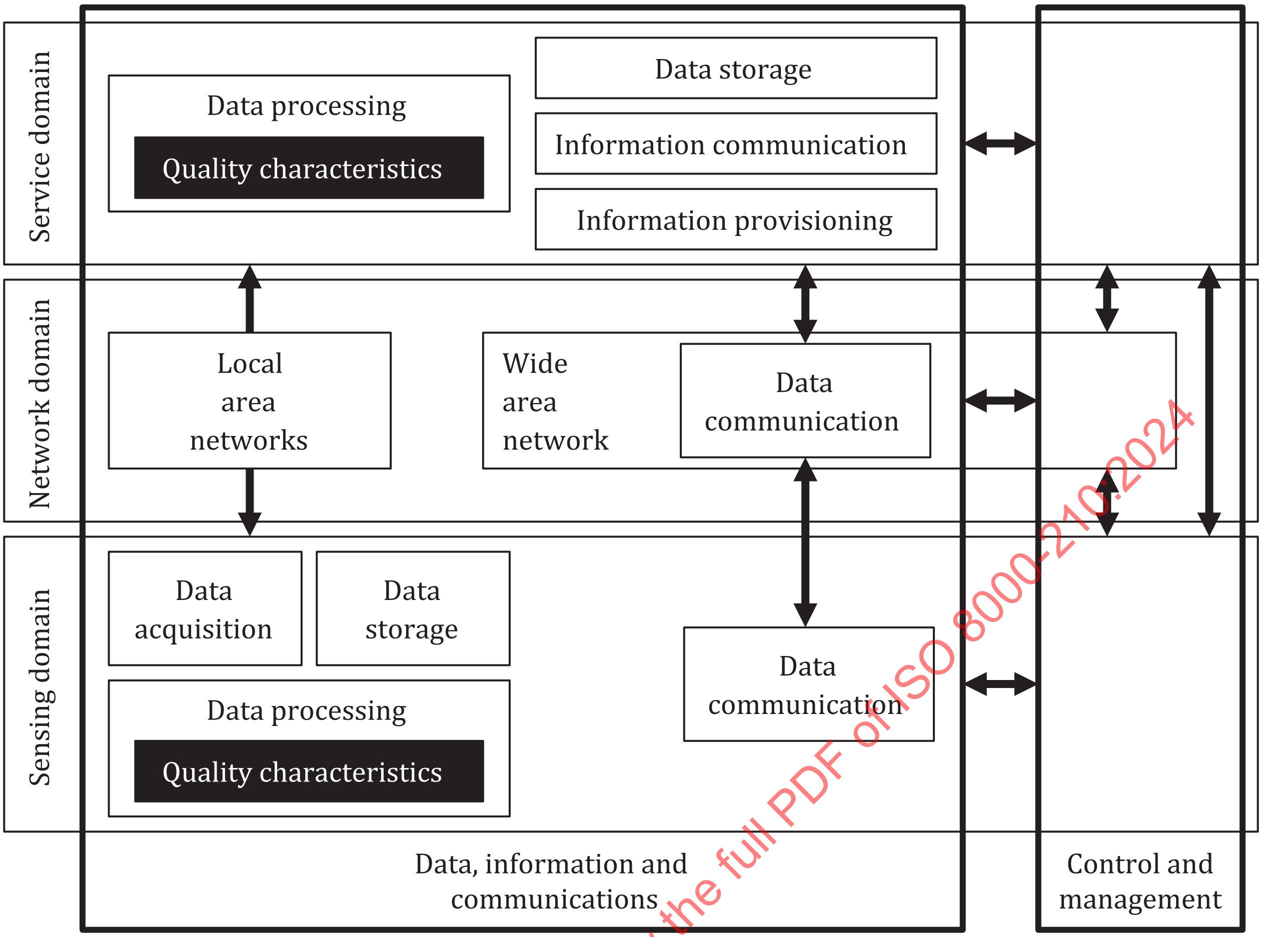
Sensor data are produced by one or more sensors connected to a sensor network or the Internet of Things (IoT) (see ISO/IEC 30141). Such networks (including the IoT) involve a set of data-related functions including acquiring, transforming, transmitting, storing and analysing data. These functions are more effective if data have appropriate levels of quality.

In any of the following scenarios, an organization shall apply and record the use of the quality characteristics specified by this document as the basis for measuring and improving the quality of the sensor data involved:

— data processing in sensor networks in either the service domain or the sensing domain (see Figure 20);

— analytic services in the application and service domain within the IoT (see Figure 21);

— data manipulation in condition monitoring and diagnostics of machines (see Figure 22).



NOTE To explain the role of sensor data quality characteristics in sensor networks, this figure uses the functional architecture specified by ISO/IEC 29182-3.

Figure 20 — Role of sensor data quality characteristics in sensor networks