22DIT02 FunSNM

D2: Guidance document on data quality metrics in sensor networks, including measurement uncertainties, common factors that influence data quality, assessment of the uncertainty-aware sensor fusion techniques and traceability to the SI system of units

Lead participant for the deliverable: FORCE Technology

Due date: 31-08-2024

Actual submission date: 31-08-2024

Confidentiality Status: PU - Public, fully open

Deliverable Cover Sheet

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

D2: Guidelines for data quality, measurement uncertainty, and traceability in sensor networks

Guidance document on data quality metrics in sensor networks, including measurement uncertainty, common factors that influence data quality, assessment of uncertainty-aware sensor fusion techniques and traceability to the SI system of units.

Prepared by:

Mads Johansen, Michael Vaa – FORCE Technology

Shahin Tabandeh, Henrik Söderblom – VTT

Peter Harris, Jonathan Pearce, Yuhui Luo, Declan Tucker – NPL

Anupam P. Vedurmudi, Maximilian Gruber – PTB

Maitane Iturrate-Garcia – METAS

Martha Arbayani Zaidan – UH

Miloš Davidović – VINS

Alexander Holtwerth, Jan Stock, André Xhonneux – FZJ

Gertjan Kok, Marcel van Dijk – VSL

Carlos Pires, João A. Sousa – IPQ

Acknowledgement: The project 22DIT02 FunSNM has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States

Funder name: European Partnership on Metrology

Funder ID: 10.13039/100019599

Grant number: 22DIT02 FunSNM

Contents

1 Introduction

As sensor networks become easier to acquire and deploy, and consequently are more common in almost all industries and in everyday life, so ensuring the trustworthiness and reliability of measurements and data in such systems becomes more challenging. Not only as the numbers of sensors grow, but also as the inaccessibility of sensors means it is infeasible to use established methods for their calibration, so the difficulties of assessing measurement uncertainty in sensor networks and establishing the traceability of measurements made by such systems increases. Furthermore, due to the large volumes of data, it is a challenge to validate the quality of data collected from sensor networks, and it is infeasible to do so without automated, efficient, and reliable methods.

The purpose of this guide is to help address the challenge of ensuring data quality for sensor networks. It is structured in two main parts, one related to data quality metrics and one to traceability.

The part on data quality metrics (Section 2) provides guidance on the importance of data quality when collecting large amounts of data from sensor networks where there is less control over the sensor environment as well as the management and architecture of the sensor network, for example, compared to a laboratory setup. This includes choosing which dimensions of data quality are most important depending on the use case, managing data requirements during the lifecycle of sensor nodes, and developing ways to measure and quantify data quality.

Different use cases have different metrological needs when it comes to traceability. The part on traceability (Section 3) addresses SI-traceability in sensor networks, providing guidance on different ways of calibrating sensors in sensor networks such as in-situ, self- and co-calibration. Furthermore, it addresses the challenge of making methods of analyzing sensor data uncertainty-aware, for example, for sensor fusion, and using different modelling techniques, for example, digital shadows and digital twins.

Different use cases are used as examples in different sections of the guide. The use cases are district heating networks, heat treatment of high-value components in advanced manufacturing, gas flow meter networks, air quality monitoring sensor networks and smart buildings. These are used to highlight certain challenges, needs, and both commonalities and differences in certain types of sensor networks within the different subjects covered in the guide.

2 Data Quality Metrics in Sensor Networks

2.1 Data quality metrics for sensor networks

This section of the guide discusses approaches to evaluate the quality of a sensor network dataset. Evaluating application quality is important to allow the development of sensor applications that are resilient and robust, and to make sound decisions based on the underlying sensor data quality. Several data quality dimensions have been identified, whereby different aspects of data quality, such as accuracy, consistency, and completeness, are measured. Sets of data quality dimensions can be used, thus, to assess the quality of a dataset. When designing a sensor network, it is crucial to keep the dimensions of data quality in mind and potentially adjust the design of the sensor network to optimize the quality of the data. Indeed, the higher the quality of the data, the more reliable the results of further analyses will be.

An overview of the data quality metrics identified after performing a literature review is included in the following subsection. Only the metrics considered most relevant for sensor networks by the authors of this report are described. Hence, the list of metrics provided is by no means exhaustive. For example, in [1] many more data quality metrics are considered besides those listed in this report such as drop rate, accessibility, compliance, etc. Furthermore, examples and data quality metrics important for the project use cases are also included.

2.1.1 Relevant data quality metrics from the literature

2.1.1.1 Consistency

As defined in [2], consistency refers to the absence of apparent contradictions in a database. Consistency is a measure of the internal validity of a database and is assessed using information that is contained within the database.

Consistency metrics help assess whether the values in a data set are consistent with the values previously recorded and stored. Consistency allows the improvement of data quality by ensuring all data remains constant. One important consistency measure is date consistency, which measures how many dates in a data set fall outside of their historical range, i.e. the time interval for which data have been measured/collected. Numeric consistency can tell how many values in a data set differ from the expected range.

In [3] consistency is described in terms of constraints, as the degree to which defined constraints are adhered. These constraints can be used, for example, to check whether a value is within a specific range or if values fall inside logical bounds (such as a humidity sensor that should provide only positive values). Consistency can be quantified by the percentage of values in the dataset that satisfy the defined constraints.

2.1.1.2 Accuracy

[4] describes accuracy as the degree to which data correctly represents the true value. In practice, the true value is often unknown. Hence, it is often required to determine a reference value in order to assess the accuracy of the data. The reference value can for example be obtained from a reference sensor, or from an aggregation of multiple sensors that measure the same quantity.

[5] quantifies accuracy simply by the fraction of fields judged "correct" (or the fraction of "correct" records, in case each datapoint contains multiple fields). Correct fields can for example be defined according to a rule-based evaluation, where a field is deemed correct if the difference with the reference is small enough. See for example [6].

In [7] and [8], accuracy of a measurement value v is defined as the maximum absolute value a, such that the real value (reference value) lies in the interval $[v - a, v + a]$. Note that this is an absolute metric, rather than a relative metric.

A general approach is described in [9], where the accuracy is calculated according to the distance between the measured quantities and the reference quantities $D(v, v')$. The distance function is zero in case of $v = v'$ and positive otherwise. The metric is defined as

$$\text{accuracy} = \frac{1}{1 + D(v, v')}.$$

In this definition, the perfect score for accuracy equals 1 and the closer the metric is to 0, the worse the accuracy of the data is.

2.1.1.3 Completeness

[10] describes completeness as referring to whether all required data is present, that is, whether any data required to deem the dataset fit-for-purpose is missing or not. Completeness of data ensures that all the information needed to run quality analytics and artificial intelligence (AI) exists. Typically, completeness is quantified as the ratio of missing values compared to the total number of values at the dataset, column (i.e. attributes) or record levels; [3], [7], [8].

2.1.1.4 Auditability

In [4], auditability (in this paper referred to as traceability) is defined as "the degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use". In other words, auditability allows to trace back where the data comes from and to track how the data has been changed over time. Having the appropriate metadata, for example, can provide auditability to a dataset. A metric to quantify auditability is the percentage of data that cannot be traced, see [11].

2.1.1.5 Timeliness

In [7] timeliness was interpreted in the context of sensor data streaming applications, as the difference between its recording timestamp and the current system time. In contrast to other data quality dimensions, timeliness takes an exceptional position as it can be calculated at runtime and must not be recorded, propagated, and processed during the data processing. In practical applications timeliness needs to be defined in a contextual manner [8] as the punctuality requirements of data depend on the task at hand.

In [1] other time-related dimensions include currency and volatility. Currency focuses on how quickly the corresponding data are updated when they occur in the real world, and volatility indicates how often the data changes over time. [6] describes volatility as the length of time for which data remains valid, whereas currency is defined as

$$\text{currency} = (t_{\text{real}} - t_{\text{ideal}}) + (t_{\text{arrive}} - t_{\text{ideal}}),$$

where t_{ideal} is the ideal sampling time, t_{real} is the actual sampling time, and t_{arrive} is the time needed to record the data.

2.1.1.6 Uniqueness

Uniqueness of a dataset refers to the absence of duplicates in a dataset. Hence, uniqueness can be measured by the number of duplicates present in a dataset. [12] quantifies uniqueness as the percentage of records having a unique primary key. Unique primary key refers to the value of a row/record in a dataset making that row unique. Duplicates of primary keys are undesirable, but some measurements can be repeated if the state doesn't change, although multiple repetitions can indicate a stuck sensor. An example of a primary key could be a timestamp in a timeseries dataset.

2.1.1.7 Correctness

Correctness refers to how well the data values correspond to actual values. It can be separated into two aspects: semantic correctness and syntactic correctness. In the context of metrology, semantic correctness is about conformity between measured values and the actual values, i.e., correctness of the content, like accuracy. For instance, validating a dataset for semantic correctness can involve checking if the physical quantities (length, mass, etc.) in a particular dataset deviate from the expected values. Syntactic correctness is about the correctness of the form or structure. It could, for example, be data format and units or dimensions. The digital calibration certificate, for instance, requires calibration data to be presented in a specified machine-readable format [13]. Adhering to this format can be considered a form of syntactic correctness.

2.1.1.8 Reusability

Reusability refers to the data being understandable and useful to others. It includes comprehensibility and consistency. Comprehensibility refers to the quality and existence of metadata and being readable and uniformly represented. Consistency meaning the absence of contradictions in the data and with referential integrity between the data and any metadata or other reference data and standards.

Moreover, reusability is one of the foundational FAIR data principles. FAIR stands for Findability, Accessibility, Interoperability, and Reuse. The principles emphasize machine readability of data because humans are relying more and more on computers for handling data, due to increase in speed of data generation, as well as increase in volume and complexity.

2.1.1.9 Redundancy

In [14] several metrics are discussed to evaluate the metrological redundancy in a sensor network. It is argued that a higher degree of redundancy is desired, as it makes the network resistant to sensor failures. Three different types of metrological redundancy are discussed: sensor replication, sensor relevance, and network redundancy. For each type of redundancy, multiple metrics are proposed.

Sensor replication considers the relationship between sensors: can readings of one sensor be derived from readings of other sensors? If there is sensor replication in the network, sensor readings can be aggregated, leading to data reduction and smaller measurement uncertainty. Three metrics are proposed to evaluate sensor replication:

- Rank (the number of linearly independent rows or columns of the matrix): the rank of the matrix containing all sensor signals provides a discrete scale between zero and the number of signals on the replication of the sensor signals. A rank equal to the number of sensor signals indicates no replication between the sensor signals.
- Condition number: the condition number of the matrix containing all sensor signals quantifies the linear dependency, and thus the replication, between the different signals. A condition number equal to 1 indicates orthogonality (so no replication between the sensor signals) and a condition number equal to infinity indicates linear dependency (so exact replication between the sensor signals). The user does need to decide on a matrix norm. A common choice is the 2-norm, but also the 1-norm and inf-norm are possible choices. The condition number under the inf-norm is typically larger compared to the 2-norm, but is also evaluated more easily.
- Cluster: sensor replication can also be evaluated by clustering the sensor time series, using a measure of 'distance' between two time series. The silhouette scores of the sensors can be calculated and the number of sensors with a score of, for example, 0.5 or 0.7 or any other appropriate value between 0 and 1, can be determined. The metric can be defined as the fraction of sensors with silhouette score higher than the threshold. A higher fraction indicates more replication between the sensors. Special consideration should be given to the construction of the clusters, as this will influence the metric.

Sensor relevance quantifies whether a sensor is relevant with respect to the measurand(s) and is determined for each sensor individually. If a sensor is irrelevant, it can be removed from the network without changing the estimate and uncertainty of the measurand(s). Two metrics are proposed to evaluate sensor relevance:

- Sensitivity coefficients: the sensitivity coefficient c_i can provide information about the relevance of sensor i. A discrete metric would deem sensor i relevant when $c_i \neq 0$. A continuous metric determines the proportion of the uncertainty of the measurand y that comes from the data x_i of sensor i: $|c_i|u(x_i)/u(y)$. The higher this value, the more relevant this sensor is. The sensitivity coefficients can be determined either algebraically or numerically, depending on the availability and form of the model function. The authors concluded that this metric is particularly suited for small models that require a small number of input data, as it is difficult to interpret the results if there are too many sensor values influencing a single output value.
- Pearson's correlation: sensor relevance can also be evaluated by Pearson's correlation coefficient between the known reference or training values of the measurand(s) and (a feature of) the sensor data. The higher the absolute value of the correlation coefficient is, the more relevant the sensor is. The Pearson correlation coefficient is the ratio between the covariance of two variables and the product of their standard deviations and measures the linear dependency between two sets of data. If a sensor or a derived feature of the sensor data is relevant for the measurand in a non-linear way, this metric may not detect this.

There is redundancy in a network if the measurand can be determined from different subsets of sensors. This requires that there is sensor replication of the relevant sensors. The paper provides two metrics to quantify network redundancy:

- Excess sensors: network redundancy can be assessed by the number of excess sensors present in the network in addition to the minimum required sensors necessary to determine the value of the measurand. The more excess sensors, the higher the network redundancy.
- Uncertainty increase: another way to assess network redundancy can be by determining the maximum increase in uncertainty of the measurand, after removing m relevant sensors from the network. The lower this value, the less the network relies on one or a few sensors, so the higher the degree of redundancy.

2.1.2 Data quality metric examples of the project use cases

This section considers different metrics for quality, divided in three subsections below, and how they relate to the different use cases used as examples throughout the guide. Not all use cases are exemplified in each of the categories.

2.1.2.1 General metrics

To measure the relevance and usefulness of sensor network data, different general metrics can be used, such as consistency, accuracy, completeness, auditability, integrity, timeliness, uniqueness, and cost. Some of these metrics have been described in detail above. Depending on the use case some metrics might be more useful than others.

As an example for the district heating use case, completeness and time resolution are the most important metrics, because when data is incomplete, it becomes difficult to analyze, due to correlation between flow and temperature at sensor points. Additionally, integrity is also an important metric, as the sensor nodes are located inside the private property, and it is difficult to access and verify if the installation is correct and no strange effects are present.

For the other use cases, accuracy, traceability, completeness and timeliness are referred as the most important metrics.

2.1.2.2 Metrics for number and quality of sensors

Different metrics can be used as a basis for optimizing the design of a sensor network in terms of number and quality of individual sensors. As mentioned before, depending on the use case and type of network some metrics are more relevant than others.

For the use case on district heating, information on accuracy is used, as well as number of heat meters and distance (pipe length) between the heat meters. The network model is also important, and the quality of this model can vary, for example, if individual coordinates on strings can properly fit each other at the connected end point. Additionally, temperature measurements from sensors in the network are compared to high-end, calibrated sensors.

For environmental monitoring the exploratory studies show that if the objective is monitoring, at least one most-exposed site and one least-exposed site, must have sensors, and additional sensors depend on population density, whereas if the objective is characterization, one sensor is placed close to the source and several sensors are installed to study the decay of the source.

For smart buildings, sensors are placed at each window/door/heater, so the number of sensors depends on the number of windows, doors, and heaters. The number of rooms is defined in such a way, that results with respect to e.g., user behavior, are statistically representative for a building. The number and variety of buildings is selected in such a way, that the buildings are representative for German office buildings with construction years between 1970 and now. The quality of individual sensors corresponds to quality applied in normal offices. Reason for this is that the developed methods should not require high quality sensors, since this would be a strong limitation for large scale deployment.

Finally for industrial manufacturing, the main requirement is sufficient temperature sensors that adequately capture the range of temperatures arising from temperature gradients in the heat treatment furnace, and for natural gas transmission accuracy and precision are the key parameters.

2.1.2.3 Consistency and redundancy metrics

To describe the consistency of individual sensors and redundancy of sensor in networks different metrics are available but across the exemplified use cases similar approaches can be used. One approach is to a comparison with typical expected values, using a reference meter or traceable calibrated sensors to validate individual sensors. Another approach was to conduct a comparison with other sensors, in the same space, to detect inconsistencies in the measurements.

2.1.2.4 Shortcomings in current metrics

From a metrological perspective the different metrics mentioned in the above sections have certain shortcomings. The value of uncertainty is not always available, and there is often missing information related to the installation and initial calibration certificates. Also, the quality of deployed sensors is not always the best and there are no duplicate measurements available for verification.

2.2 Data requirements in a sensor network

This section of the guide discusses data requirements for sensor networks both for the complete lifecycle of a typical network node as well as for the complete sensor network. A typical sensor lifecycle may include steps of factory calibration, field calibration, development of multi-predictor calibration models (in case of multi-sensor nodes or soft sensor capabilities), deployment and continuous operation, periodic recalibration (e.g., machine learning (ML) retraining/continual learning), end-of-life or repurposing. Each of the stages have different data requirements and produces different insights into sensor performance from a metrological standpoint. Furthermore, extent of needed data is different for different sensor networks, influencing operational costs and loss benefits considerations. In the following subsections two use cases are used to describe data quality requirements.

Air quality sensor networks enable continuous monitoring for reporting and facilitating the creation of spatial interpolations for key air quality parameters. These networks can produce high-resolution maps of $PM_{2.5}$ concentrations, offering valuable insights into air pollution distribution. Pervasive air quality monitoring can be used in both fixed and mobile nodes. The primary aim is to generate high-resolution spatial and temporal data on air quality Low-cost sensor air quality monitoring network nodes can be calibrated effectively using a range of ML algorithms, from basic models to advanced techniques.

Use of sensor networks in industrial applications includes exploratory use of multi-wire thermocouples for possible in-situ drift detection. Each wire has a different Pt-Rh composition. Each pair of wires formed an individual thermocouple. In e.g., a 5-wire thermocouple, there are 10 possible pairs of wires and so 10 thermocouples., each with a different calibration drift rate. Approaches using data mining, physical modelling, and a combination of the two offer the possibility of deducing the calibration drift in-situ.

2.2.1 Data requirements expected during the lifecyle of a sensor node in a specific sensor network application

a) Because each phase of a sensor node lifecycle has different data requirements, it is important to understand and identify each of them. A typical lifecycle of a sensor node in an air quality sensor network includes the following phases: Integration of the sensor node: this phase is the node fabrication step in which transducers, analog front end, microcontroller boards, charging board, battery, and data transponders are connected and integrated with microcontroller firmware to become a full featured sensor node.

b) Characterization of sensor node response to chemical/physical pollutants in lab conditions: in this phase, information about sensor response model (linearity, sensitivity, limit of detection, response time, etc.) are collected. Typically, though they may depend to a certain extent from the specific sensor node, they are collected for a very limited number of devices (1-3) and generalized.

c) Pre-deployment for calibration data gatherings: In this phase the sensor is collocated in field conditions together with a reference analyzer to provide a reference dataset for data driven calibration derivation.

d) Calibration: it is the process of deriving a transducer model for translating raw data in actual pollutant amount of substance fraction or concentration data. This phase may include the factory calibration, field calibration (based on co-location with regulatory air quality monitoring stations) or the development of multi-predictor calibration models based on ML approaches. During operation, periodic field recalibration is recommended, which can be further used to determine the level of performance and/or stability over time.

e) Deployment: in this phase the sensor node is installed in operational conditions for fulfilling actual data gathering.

f) Data gathering and processing: this phase is the main part of the operative deployment and represents the actual data gathering, processing and transmission of the quantity measured data to operative control.

Many of these lifecycle phases will be part of other sensor networks used in different applications, such as industrial applications (see the brief description of the use case of sensor networks for industrial applications included in the introduction of this section).

Data requirements expected during the lifecycle of a sensor node in a specific application of a sensor network. Factory calibration for air quality sensor node(s) (particularly, low-cost air quality sensor node that estimates particulate matter concentration, or gaseous pollutant amount of substance fractions) may consist of a simple zero check, such as one data point check or similar. This step usually has very limited data requirements, usually only a few data points.

Field calibration serves to test the sensor in real word operating conditions. For air quality sensor field calibration, is usually conducted by collocating sensor nodes with reference instruments (e.g., regulatory automatic monitoring station), and simultaneously collecting data from sensor nodes and reference instruments. Collocation typically lasts few weeks (about three weeks is the recommended period in the literature), in order to capture enough dynamic range of the air quality parameters of interest. After a four or five-deployment period of the sensor node, it is recommended its recalibration by collocation with the regulatory reference instruments (maximum every six months during three weeks). Data requirements are the following: both the data from the sensor node being calibrated and the data from a reference are required to compute the calibration model. According to the literature, the most commonly used is a simple linear regression model. Data requirements for more complex calibration models could be more demanding. Quality of calibration is typically reported via the coefficient of determination (r^2), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). The calibration models are considered valid if the sensor node is used in conditions that are similar to the conditions in which the sensor was calibrated.

Use-case of air quality sensor network: In this application, when taking the entire lifecycle of a sensor node into consideration, requirements are linked to the characterization of the sensor node response including.:

a) Full response to at least the target pollutant(s) and primary interferents (i.e., forcers, such as CO and temperature) in at least two different quantitative levels are required. Typically, half an hour of 1Hz float readings for the raw data samples for each single exposure cycle characterized by a quantitative level forcers tuple, is needed. After this first step, to fully characterize the sensor node response, a Latin hypercube exploration of the combinations of the different quantitative levels of the forcers is recommended. Raw data sample for each transducer may be a single scalar value (i.e., electrode potentials for electrochemical sensors) or vector (see temperature modulation MOX sensors).

b) In the deployment phase the single node is expected to output a stream of amount of substance fraction data with sampling frequency of 0.1 to 1Hz. Data tuple typically include at least two or three gas amount of substance fraction readings, and a measure of primary environmental drivers (e.g., temperature and relative humidity). Depending on the final data application, wind direction and wind speed might be needed. Data coverage should exceed 95%. The most important requirement is accuracy. In EU, reference guides do exist and measure the accuracy in terms of relative expanded uncertainty of the node in relevant ranges of concentrations.

Use-case of a thermocouple sensor network contains simple data format in the form of temperature versus time. The time interval is highly variable but typically 30 seconds or 1 minute, for time scales of between days and years. The scatter is generally low, within about 0.2°C. There are generally multiple channels, each channel representing one thermocouple from two to more than 50.

2.2.2 Typical and possible use cases of a specific sensor network

Some typical use cases of air quality sensor networks include mapping of urban air pollution, which if the network is dense enough can be achieved by spatial interpolation. Data requirements in this example may be sufficient data coverage for obtaining representative averages, and sufficient spatial density, since kriging itself will introduce error if the estimate is not based on a sufficient number of data points.

The networked devices are typically used in two deployment modes: mobile and fixed. Use cases for mobile deployment include pervasive emission monitoring (along the streets), pervasive air quality monitoring, personal exposure monitoring, etc. They differ in terms of data quality level and sampling frequency parameters. Personal exposure requires sampling frequency in the range of one sample per 30 sec or minute. Mobile monitoring with bicycles, cars, or buses requires several samples per minute (ca. 10), this allows for a sufficient spatial resolution. Coverage can be more stringent especially when pursuing single pass routes.
Fixed deployments include pervasive air quality monitoring for high spatial resolution information integrating the regulatory grade network with low-cost sensor systems. Data requirements have already been described above. In both cases, sampling averaging procedures with raw sampling in the range of several samples per second will allow for implementation of noise reduction algorithms.

Use case of industrial temperature sensor network: This use-case is primarily high-value heat treatment processes. This encompasses areas such as aerospace and automotive, where the sensors are required to exhibit high stability, i.e., low calibration drift, in harsh environments with temperatures up to 1500 °C for long time periods up to years. Here, physical modelling is being used in one of the potential approaches to deploy multi-wire thermocouples (the drift is mainly caused by vaporization of Pt and Rh oxides which results in changed local wire composition). Also, data mining approaches are being investigated, and hybrids of the two approaches.

2.2.3 Extent of needed data in specific sensor networks

In air quality monitoring, sensor network's lack of data coverage can disrupt reporting. Some common problems can be met with mitigation strategies such as duplication of sensor nodes. Field calibration campaigns can fail to produce sufficient data to give confidence in derived calibration models. This can be mitigated by increasing the duration of the calibration effort. However, monitoring the uncertainty of sensors over their lifetime in such application scenarios is challenging.

Air quality monitoring sensor networks can play a crucial role in administrative decisions. Accuracy is a mandatory requirement that can be attained to a certain extent with low-cost sensors in a trade off with costs. While low-cost sensor systems remain an attractive solution, ensuring their accuracy through current state-of-the-art in-field calibration methods can significantly increase costs. Thus, while sensors themselves are affordable, achieving high accuracy requires additional investment. Several strategies are under study to allow for a feasible way to guarantee accuracy including multiunit (universal) calibration coupled with calibration transfer strategies [15], which reduces the number of samples used for (re)calibration of each unit. Another solution under study implies the use of continuous recalibration strategies using distant reference grade monitoring station data as reference, hence avoiding field colocation and saving on the logistics costs. However, long term and general figures on the attainable accuracy are rare in literature and unavailable for commercial systems.

2.3 Data quality and validation: Methods, Processes, and Best Practices

This section explores how to perform data validation, i.e., ensuring data is "fit-for-purpose" and of high enough quality. The exact definition of "high quality" can vary from case to case and there are many different approaches to defining, validating, and maintaining it. There is however no doubt that data quality is an important subject. In today's world where Big Data, Artificial Intelligence (AI), Internet of Things (IoT), sensors etc., are becoming more and more common in society – not only for research and big companies – it is becoming more and more important that the vast amounts of data being collected and exploited, meets the quality required for its application. There are several risks involved in analyzing data as well as in drawing conclusions and taking decisions based on data of poor quality. For the creation of data quality requirements and data validation rules it is important to have a good understanding of the data itself. From here it is necessary to prioritize which data is most important and which characteristics of the data are most important. Subsequently, rules and metrics should be applied to data relevant for those, i.e. some rules might only be applicable to certain data.

2.3.1 Importance of data quality

It is important to make sure data is of high quality due to risks of drawing wrong conclusions or taking the wrong decisions based on data of poor quality. Data quality needs to meet the needs and expectations of data consumers, creators, and other stakeholders. Risks are a large factor in assessing data quality from both business and health perspectives depending on the system, for example, in gas flow meter (or sensor) networks.

In sensor networks errors can originate from many sources, e.g., architectural, data flow, edge devices, data transfer and processing, cloud, storage, and analysis. Possible correlations and relationships, both temporal and spatial, between sensors and measurements can also give rise to both challenges and insights. It is important to monitor any issues related to sensor data and address these as they arise.

2.3.2 Variations of data quality dimensions

The data quality dimensions vary from source to source both regarding the number of dimensions as well as the terminology itself for each dimension where the meanings are sometimes mixed between dimensions and overlap with each other.

Table 1: Overview of selected sets of data quality dimensions.

Source	Dimensions	Note
[16]	14	Hierarchy with two layers (5 in first layer, 14 in second layer)
[17]	21	Dimensions in total, from literature review.
[1]	24	Dimensions in total, from literature review.
[10]	8	Common dimensions. Mentions a few more.
[18]	4	Core dimensions
[19]	15	Separated in two points of view: inherent, system dependent. Some dimensions are in both.
[20]	18	Grouped in 4 categories. Literature survey
[21]	20	Grouped in 4 categories. Literature survey

As seen in the table above there are many variations when it comes to data quality dimensions. A subset of the data quality dimensions is recurring and has commonly agreed definitions such as accuracy and completeness, while others only appear in one or few sources. Furthermore, some dimensions vary in the way they are defined. Not all sources have time series or sensor networks as their focus and instead look broadly at data in general. The literature review of [1] has a focus on IoT related papers and it is thus relevant when looking at sensor networks.

2.3.3 Existing approaches

Several approaches exist for performing data validation and defining data quality requirements. Data quality dimensions are at the center of defining data quality requirements, and as seen above, there exist many different versions of these along with different approaches to work with data quality. Researchers and different types of organizations address different aspects, methods, approaches, and processes for performing data quality management in general. The following subsections will briefly describe some standardized processes for data quality management. These are also used as inspiration for some of the best practice approaches described later.

2.3.3.1 ISO 8000-61

The ISO 8000 series is a series of standards for data quality management. Specifically, ISO 8000-61 describes an overall approach to data quality management. The core process cycle of ISO 8000-61 is based on the Plan-Do-Check-Act cycle of ISO 9000 which is a well-known series of standards on quality management systems.

The ISO 8000-61 process consists of three areas where the first is "Implementation" consisting of a modified version of the Plan-Do-Check-Act cycle. The second and third areas are Data-Related Support and Resource Provision, which support the core process cycle. The full model can be seen in Figure 1.

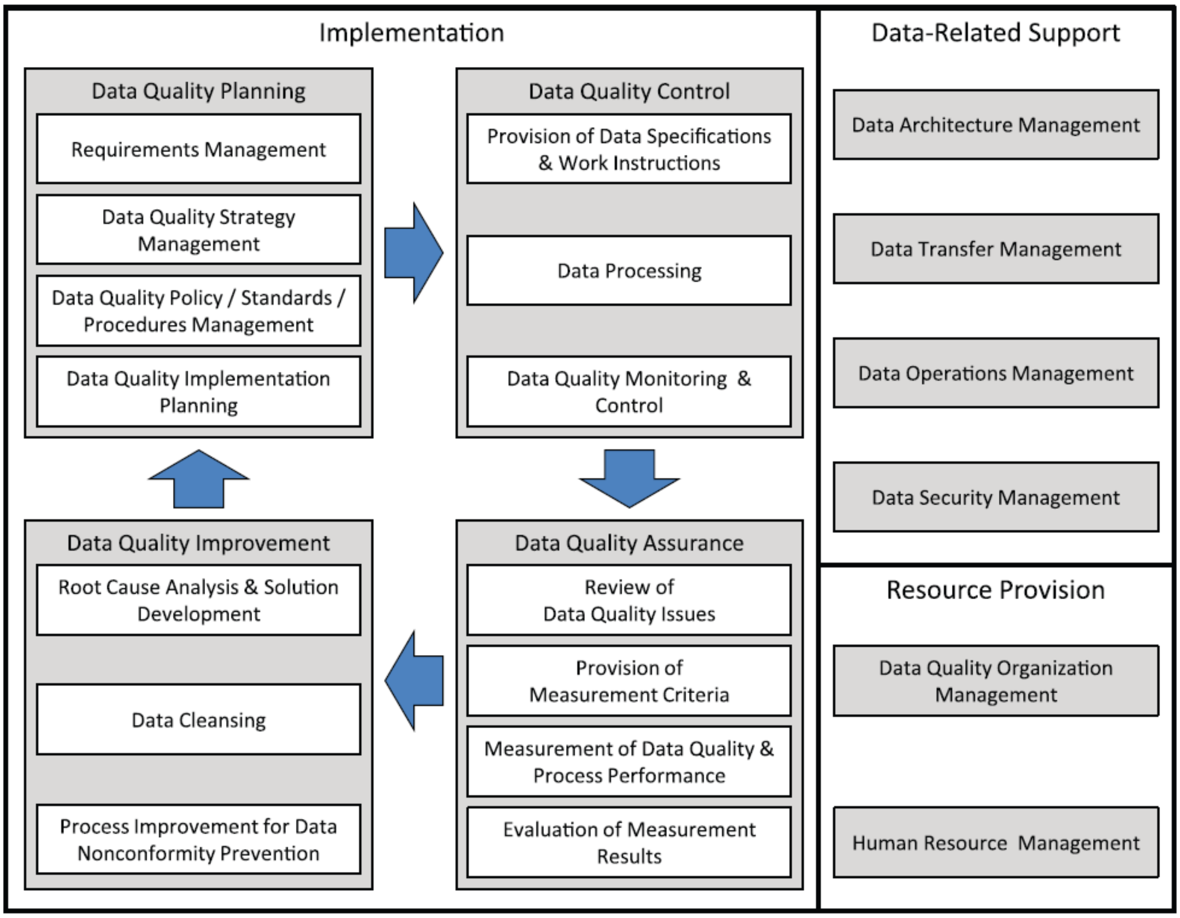


Figure 1: ISO 8000-61 data quality management process.

The Implementation process follows the cycle where Data Quality Planning is the first step, then Data Quality Control, Data Quality Assurance, and finally Data Quality Improvement before the cycle repeats. The Data-Related Support process enables the Implementation process with information and technology related to data management and the Resource Provision process improves the efficiency of the two other processes by providing resources and training services on an organizational level. Within the model described in Figure 1, there are 20 lower-level processes which will not be described here, but it should be noted that Requirements Management is the first of the lower-level processes and is where data requirements are defined.

2.3.3.2 Methodology for data validation

The Methodology for Data Validation from the European Commission describes a process' life cycle to define and execute data validation (see

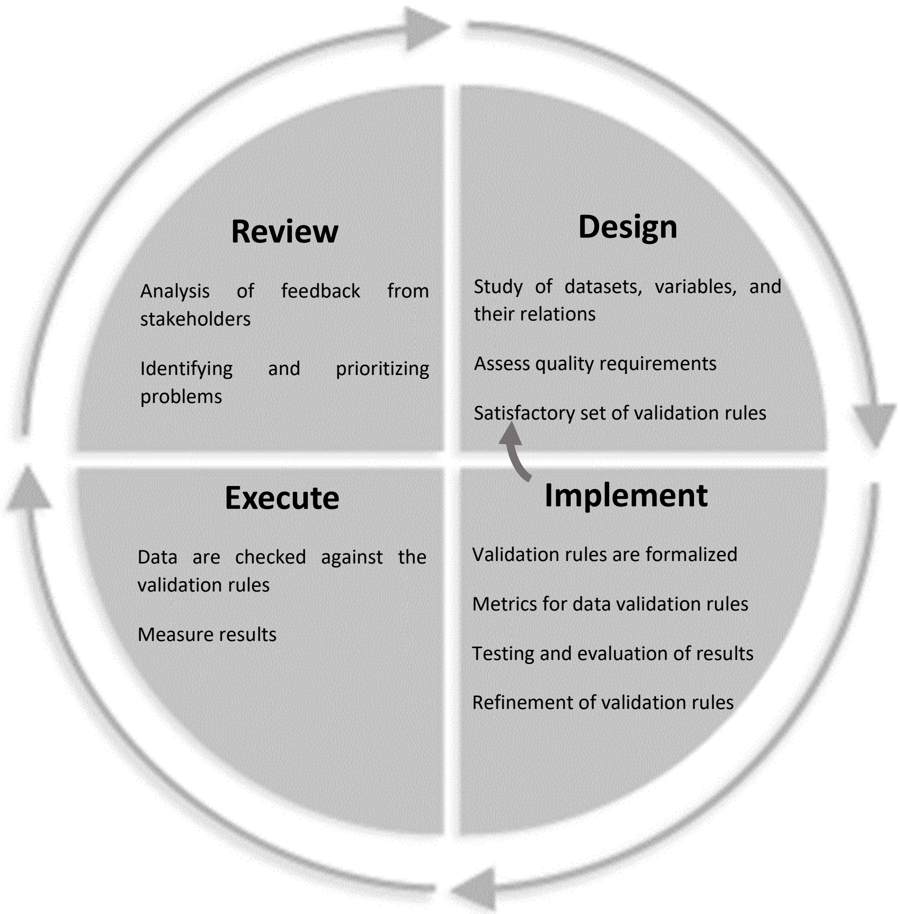


Figure 2.

Figure 2: Data validation process life cycle from [22]

The cycle has four main phases and starts with the design of the data validation process. The Design phase includes the familiarization with the dataset (through the study of the datasets, variables, and their relations) and the assessment of quality requirements, as well as the definition of data validation rules. In the Implementation phase, the validation rules are described, formalized, tested, refined, and discussed by stakeholders. In the Execution phase, data is checked against the validation rules and the results are measured and quantified. The final phase is the Review phase where the validation rules are improved, based on the feedback from stakeholders, before the cycle can start over.

2.3.3.3 DAMA International

The Data Management Body of Knowledge (DMBoK) of DAMA International (a global Data Management organisation) [10] is concerned with everything related to data management. Data quality is one of the knowledge areas in the Data Management Framework. The DMBoK goes through what Data Quality Management is, what the activities are, as well as both the inputs for those activities and the outcomes. It also describes different tools, techniques, and metrics. It gives some examples of rules and metrics but is mostly concerned with organizational aspects, the framework, procedures, people, etc. around these.

Thus, in the context of metrology, DMBoK is too high level but the activities described therein are relevant for best practices in defining data quality requirements and working with it in general. The activities mentioned include defining what high quality data is, defining the critical data and business rules, and doing an initial quality assessment.

The DMBoK also highlights the Plan-Do-Check-Act cycle as the Data Quality Improvement Life Cycle (see Figure 3. This cycle is used to improve data quality and it starts by scoping and prioritizing data issues in the planning step. Then in the "Do" step, root causes of issues are addressed and a plan for continuous monitoring is made. The "Check" step is about actively monitoring data quality as it is measured against requirements. If data quality falls below accepted levels additional actions must be taken to reach acceptable levels. The "Act" stage involves activities for addressing and resolving emerging data quality issues. When issues are assessed and solutions proposed, the cycle restarts.

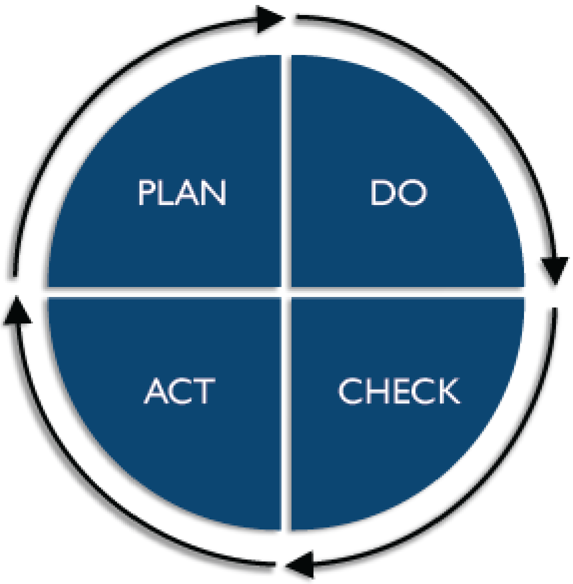


Figure 3: Plan-Do-Check-Act cycle from [10].

2.3.4 Defining data quality metrics

A data quality requirement or a data quality rule needs to be translated into a metric that can be measured. Ideally, the metric is presented in a machine-interpretable form to enable its processing by automated systems [9]. Within each data quality dimension, several different metrics can be defined. Defining metrics is about figuring out how to quantify or how to measure a certain requirement. Below are a few examples of rules and corresponding metrics. Each example uses a data quality dimension as the foundation and exemplifies a rule within the given dimension and a metric for calculating it. The thresholds in the examples are randomly chosen. In a real-world scenario rules, metrics, and threshold values would depend highly on the use case.

Summary of defining metrics (continued from the defined requirements):

1. Make sure the requirement or rule is quantifiable and measurable
2. Identify the variables for the metric
3. Identify relationships between variables
4. Adjust to give the correct output format or unit

<u>Example 1</u>

Dimension: Completeness

Rule: Missing data: No less than 95% of data points should be present in the time series

Metric: $time\ series\ completness\ [\%] = \frac{\text{actual number of rows}}{\text{theoretical number of rows}} \times 100$

Description and example: Given a start and end time, as well as a frequency of measurements for a given sensor, the theoretical number of rows can be calculated. For example, one measurement per minute from 1st April to 30th April (both included), will give 43200 data points, theoretically. If for example the actual number of data points is 42000 it means 1200 data points have been lost and gives a completeness of ~97.2% and the completeness lives up to the 95% requirement. (The scope of this rule is a time series from one sensor, but its scope could easily be extended to a sensor network collecting data from multiple sensors).

<u>Example 2</u>

Dimension: Accuracy

Rule: No more than 1% of values in a time series should be beyond absolute threshold for reference sensor

Metric: $sensor\ accuracy\ [\%] = \frac{number\ of\ values\ in\ time\ series\ beyond\ threshold}{\text{number of values in time series}} \times 100$

Description and example: Taking a time series of sensor values and comparing them to the time series of a reference sensor (the two time-series should of course be aligned (i.e., same start and end time, same measurement frequency, etc.). With a defined threshold of, for example, ±1°C, the difference between a sensor value and its corresponding reference value cannot go above or below the threshold. The rule then says to count the number of values exceeding the threshold and see if this is more than 1% of the total number of values in the time series.

<u>Example 3</u>

Dimension: Timeliness or Currency

Rule: The interarrival time between sensor measurements cannot exceed 30 minutes.

Metric: $sensor\ timeliness\ [Minutes] = timestamp_n - timestamp_{n-1}$

Description and example: Depending on the focus, this rule can both be related to timeliness or currency. Timeliness can be viewed as the time from sensor measurement to it being stored in the database and then made available to the user. This rule checks the time passed from the most recent measurement and the measurement immediately before that. If this duration is too long it might violate the timeliness. This could be due to long transmission time between the edge device, gateway, and server. If data quickly becomes outdated due to quick changes in the environment of the sensor, a long time span between measurements might also violate currency requirements since the latest measurement has been outdated before the next one arrives. Also, if the new measurement arrives late, it might already be outdated.

2.3.5 Maintaining data quality

The cyclic nature of the processes mentioned in a few of the previous sections is an important aspect of maintaining data quality, since it is a continuous process and not a one-time project. Data is continuously generated, which means there is constantly new data that needs to be validated to make sure the quality is good enough. This is especially the case in sensor networks, where large amounts of data can be gathered at a high frequency. Furthermore, there can be new uses of the data, or the requirements can be changed. Both affecting the requirements and process.

Sections 2.3.3 and 2.3.4 describe how to get from a dataset to a set of requirements and from there how to define metrics for these requirements. The basis for these can be the data quality dimensions. The steps mentioned can be carried out whenever requirements change, new issues arise, or new uses are found for the data. Furthermore, it is beneficial to follow the common or standardized approaches as described in previous sections.

Not only is it beneficial to use a cyclic process for updating requirements and maintaining data quality, but it is also desirable to implement an automatic process for executing the rules on data. It is infeasible to do it manually every time new data is generated or to go back and do it on historic data whenever requirements change. An automatic process becomes even more crucial when considering sensor networks and the amount of data which can be generated.

Maintaining data quality can also be structured by both defining the most important data quality dimensions and defining critical data. In the use cases used as examples throughout the guide the most important characteristics of data are completeness and accuracy. In addition, it is important that the data is traceable (traceability) and not outdated (timeliness/currency). Choosing the most important characteristics makes it easier to focus on the most important requirements. The same is the case for defining critical data. If data can be categorized like this, stricter rules can be put on critical data and more lenient rules on less important data, making the validation task lighter and avoiding setting strict rules for all data when only a subset of data has such strict requirements.

2.3.6 Global calibration models for air quality sensor

Low-cost air quality (AQ) sensor networks introduced a promising paradigm shift, going from traditional monitoring equipment, which has high accuracy but also associated high costs of initial installation and maintenance, to a much more cost-effective solution of low-cost compact devices with IoT features integrated into a network. This shift can, due to cost-effectiveness, increase spatial resolution compared to traditional monitoring.

However, for these low-cost devices to reach sufficient accuracy to be deployed as indicative measurement devices in AQ monitoring networks, the calibration process needs to be periodically repeated. This frequent calibration has as purpose to mitigate the problem of sensor drift. Moreover, if the calibration location and deployment location are similar it can also eliminate the concept drift, which is also known as calibration location-deployment location mismatch (De Vito et al 2023, Topalovic et al 2019). These additional efforts, especially needed when accuracy is a concern, can significantly contribute to overall costs, which hinders and complicates massive deployment.

One possible solution for reducing the cost of calibration is to perform global calibration models (GCMs). GCMs are used when multiple sensor units, typically originating from the same fabrication process, need to be calibrated in a batchwise manner. GCMs are often useful for low-cost sensors with a tendency towards high inter-device variability [24]. In this approach a calibration model is built with the matrix of responses of multiple sensor units (i.e., a subset of network nodes) exposed to the same calibration conditions. The resulting global model is then applied to new uncalibrated replicas after being optimized to maximize prediction accuracy in new replicas [25]. By transferring the GCM to the complete network, the complexity of calibration/recalibration campaigns and subsequent related efforts on network quality assurance are significantly reduced.

Here recently obtained results by several research groups that examine GCMs when applied to several different types of sensors used in air quality monitoring are summarized. In addition to targeting different pollutants, these global calibration approaches also utilize several different GCMs, such as linear global calibration model for particulate matter (PM) equivalent mass concentration of $PM_{2.5}$ and PM_{10} (De Vito et al, 2023), GCMs that target different gaseous pollutants measured by electrochemical sensors (namely limited quadratic regression calibration models for carbon monoxide (CO), neural network models for nitric oxide (NO), hybrid models for nitrogen dioxide (NO_2) and ozone (O_3) (Malings et al, 2019), machine learning (ML)-based GCMs for NO_2 and NO low-cost sensors [26]), by metal oxide (MOX) gas sensors like the GCMs for temperature-modulated MOX CO gas sensors (Miquel-Ibarz et al, 2022), parallel machine learning based calibrations for $PM_{2.5}$ sensors [26] and deep learning based calibration for Metal Oxide Semiconductor (MOS) gas sensors. An example of the latter is found in [28], where gas mixtures including seven target Volatile Organic Compounds (VOCs: acetic acid, acetone, ethanol, ethyl acetate, formaldehyde, toluene, and xylene) and two background gases (CO and hydrogen), as well as the relative humidity, which was analyzed by using MOS gas sensors (SGP40, Seinsirion AG, Switzerland).

Appendix B gives an overview of global calibration approaches as applied by several research groups (De Vito et al., 2023, Malings et al., 2019, Miquel-Ibarz et al., 2022, [26], [28]. Calibration models that were used range from simple linear regression models to deep neural networks. Majority of the examples are derived from in field collocation campaigns, while two examples for VOCs [28] and CO (Miquel-Ibarz et al., 2022) are derived using data obtained in laboratory conditions. Model features for simple linear models include low-cost sensor signals, typically alongside meteorological parameters. Relative humidity (RH) is the meteorological parameter included in the models for PM, and RH and temperature (T) are the parameters added to the hybrid models for gaseous pollutants in Table 2 (note that here linear models are only used for large concentration that are near the ones observed during training). For more complex models, the feature set often includes all environmental signals produced by the sensor. Most common performance metrics are the root mean square error (RMSE), mean absolute error (MAE) and the coefficient of determination R^2 (between the calibration model prediction and reference concentration) [27].

2.3.7 Using data quality metrics in the design of a multi-wire thermocouple

In this section we describe an example where data quality metrics are used to inform the design of a sensor network, here taking the form of a multi-wire thermocouple used to measure temperature in high value manufacturing applications.

A thermocouple is a device for measuring temperature. A thermocouple comprises two dissimilar metal wires joined together at one end to form a measurement junction. A voltage is developed across the wires, which is measured at the open end and is a function of the temperature gradient between the two ends. Each wire develops an electromotive force (emf) in a temperature gradient, which is the thermoelectric effect. For any small length of wire, this emf is the product of the Seebeck coefficient (defined as the voltage generated per unit temperature change along the wire, which is characteristic of a given metal) and the temperature difference from one end of that length of wire to the other.

Thermocouples made of platinum (Pt) and its alloys with rhodium (Pt-Rh) are widely used in high value manufacturing applications as they offer relatively high thermoelectric stability in comparison with other thermocouples. The principal cause of instability, and therefore calibration drift, is the vaporization of platinum and rhodium oxides from the wires, which causes a local change in composition, and hence a local change in Seebeck coefficient. This in turn changes the emf generated in a given temperature gradient, which results in a temperature measurement error because the measured emf is different to the emf that would have been generated during the original calibration for the same temperature gradient.

The magnitude of the effect depends on the Pt-Rh composition of the wire. A Pt-30%Rh wire drifts more slowly than a Pt-6%Rh wire because the same quantity of rhodium lost in a given time interval is a smaller proportion of the total amount in Pt-30%Rh. A thermocouple assembly comprising of several wires allows measurements to be made simultaneously by different thermocouples defined by different pairs of wires. For example, a 5-wire thermocouple with the widely available compositions Pt-0%Rh, Pt-6%Rh, Pt-10%Rh, Pt-13%Rh and Pt-30%Rh offers 10 possible pairs of wires and so 10 thermocouples.

Since the different thermocouples have wires in common, e.g., a Pt-6%Rh versus Pt-30%Rh thermocouple shares a Pt-6%Rh wire with the Pt-6%Rh versus Pt-13%Rh thermocouple, etc., there is the possibility to use the resulting correlations between the measurements made by the thermocouples in the ensemble to provide information about calibration drift. Additionally, there is the possibility to use the measurements to provide a measurement of temperature that is 'better', e.g., has lower uncertainty, than that provided by any individual thermocouple. These possibilities, which are the subject of on-going research and development, are the motivation for the so-called multi-wire thermocouple. The multi-wire thermocouple is an example of a sensor network in which the individual sensors (thermocouples) are co-located and are measuring the same measurand (the temperature at the measurement junction).

Various methods are being studied for using the data recorded by such a multi-wire thermocouple to estimate the drift of the individual thermocouples and to obtain an estimate of the (common) temperature measured by the thermocouples and its associated standard uncertainty. Here, a study is made of the influence on the measurement results obtained from one of those methods from the choice of the design of the multi-wire thermocouple in terms of the number of wires and the compositions of the wires. The influence is assessed in terms of various data quality metrics, including the difference between the estimate of temperature and its known value, the uncertainty of the estimate of temperature, an estimate of the magnitude of the noise in the measured emf data, and the cost of fabricating and calibrating the multi-wire thermocouple. An optimal design is one that balances the quality of the information delivered against the cost of obtaining that information.

For illustration, data is recorded by a multi-wire thermocouple comprising of the seven wires Pt-5%Rh, Pt-8%Rh, Pt-10%Rh, Pt-13%Rh, Pt-20%Rh, Pt-30%Rh, and Pt-40%Rh. The data was obtained while the thermocouple was immersed in a calibration artefact (a cobalt-carbon fixed point with a melting temperature of 1324.29 °C), enabling periodic re-calibrations of all 21 thermocouples and hence yielding direct measurements of thermocouple drift *in situ* [30]. The data comprised measured values of emf for each thermocouple made of two different wires. Data was recorded over a time period of about 1,520 hours, but the study focused on the time period between about 200 hours and 650 hours with data at the start of the time period omitted as in that early period there can be mechanisms other than evaporation that cause the observed calibration drift.

For each design of the multi-wire thermocouple, the following measurement results are calculated: the estimates $T_i, i = 1, \dots, N,$ of the known temperature $T^* = 1324.29\,℃$, the standard uncertainties $u(T_i), i = 1, \dots, N,$ of those estimates, and estimates $s_i, i = 1, \dots, N,$ of the standard deviations of the noise in the measured values of the emf. These results are then summarized by the following data quality metrics:

$$M_1 = \left| \frac{1}{N} \sum_{i=1}^{N} (T_i - T^*) \right|, \qquad M_2 = \frac{1}{N} \sum_{i=1}^{N} |T_i - T^*|, \qquad M_3 = \sqrt{\frac{1}{N} \sum_{i=1}^{N} s_i^2}, \qquad M_4 = \sqrt{\frac{1}{N} \sum_{i=1}^{N} u^2(T_i)}.$$

Additionally, the cost of fabricating and calibrating the multi-wire thermocouple is calculated using the following information (correct at time of writing): wire prices (per metre) are £147 (Pt-0%Rh), £176 (Pt-10%Rh), £186 (Pt-13%Rh, estimated), and £243 (Pt-30%Rh). Each wire is generally two meters long, and the cost increases linearly with rhodium content as rhodium is more expensive than platinum. A linear interpolating function of wire cost with rhodium content is considered appropriate to calculate the cost of other wires. The cost of calibrating a thermocouple is approximately £1000, independent of thermocouple type, and would be a negligible amount for several thermocouples in a multi-wire assembly.

A method for drift estimation is applied to the data corresponding to the complete multi-wire thermocouple comprising all seven wires and made up of 21 individual two-wire thermocouples. Additionally, the method is applied to each multi-wire thermocouple formed by omitting a single wire, which reduces the number of individual two-wire thermocouples to 15. Table 2 gives the values of the various data quality metrics for each design of multi-wire assembly, including the cost of fabrication and calibration.

For the cases when the wire Pt-20%Rh or Pt-40%Rh is removed, the closeness of fit between the data and the model is better, e.g., as measured by M_3 in Table 2 but this might be a sign of overfitting of the model to the data. Removing these wires, and also the Pt-30%Rh wire, has an impact on the quality of the temperature estimates as measured by the metrics M_1 and M_2 (Table 2), suggesting that the estimates of the temperature in these cases are poorer overall. For these cases, the standard uncertainties of the estimates of temperature are also noticeably different than for the other cases (metric M_4 in Table 2). The small values for the standard uncertainties when the wire Pt-20%Rh or Pt-40%Rh is removed is likely linked to the smaller estimates of the standard deviations of the noise.

For these reasons, it is concluded that it is important to include the wires Pt-20%Rh, Pt-30%Rh and Pt-40%Rh (which coincidentally are least affected by calibration drift) whereas removing one of the other wires has little impact on the results. In terms of cost, the resulting multi-wire thermocouple assemblies are more expensive to fabricate and calibrate. The results suggest that an 'optimal' choice for removing a single wire is to remove the wire Pt-13%Rh.

Table 2 Values of data quality metrics for different designs of multi-wire thermocouple, including a design with all the wires and designs with an individual wire removed.

Wire removed	M_1/℃	M_2/℃	M_3/μV	M_4/℃	Cost/£
None	0.018	0.078	5.223	0.314	3,839
Pt-5%Rh	0.024	0.090	5.941	0.458	3,521
Pt-8%Rh	0.018	0.077	5.931	0.389	3,501
Pt-10%Rh	0.019	0.078	5.978	0.389	3,487
Pt-13%Rh	0.023	0.081	5.916	0.393	3,467
Pt-20%Rh	0.030	0.090	2.850	0.187	3,420
Pt-30%Rh	0.063	0.129	5.976	0.594	3,353
Pt-40%Rh	0.099	0.207	1.045	0.122	3,286

3 Metrological Traceability

3.1 Methods and Guidelines for In-Situ Self-Calibration or Co-Calibration with Reference Sensors in a Sensor Network

Establishing the traceability of measurements to the international system of units (SI) in sensor networks (SN) is essential from a metrological standpoint [26]. One of the key aspects in this regard are traceable calibration operations, which ensure the link from a measurement to its appropriate SI unit via an unbroken chain of calibrations [27]. Conventional calibrations, which are carried out in specialized laboratories, involve the comparison of the values delivered by the device under test with the measurement values provided by a reference standard. A reference standard can be another device or physical artefact with known properties or quantities derived using fundamental physical constants. In the case of a sensor, another reference sensor with known uncertainty can be used as a reference sensor. Such a calibration is typically very expensive and not cost-effective, particularly for low-cost sensors. The deployment of sensor networks in hard-to-reach locations and carefully controlled environments further reduces the feasibility of regular laboratory calibrations.

In-situ calibration is a common work-around in such cases. In this context, in-situ refers to the characterization of the measurement model of the sensor and its uncertainty, i.e., its calibration being performed at the location of its deployment without having to disassemble and transport it to a calibration laboratory or factory [28]. For this purpose, a transportable device with known accuracy can be used. A co-calibration, on the other hand, can be considered as a special case of an in-situ calibration where nearby sensors already present in the network are used as reference devices [29]. In the case of co-calibration, the reference value itself may have to be estimated at the position of the device under test using appropriate interpolation and sensor fusion techniques.

The report provides an overview of currently available methods to (self-/co-)calibrate sensors within a sensor network and assesses the suitability for metrological use-cases. The insights from the literature review are joined with the discussions made with the project use-cases to provide guidelines and practical considerations within real-world SN.

3.1.1 State of the Art

An excellent review on in-situ based co-calibration in sensor networks is provided by Delaine et al. [28]. Macro-calibration and blind-calibration are two concepts that are closely related to in-situ and co-calibration. Macro-calibration refers to the calibration of an entire sensor network based on its total response without having to calibrate each individual sensor node

[30]. Blind-calibrations [31] refer to the steps taken to achieve homogeneous behaviour of all sensor nodes by possibly enforcing the dominant influence of sensors that are a priori known to provide sufficiently good (calibrated) measurements. This is typically done in cases where there are no reference signals/sensors, or other sources of ground-truth information about the measured process. A consolidated review of algorithms relevant to blind- and macro-calibrations as well as a consensus-based extension to the distributed case can be found in [32]. An extension of the algorithm with uncertainty evaluation as well as a treatment of traceable co-calibration was provided in [33].

3.1.2 Guidelines and Considerations in Real-World Sensor Networks

Based on discussions within the project consortium and information from the literature review, guidelines and practical considerations for the application of co-calibration methods in real-world sensor networks are provided. In addition to some general remarks, these guidelines cover generic scenarios of prototypical sensor network configurations. The scenarios were chosen based on their relevance to the use cases within the project. Each scenario is described, relevant co-calibration methods are presented, and their applicability discussed. Concerning the co-calibration methods, the approach, strengths, weaknesses, and complexity of the transfer behaviour are described.

3.1.2.1 General Remarks

It is encouraged that the data from the sensor network is available/retrievable in a structured and processing-friendly way, e.g., REST, JSON, HDF5, and SQL. If traceability of reference sensors is of importance, this data also needs to include information about the measurement uncertainty. Moreover, it is important to provide a way to align and relate datapoints, meaning that, i.e., measurement data in sensor networks are timeseries of datapoints and the timestamps are based on the same timescale across the network. Ideally, metadata (e.g., sensor identifier, location, unit, calibration status, etc.) is available alongside the (numerical) measurement data to simplify the selection of reference sensors and facilitate automation at a later stage.

3.1.2.2 Detecting Sensors to be Calibrated

As with physical sensors, the intervals at which a co-calibration needs to be performed also needs to be determined. Such intervals depend on the quality of sensors involved, the application for which they are used, and the accuracy required. The interval can be fixed based on a particular standard or it can be flexible, relying on an online monitoring of the constituent sensors [34].

Sensor transfer behaviour: The transfer behaviour of a sensor refers to the relationship between the physical quantity being measured and the value provided by a given sensor. The relationship can usually be expressed by means of a parametric expression. The most common choice is a linear affine model, i.e., the output y and input x, are linked by $y = a * x + b$ with potentially multivariate gain a and offset b. More complex transfer behaviour is rarely encountered. Another general observation is the lack of uncertainty evaluation / sensitivity analysis and with that missing traceability of the estimated transfer behaviour. An understanding of the transfer behaviour of a sensor is indispensable when measuring time-varying quantities requiring the estimation of dynamic measurement uncertainties.

3.1.2.3 Scenario A: Dense Sensor Network with Stationary Sensors

A sensor network can be called "dense" if the specific quantity of interest does not change much between (spatially) neighbouring sensors. This can be leveraged to obtain reference values from the network. [35]The review paper by Delaine [28] contains a list of co-calibration methods are designed for such networks. For example, Stankovic et al. [32], [36], [37] proposed a consensus-based calibration algorithm for a set of co-located sensors. Computation complexity in this case is rather low due to a gradient-decent optimization approach. It estimates the gain and offset of a linear affine transfer behaviour. Weight factors can be used to include only traceable reference values into the calculations. Uncertainty evaluation is not covered in the original papers but is straightforward and was added in [33], but calibration performance on transient signals could be limited.

Kizel et al. [38] proposes a node-to-node calibration approach, assuming pairwise co-location. The in situ calibrated sensor therefore form a chain of calibration dependencies. In each link of this chain the gain and offset are estimated using least square regression methods. Sensitivity analysis of the estimated parameters is provided, although this is not an uncertainty propagation in the strict sense. Moreover, the influence of the chain length onto the "uncertainty" is investigated.

Gruber [33] assumes the availability of a virtual reference for the sensor to be calibrated. In a simple case, this would come from an appropriately robust mean of co-located sensors. Uncertainty of the virtual reference is directly included in the parameter estimation process by following a Bayesian approach, which also leads to probabilistic distributions of the sought parameters of a linear affine model with an error term. The method provides traceable results according to the definition in the VIM.

3.1.2.4 Scenario B: Sparse Sensor Network with Stationary Sensors

Within a sparse sensor network, the specific quantity of interest can change considerably between neighbouring sensors. It is therefore necessary to fill potential information gap with model-based approaches. Examples of such sensor networks can be found in energy networks [39] and chemical production plants [40]and smart buildings [41]. A generic approach to handle such sensor networks is to make use of the methods listed in scenario A. This is done by first applying an interpolation model that quantifies the available knowledge about the spatial, temporal, scientific or technical relations, e.g., a Gaussian process interpolation that also takes correlations in the values into account [42]. Another common approach involves performing a spatio-temporal k-nearest-neighbour gradient-model interpolation [29].

Moltchanov et al. [43] and Tsujita et al. [44] propose simple but effective heuristics to find subsets of urban environmental pollution monitoring data that allows to be used for in situ calibration. While Moltchanov assumes that nightly measurement values are almost equal due to a lack of anthropogenic influence, Tsujita concludes from almost equal measurements of reference stations across the city that the overall distribution is uniform. The selected data is then used to estimate the transfer behaviour using least-squares techniques. No uncertainty evaluation was given but could be added according to [45] or Monte Carlo approaches. Lin [46] proposes to regularly bring a reference sensor in the vicinity of the sensor for a specified time duration to be in-situ calibrated (opposed to a classic calibration, which brings the sensor to the (laboratory) reference).

Further interesting ideas are found across the literature. For gas sensors, Sun et al. [47] develops a sensor that periodically can be inserted into / surrounded by a known gas mixture, providing known and traceable reference measurements. Bychkovskiy et al. [48] does not directly compare time-series or datapoints, but the histograms of periods in time that are highly correlated between two sensors. Martin et al. [49] allows for additional model terms (e.g., dependency on temperature or humidity) and only includes them, if they increase the transfer model performance.

3.1.2.5 Sparse Sensor Network with Mobile Sensors and Stationary Reference Sensors

Another prototypical situation is the use of mobile sensors (e.g., mounted on a vehicle). The sensor density of such networks is typically sparse but offers the possibility of a sensor "rendezvous" – a limited subset of datapoints that fulfils the properties of a dense sensor placement. Based on the review paper by Delaine [28], the following co-calibration methods are designed for such networks. The mobility is usually leveraged by search for temporary co-location or "rendezvous" of sensors in the data.

Miluzzo et al. [50] propose that uncalibrated nodes/sensors can ask for nearby reference readings. These readings are then distance weighted to obtain a virtual reference reading to estimate the offset (but no gain). Although no uncertainty evaluation is carried out, doing so is likely to be straightforward using the GUM LPU (law of propagation of uncertainty).

The approach chosen by Hasenfratz et al. [51] and Saukh et al. [52], [53] is rendezvous based. Rendevouz in the context of mobile sensor nodes refers to a situation in which two or more sensors are in the temporal and spatial vicinity of each other, i.e., in a given spatial location at the same time. In contrast to [50], the datapoints selected for the in-situ calibration are age weighted, reducing the influence of older rendezvous. The parameter estimation is carried out using least square regression. Although a RMSE-value with regard to the (simulated) ground truth is provided, no uncertainty evaluation is documented. A very similar approach is proposed by Maag et al. [54].

3.1.3 Applicability To Real-World Use-Case of Air-quality Monitoring Networks

Sensor networks used for air-quality monitoring consist of many low-cost sensors with a significant number of mobile nodes and are an ideal use-case for the application of methods for re-calibration, self- and co-calibration developed as part of earlier tasks. Low-cost gas sensor systems can potentially increase spatial and temporal resolution in air quality monitoring networks in smart cities, but suffer from cross-sensitivities, interference with environmental factors, and ageing. These problems are compounded as these sensors usually operate under non-static conditions. The main requirement of this task is the availability of methods for uncertainty-aware sensor fusion, drift detection, dynamic uncertainty estimation, and optimized traceability paths. Given the large-scale use of low-cost sensors, e.g., 600 sensors distributed around the Parisian region [55], co-calibration has the potential to play a significant role in this use case. Moreover, a large majority of the sensors (500) in the aforementioned network are placed on postal service vehicles and don't have a fixed location as a consequence. The co-calibration methods developed will have to take this fact into account by potentially adapting the methods outlined in the preceding section.

3.1.4 Summary

The need and current state of in-situ calibration in sensor networks are described. To formulate guidelines and practical considerations for the application of such methods in real sensor networks, three prototypical sensor network scenarios are proposed. For each scenario, multiple promising and existing co-calibration methods are presented. The advantages of these methods regarding metrologically sound results are briefly discussed for each method, revealing a lack of uncertainty evaluation in many methods. Moreover, general remarks are provided that enhance the data quality of suitable datasets and prepare the automation of in-situ calibration methods. Finally, the applicability in rea-world use cases were discussed for three generic scenarios corresponding to common sensor network configurations: dense networks with stationary sensors, sparse networks with stationary sensors and sparse networks with mobile sensors and stationary reference nodes. The latter of the three aforementioned scenarios was further discussed for the specific case of air-quality monitoring networks. It was shown that the methods for co-calibration and *in situ* calibration must take the mobility of individual sensor nodes into account. The use of low-cost sensors in such networks further increases the need to develop methods for uncertainty-aware sensor fusion, drift detection, dynamic uncertainty estimation, and optimized traceability paths.

3.2 Methods for Uncertainty-Aware Sensor Fusion in Dynamic Measurements

A dynamic measurement can be defined as one where the physical quantity being measured (the measurand) varies with time and where this variation may have a significant effect on the measurement result (the estimate of the measurand) and the associated uncertainty [56], [57]. Typically, sensors used in industrial measurements have been calibrated under static conditions. In practice, however, the measurements are usually performed under dynamic conditions, i.e., the measured signal is non-stationary or transient. The use of a sensor in a different mode from that in which it was calibrated can significantly affect the reliability and uncertainty of the measurement result.

A measurement system necessarily has a finite response-time to a change in the physical quantity that is being measured, i.e. the measurand. In case the system - in our case a sensor – responds much faster than the rate at which the measurand changes, it is possible to directly analyze the measurement and compute the uncertainty by the conventional static means as defined by the guide to the expression of uncertainty in measurement (GUM) [58]. On the other hand, if the measurement system responds slowly to the rate of change of the measurand, the uncertainty determined using conventional static means is no longer accurate. In such cases the measurement uncertainty itself may be time-dependent [59].

3.2.1 Sensor Fusion

Sensor fusion can be defined as "the combining of sensor data or data derived from sensor data such that the resulting information is in some sense better than would be possible when these sources were used individually" [60]. In the context of sensor networks, information obtained from multiple sensors, often measuring different physical quantities, can be combined based on a mathematical model to generate values that cannot be directly measured. For instance, a common application of sensor fusion is to capture industrial processes in the form of a digital twin, i.e. virtual representations of sensors and sensor networks in the fields of discrete manufacturing and process engineering [61], [62]. In addition to measuring quantities that aren't measurable by conventional means, combining information from multiple sensors can [63]

- Increase the quality of data,
- Increase reliability and,
- Increase the coverage area of a measured quantity.

From the point of view of metrology, ensuring the traceability of the resulting "fused" or derived measurements by appropriately propagating the uncertainties is of utmost importance. In this report, a brief review of methods and literature relevant to sensor fusion will be explored with an emphasis on he added challenge posed by dynamic measurements and calibration. The results are informed by a survey conducted among the use case owners to address specific issues pertaining to the use of sensor fusion in their respective areas of expertise.

3.2.2 State of the art

3.2.2.1 Sensor Fusion

Over the years, a substantial amount of research in sensor fusion has focused primarily on the methods and their applications in diverse areas [64], [65], [66]. In comparison, literature on the incorporation of metrological principles in sensor fusion is limited. This problem is compounded when seeking sources relevant to sensor networks. An exception to this rule is the method to compensate for outliers while reducing the effect of sensor failure and drift in the case of homogeneous sensor fusion, which was presented in [61]. A general review of sensor fusion (referred to here as multisensory fusion) and consensus filtering was presented in [67]. Consensus filtering refers to a distributed algorithm that allows the nodes of a sensor network to track the average of all their measurements [68] in such a way that the information exchange only happens between neighbouring nodes. Consensus filtering is in fact a dynamic version of the average consensus algorithm which allows a network of agents (in our case, sensors) to agree on the average of a set of initial values [69].

Perhaps the most widely used method in sensor fusion is the Kalman filter and its extensions and derivatives [70]. The Kalman filter is used to produce estimates of unknown quantities over time using measurements from multiple sources along with statistical noise. The estimate thus produced is better than one obtained from a single measurement. The Kalman filter recursively updates the value of an unknown quantity under observation by combining the predicted value based on its previous state and a physical model describing its dynamic evolution along with measurements provided by sensors. The Kalman filter is an established method in sensor fusion and has given rise to variations such as the extended and unscented [71] Kalman filters in order to model nonlinear systems.

3.2.2.2 Dynamic Measurements

A measurement is considered to be a dynamic measurement when the value of the quantity of interest varies over time [56]. While the guide to the expression of uncertainty in measurement GUM/GUM-S1 [72], [58] provides methods for measurements that are constant in time, a dearth of similar techniques was identified for dynamic measurements, particularly in the context of traceability [57]. Traceability is typically established via a calibration of the measurement device or sensor and, in the dynamic case, must be performed under appropriate conditions. The resulting methods must also be consistent with the static case. As measurement systems for dynamic measurements in metrology can often be assumed to be linear and time-invariant (LTI), a sizable body of research is focused on developing methods for such models [73], [74], [75]. Typically, the sensor output is represented by a digital finite or infinite impulse response (FIR/IIR) filter such that the sensor output $y[n]$ at a discrete time point n is related to the physical stimulus $x[n]$ by

$$y_{\mathrm{fir}}[n] = \sum_{m=0}^{M} b_m x[n-m] \qquad y_{\mathrm{iir}}[n] = \sum_{m=0}^{M} b_m x[n-m] - \sum_{l=1}^{L} a_l y_{\mathrm{iir}}[n-l].$$

Equivalently, the sensor behaviour can be represented by the transfer function $H(z)$ in the frequency domain as

$$H_{\mathrm{fir}}(z) = \frac{Y_{\mathrm{fir}}(z)}{X_{\mathrm{fir}}(z)} = \sum_{m=0}^{M} b_m z^{-m}\,,$$

$$H_{\mathrm{iir}}(z) = \frac{Y_{\mathrm{iir}}(z)}{X_{\mathrm{iir}}(z)} = \frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{l=1}^{L} a_l z^{-l}}\,.$$

Performing a dynamic calibration would require the coefficients a_l, b_m as well as their respective uncertainties to be determined [76]. A brief outline of the basic principles of dynamic measurement analysis that combines inputs from different fields such as measurement science, statistics, mathematics and signal processing can be found in [77] and [78]. Methods for the analysis of dynamic measurements and dynamic calibration have already been implemented in several use cases. For instance, a method to incorporate dynamic uncertainty in real-time systems and compensate for jitter prior to sensor fusion was explored in [79]. Other areas where dynamic calibration has been explored are waveform metrology [80], the calibration of hydrophones used as medical ultrasonic instruments [81] and in the case of high-g shock-accelerometers [82].

3.2.3 Use case specific considerations

3.2.3.1 District heating

Sensor fusion is especially important to the case of district heating as typical networks are sparsely populated, i.e. physical sensors are located at only a few points. In order to optimize the network, it is important to deduce parameters at several other locations using sensor fusion. Kalman filtering is the most used method and accounts for the uncertainty of the estimated value, when set up correctly. An example of a sensor fusion application is the spatial interpolation to determine values of a quantity at points without a physical sensor.

3.2.3.2 Industrial Manufacturing

The main application of sensor fusion for the heat treatment of high-value components in advanced manufacturing is to detect and quantify calibration drift in sensors, specifically thermocouples comprising wires made of platinum (Pt) and its alloys with rhodium (Pt-Rh). The uncertainty of the estimates of the calibration drift for the individual sensors is an important metric for quantifying the reliability of the sensor network, which is in the form of a multi-wire thermocouple, as well as the drift detection algorithm. Currently, sensor fusion techniques for this use-case are under development with the aim to treat data fusion for sensors with different systematic drift characteristics as well as the handling of correlations and uncertainty in recorded sensor data. The fused values here do not correspond to interpolated measurements, but to an estimate of the common temperature measured the sensors. Some of the methods for sensor fusion under consideration are 'data-driven', in the sense that they make very few assumptions about the sensor network that generates the data, and they make no use of knowledge about the physical mechanism that leads to calibration drift. Other methods use, to varying degrees, knowledge about the sensor network and those mechanisms. Furthermore, consideration is given to how those methods can be made 'uncertainty-aware', i.e., to account for the (measurement) uncertainties associated with the different sources of information used by the methods, including measured data obtained from observation and data obtained from physics-based models. Those uncertainties are then used as the basis for evaluating the uncertainties for the estimates of the quantities inferred or predicted by the methods.

Smart buildings: Sensor fusion techniques are also very important in smart building applications. There is a distinct need for guidelines for the validation of models and algorithms as well as for the development of standards and benchmarks. As in the case of industrial manufacturing, the use of Kalman filters is prominent. In addition, Bayesian networks are employed frequently [83].

Environmental monitoring: The role of sensor fusion in environmental/air-quality monitoring needs to be studied further. The main issue is here is to assess whether model-based sensor data assimilation can be considered as a form of sensor fusion. Data assimilation in this context involves estimating the error of a model through the interpolation of sparsely observed errors. In other words, one could say that interpolation tasks are a form of sensor fusion with model outputs. In this case, methods developed for uncertainty propagation and traceability would find direct application.

3.2.4 Summary

Sensor fusion techniques are known to be integral to several domains with sensor network use cases. The applications range from interpolation in the form of deducing parameters at different locations to the use of sensor fusion for drift detection. In order for the developed methods to be applied in a trustworthy manner, ensuring the uncertainty awareness and hence the traceability of the methods is of utmost importance. As such methods will almost certainly involve the use of time-varying quantities, the uncertainty awareness must also account for the dynamic nature of the system. A brief literature review along with basic concepts relevant to sensor fusion and dynamic calibration in sensor networks was presented. A special emphasis was placed on consensus filtering and Kalman filters as commonly used data fusion methods. A brief overview of the basic concepts of dynamic calibration and the estimation of the transfer behavior of a sensor were also presented. Finally, a discussion of a set of real-world use cases and their individual requirements with respect to sensor fusion and dynamic calibration were also provided. The potential application of sensor fusion to sensor network use cases is varied. The combination of data-driven and physics-based models for drift detection was found to be a particularly important subject. In the context of metrology, the propagation of uncertainty to the fused value is of utmost importance.

3.3 Digital Twins and Digital Shadows as Potential Modelling Techniques for Case Studies

This section summarizes different modelling techniques and relevant literature on digital twins and digital shadows. Definitions of digital twins and digital shadows will be given and afterwards the sections will focus on the different project use case and any potential modelling techniques within that.

3.3.1 General definitions

A digital twin is a technology that is more than just the digital representation of the real object; it also enables bi-directional data exchange and real-time management [84]. Therefore, a digital twin differs from models like Building Information Modeling (BIM) models. A Digital Shadow is similar to a digital twin but not as advanced, as it only allows a one-way transfer of data [84].

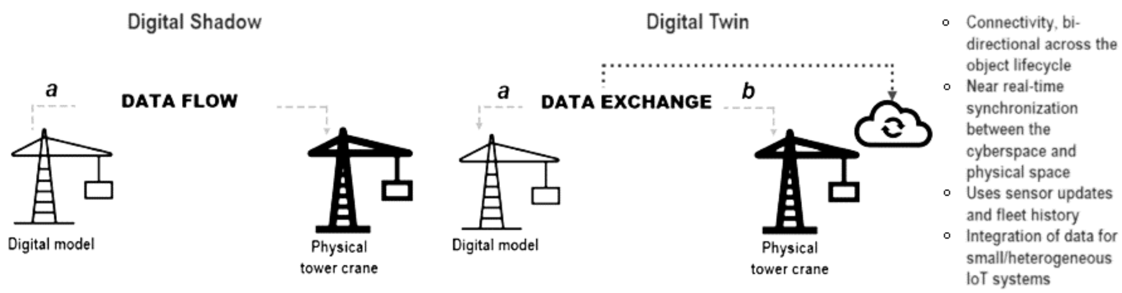


Figure 1: Comparison between digital twin and digital shadow. [84]

In recent years, digital twins have gained increasing attention due to their versatile applications. Digital twins provide benefits throughout the entire lifecycle of a product [85]. Therefore, digital twins are currently used in important industries, including environmental protection, urban management, oil and gas, aerospace, electricity, automotive, healthcare, rail transport, manufacturing, construction, and shipping [86]. In the buildings sector, digital twins are mainly used for four different topics: to optimize the design of a building, to increase the comfort of occupants, to evaluate and increase the building performance, and finally, to simulate and forecast future situations. [87]

The digital representation of a specific object enables optimized decision-making, which can be analyzed in digital space [85]. However, a digital twin not only consists of a model of the real object but also includes a data link between the real and digital object [87]. Therefore, three main elements are required for a digital twin: the real-world entity, a digital representation of the real-world entity, and a linking mechanism that allows an automatic bi-directional data transfer between the two entities. [87]

3.3.2 Potential modelling techniques from literature

In the following, potential modelling techniques, e.g., digital shadow and digital twin and relevant existing bibliography, are described for a few of the use cases used in this guide.

3.3.2.1 Use-Case “Environmental monitoring”

The concept of digital twins is relatively new and involves substantial funding to develop real-time models for environmental applications. Research has predominantly focused on exploring their potential in environmental modeling, yet there are some promising examples where digital twins have proven effective in environmental monitoring.

[88] described the development of a digital twin focused on advanced modeling of soil moisture, river discharge, evaporation, and precipitation. This digital twin collects and disseminates current data from the Mediterranean Basin, encompassing countries like Spain, France, Italy, Greece, Turkey, and others. It serves purposes such as predicting landslides, managing irrigation water resources, and forecasting forest fires. However, the study underscores several challenges, including the need for high-resolution monitoring (1 km, 1 hour) and sophisticated artificial intelligence (AI) to grasp human impacts on hydrological processes, alongside uncertainties in data accuracy.

[89] proposed a similar framework for visualizing environmental sensor data within the context of digital twins, applying it to create a digital twin of Poyang Lake in China. They employed scalar and vector visualization methods to present collected environmental data and utilized video fusion technology for real-time display of environmental surveillance videos. Their study underscored the framework's practical benefits in enhancing the efficiency of Poyang Lake's environmental monitoring, suggesting its adaptability to other lake environments.

[90] explored the creation of a digital twin for air quality monitoring networks in smart cities, relying on mathematical models grounded in differential equations. They highlighted the application of these models for simulating pollutants like carbon monoxide (CO), ozone (O_3), nitrogen oxide (NO), and nitrogen dioxide (NO_2) across spatial and temporal dimensions, contingent upon initial data for predictive purposes. The digital twin operates through continuous data exchange between physical sensors and the digital model, facilitating real-time updates. Expected advantages include predictive maintenance, risk assessment, and operational improvements, though challenges such as sensor data quality and computational limitations are acknowledged.

[91] presented a case study on developing a digital twin for Jakarta, Indonesia, integrating Digital Twin and Mixed Reality technologies to advance Smart City initiatives. This integration enables planners and decision-makers to visualize and implement solutions for optimizing transport routes, implementing greener energy policies in highly polluted areas, and expanding urban green spaces effectively. Their approach leveraged existing datasets such as meteorological records, air quality metrics, and traffic data to build the digital twin infrastructure.

[92] investigated the evolution of digital twins in urban air quality management, emphasizing real-time sensor data integration and predictive modeling. Their research enables cities to simulate air pollution scenarios based on factors like traffic patterns and weather conditions, supporting informed decision-making to mitigate pollution hotspots and enhance public health. For instance, digital twins predict pollutant levels in specific urban zones, guiding policies aimed at reducing emissions and improving air quality standards.

3.3.2.2 Use-case “Smart buildings”

The building sector accounts for nearly one-third of the global final energy consumption [93]. Thus, increasing the operational energy efficiency of buildings is critical to achieving carbon neutrality. Digitalization of buildings can reduce energy consumption by approximately 10 % using real-time data to improve the operational efficiency of buildings, according to the International Energy Agency [94]. Thus, many concepts, such as cyber-physical systems and digital twins, have been proposed [95]. A digital twin differs significantly from a static 3D model derived from building information modeling (BIM) [96].

However, BIM models are currently widely utilized as a basis for the derivation of digital twins [87]. Furthermore, building energy model (BEM) models are derived and combined with GIS datasets in cases of performance simulations for multiple buildings up to the urban level [87]. Furthermore, Internet of Things (IoT) devices such as temperature, humidity, and sensors are widely utilized in literature [87] [97] [98] [99]. The integration of an IoT infrastructure enables a digital twin to process and visualize measurement data [87]. Therefore, IoT is utilized in combination with machine learning to derive a digital twin in the building sector [6,7,8]. In conclusion, in the context of building energy efficiency and control, a BIM model is utilized as a starting point for a digital twin, and cloud computing and Internet-of-things (IoT) technologies are integrated into the digital twin platform [87].

The derivation of a digital twin for buildings remains challenging due to the inherent characteristics of buildings, such as the design differences, the large building sizes, and their long operational period [95]. Yoon [95] divides the process of deriving a digital twin into 3 phases. First, in the design stage, BIM information is utilized in combination with prebuild models and physics-based white box modeling to derive surrogate models based on data-driven methods. These models are calibrated in a later step during their life cycle. Second, intrusive data is collected during commissioning to verify or update the white-box, surrogate, or prebuild models from the design phase. During the third phase, the operation phase, data is obtained non-intrusively from physical sensors. Prebuild models are continuously applied and calibrated with correlational techniques. Thus, the gap between the real building and its digital twin can be reduced so that the real building can be operated optimally [95].

Furthermore, energy-related occupant behavior has a large impact on the predicted and observed energy consumption phase [100]. Thus, there is a need for an intelligent, optimized, and personalized control of the indoor environment that acknowledges the occupant's preferences [101]. Therefore, an intelligent energy management system should communicate with the occupants and have up-to-date information [101]. Furthermore, external conditions such as irradiance and outdoor temperature can play a key role in energy-related decisions and should be processed to make decisions for intelligent energy and comfort management [101].

In [101] divides energy management strategies into three non-mutually exclusive categories: conventional control strategies, intelligent control, and multi-agent-based modeling. Classical controllers encompass P, PI, and PID controllers that are closed-loop controllers. However, these controllers are non-optimal and lead to energy consumption waste [101]. Thus, adaptive controllers were designed, integrating fuzzy logic controllers into the control loop. Similarly, least-square estimations were introduced as an alternative to fuzzy controllers to keep the performance stable when facing uncertainties [101]. However, these controllers depend on the building model, do not have the flexibility to deal with varying occupant comfort, and have limited learning capabilities [101].

Intelligent controllers encompass many different approaches to derive control strategies, such as computational intelligence (CI), including fuzzy logic, artificial neural networks, genetic algorithms, or model-based model predictive control (MPC) [101]. The main features of CI are the learning capability, the interaction with the occupant to receive feedback, their adaptability to the environment, and the ability to operate under uncertainty [101]. MPC is a control strategy that can handle uncertainty in parameters, occupancy, comfort conditions, and weather predictions using dynamic models [101]. Previous outputs of the system are utilized to predict future control signals and optimize them according to an objective [101]. Finally, muti-agent-based modeling techniques are characterized by multiple agents that have the ability to act autonomously. Each agent can cooperate with other agents while fulfilling their specific goal. In intelligent energy system management, a central agent is generally responsible for supervisory control. [101]

In the following, the simulation-based, agent-based and machine learning modelling techniques are summarized.

White-Box / Simulation-Based Modelling

White box modeling, also known as physical or deductive modeling, involves creating a detailed representation of the building's physical characteristics and systems. This method relies on first principles and explicit knowledge of the building's structure, materials, and dynamics. Key features of white box modeling include its transparency, as all aspects of the model are known and can be examined in detail. White box models are highly detail-oriented, including comprehensive information about the building's geometry, thermal properties of materials, HVAC systems, and occupancy patterns. However, this approach requires extensive and accurate data about the building's physical characteristics and operational parameters. The high level of detail ensures high accuracy in simulating building performance, making this method especially suitable for new buildings or those with well-documented characteristics.

Simulation models are closely related to white box modeling in the context of building performance analysis and design. Simulation models are widely used in building automation to create virtual replicas of building systems. These models allow for the testing and analysis of different scenarios without physical implementation. Simulation-based modelling is applied in performance evaluation of building automation system (BAS), scenario analysis, and system design and optimization. The advantages of this technique include providing a risk-free testing environment and the ability to model complex systems. On the downside, it has high computational requirements, and the model's accuracy depends on the quality of the input data.

Modeling the whole building to simulate the overall building performance is commonly done using simulation tools such as EnergyPlus (developed by the U.S. Department of Energy), TRNSYS, DeST, or Modelica [102]. At Forschungszentrum Jülich, Energy Systems Engineering (ICE-1), the focus is currently on the usage of Modelica models, making use of the open source Modelica model library AixLib.

However, the performance of the derived models depends on the input parameters. Inaccurate or incomplete data can lead to erroneous predictions and suboptimal decisions. Therefore, it is crucial to ensure that the models are calibrated with accurate data from the actual building systems. Thus, calibration of the parameters is needed to fit the simulation model to the actual physical system [102]. Additionally, simulation models can be computationally intensive, requiring significant processing power and time, especially for large and complex buildings.

Furthermore, the current physics-based simulation models utilized for digital twins often ignore the object's specific geometric shape and volume [102]. Such dynamics could be integrated by using CFD simulation models. Finally, artificial intelligence can be combined with digital twins to describe the relations between physical parameters and energy consumption more accurately. Thus, the energy efficiency can be improved [102].

According to Pan et al. [102], simulations in digital twins are most commonly applied in the construction and operational phases. In the construction phase, simulations are utilized for monitoring, workers' safety management, and compliance checks of materials [102]. During the operation, a digital twin is utilized to improve the energy efficiency of the building, as well as to monitor thermal management and monitoring [102]. However, according to Pan et al. [102], digital twins are not yet fully used on a large scale.

In addition to traditional simulation tools, there has been a growing interest in using co-simulation approaches to enhance the accuracy and flexibility of building models. Co-simulation involves coupling multiple simulation tools to model different aspects of a building system simultaneously. This approach allows for more detailed and integrated analyses, capturing the interactions between different subsystems more accurately.

Wetter [103] studied the use of the Building Controls Virtual Test Bed (BCVTB), a co-simulation platform that integrates EnergyPlus with other simulation tools such as MATLAB and Simulink. This integration enables the detailed simulation of building energy systems and control strategies, providing a more comprehensive understanding of system performance and potential optimization opportunities.

Agent-based models

Agent-based models (ABM) use autonomous agents to represent individual components within a building system. Each agent follows a set of rules and interacts with other agents, leading to emergent system behavior.

ABM is used for decentralized control systems, occupant behavior modelling, and energy management. Its main advantages are flexibility in modelling diverse behaviors and suitability for complex adaptive systems. However, it requires detailed definition of agent rules and can be computationally expensive.

Wang et al. [104] utilized ABM for modeling occupant behavior and its impact on energy consumption, providing insights into how different usage patterns affect overall building performance. D'Oca and Hong [105] explored the application of ABM in simulating occupant interactions with building systems, demonstrating how these interactions can be optimized to enhance energy efficiency and occupant comfort.

Black box Modelling / Machine Learning Models

In building automation, black box modelling refers to a modelling approach where the internal workings or mechanisms of the system are not explicitly understood or modeled. Instead, the emphasis is on capturing the input-output relationship of the system using empirical data or observations. Black box models are typically used when the underlying processes are complex or not fully understood, and the focus is on predicting outputs based on inputs without delving into the detailed mechanisms.

Machine learning (ML) techniques often function as black box models in building applications. They are nowadays increasingly applied in building automation to analyze large datasets and identify patterns for system optimization and predictive maintenance. Machine learning (ML) techniques are transforming building automation by enabling systems to analyze large datasets, identify patterns, and optimize operations. These models are being applied in fault detection and diagnosis, predictive maintenance, and energy consumption forecasting.

ML models are particularly effective for fault detection, diagnosis and control purposes in BAS. They can process vast amounts of sensor data to identify anomalies that indicate equipment malfunctions or inefficiencies. For instance, supervised learning algorithms such as decision trees and support vector machines (SVMs) can be trained to recognize fault patterns using historical data. Once trained, these models can quickly detect faults in real-time, reducing downtime and maintenance costs.

Predictive maintenance is another critical application of ML in building automation. Traditional maintenance schedules are based on fixed intervals, which can be inefficient. ML models can predict when maintenance is actually needed by analyzing trends in sensor data. Techniques such as regression analysis and neural networks are used to predict the remaining useful life of equipment. This predictive approach minimizes unnecessary maintenance, extends the lifespan of equipment, and reduces costs.

Energy consumption forecasting is essential for optimizing energy use in buildings. ML models, particularly time series forecasting techniques like ARIMA and LSTM networks, can predict future energy consumption based on past usage patterns, weather data, and occupancy information. Accurate energy forecasts enable more effective demand response strategies and energy purchasing decisions, ultimately leading to cost savings and reduced environmental impact.

The primary advantages of ML models include their ability to handle large datasets, adapt to new data, and identify complex, non-linear relationships between variables. However, they require extensive training data and computational resources. The interpretability of ML models can also be challenging, particularly with complex models like deep neural networks.

In [106], a review of the use of machine learning for energy prediction in smart buildings, highlighting how different ML algorithms can accurately forecast energy consumption based on historical data is presented. They found that ML models outperformed traditional statistical methods in terms of accuracy and adaptability. In [107], the application of deep learning techniques for fault detection in HVAC systems, showing how these models can identify anomalies and predict potential failures, thereby reducing maintenance costs and improving system reliability is demonstrated.

Another significant development in ML for building automation is the use of reinforcement learning (RL). RL algorithms can learn optimal control strategies through trial and error, making them suitable for complex systems where predefined control strategies are impractical. In [108], RL to optimize HVAC control, resulting in significant energy savings while maintaining occupant comfort is applied. RL models are particularly valuable in dynamic environments where system behavior changes over time.

Hybrid Models

Hybrid models combine multiple modelling techniques to leverage their strengths and mitigate their weaknesses. These models are particularly useful in capturing the complex interactions within building systems.

Hybrid modelling approaches are applied in integrated building energy management, enhancing system reliability, and providing comprehensive system analysis. The primary advantages include improved accuracy and greater modelling flexibility. However, these models come with increased complexity and require expertise in multiple modelling techniques. In [109], mathematical and simulation models for HVAC control, demonstrating how this hybrid approach can enhance system performance and energy efficiency are combined. Furthermore, hybrid models integrating physical and data-driven approaches for building energy prediction, highlighting how this combination can provide more accurate and robust predictions compared to using a single modelling technique are discussed.

3.3.3 Summary

The modelling techniques for buildings are diverse, each with its own set of advantages and limitations. Mathematical and simulation-based models provide a strong basis for system understanding and performance prediction, while agent-based and machine learning models offer innovative ways to handle complex behaviors and large datasets. Hybrid models present a promising direction for future research by integrating the strengths of various techniques to address the multifaceted challenges in building modelling.

4 Conclusions

This document provides guidance on data quality in sensor networks. It provides an overview of metrics that can be used to assess the data quality of a sensor network. There are many distinct aspects of data quality one can assess, and which metrics are important depends on the use case. Based on the use cases used in the guide it is most important that data is complete (completeness) and correct (accuracy). In addition, it is also important that the data is traceable and not outdated (timeliness).

Many aspects of data quality depend highly on the specific application or use case. One thing is the required number of sensors in a sensor network affecting the spatial coverage whilst the quality of the individual sensors affect the accuracy. In the case of large-scale deployments, the cost of the sensors cannot be too high, which means that the sensors will be of lower quality. In other cases, the aim is typically to employ high-quality sensors.

In sensor networks, lack of data coverage can disrupt reporting. Some common problems can be met with mitigation strategies such as duplication of sensor nodes. Field calibration campaigns can fail to produce sufficient data to give confidence in derived calibration models. This can be mitigated by increasing the duration of the calibration effort. This can be critical if continuous reporting is required since network nodes being used in the field calibration campaign cannot be used at the deployment locations. Accuracy is a mandatory requirement that can be attained to a certain extent with low-cost sensors in a trade off with costs.

When it comes to validating data there are a few common steps. First it is necessary to get a thorough understanding of the data to capture the relationships and intricacies. From here different data quality dimensions can be selected based on their importance and used as the foundation of creating data validation rules. Before creating any rules, it can also be beneficial to define any critical data, i.e. which data is most important for example from a risk perspective. The validation rules should consider the business rules, risks, needs, and expectations of the data consumers. Furthermore, they should be measurable. With a set of measurable rules, it is possible to define metrics to capture these. The process of creating validation rules can be repeated continuously. In the beginning, it is to ensure the right rules are formulated and later for maintaining rules as requirements change or new uses of data are discovered. Implementing data validation rules as an automated process in a given system is desirable to ensure constant validation of new data and due to the volume, speed, and variety of data.

The typical lifecycle of a node is similar in air quality monitoring networks, networks of temperature sensors for district heating, and sensors in smart buildings. Very long lifecycles are expected in sensors in gas flow sensor networks (-20 years), district heating networks (-16 years), and smart building sensor networks (large networks with difficult recalibration processes). The lifecycle of sensor network in heat treatment of high value components is very repetitive and happens under harsh conditions, while network operators are aiming for good performance over an exploitation period of several years.

Three prototypical sensor network scenarios are proposed to formulate guidelines and practical considerations for the application of such methods in real sensor networks. For each scenario, multiple promising and existing co-calibration methods are presented. The advantages of these methods regarding metrologically sound results are briefly discussed for each method, revealing a lack of uncertainty evaluation in many methods. Moreover, general remarks are provided that enhance the data quality of suitable datasets and prepare the automation of in-situ calibration methods. Finally, the applicability in real-world use cases were discussed for three generic scenarios corresponding to common sensor network configurations: dense networks with stationary sensors, sparse networks with stationary sensors and sparse networks with mobile sensors and stationary reference nodes. The latter of the three scenarios was further discussed for the specific case of air-quality monitoring networks. It was shown that the methods for co-calibration and *in situ* calibration must take the mobility of individual sensor nodes into account. The use of low-cost sensors in such networks further increases the need to develop methods for uncertainty-aware sensor fusion, drift detection, dynamic uncertainty estimation, and optimized traceability paths.

Sensor fusion techniques are known to be integral to several domains with sensor network use cases. The applications range from interpolation in the form of deducing parameters at different locations to the use of sensor fusion for drift detection. In order for the developed methods to be applied in a trustworthy manner, ensuring the uncertainty awareness and hence the traceability of the methods is of utmost importance. Since such methods will almost certainly involve the use of time-varying quantities, the uncertainty awareness must also account for the dynamic nature of the system. In a brief literature review special emphasis was placed on consensus filtering and Kalman filters as commonly used data fusion methods. A discussion of a set of real-world use cases and their individual requirements with respect to sensor fusion and dynamic calibration showed the potential application of sensor fusion to sensor network use cases is varied. The combination of data-driven and physics-based models for drift detection was found to be a particularly important subject. In the context of metrology, the propagation of uncertainty to the fused value is of utmost importance.

The modelling techniques for buildings are diverse, each with its own set of advantages and limitations. Mathematical and simulation-based models provide a strong basis for system understanding and performance prediction, while agent-based and machine learning models offer innovative ways to handle complex behaviors and large datasets. Hybrid models present a promising direction for future research by integrating the strengths of various techniques to address the multifaceted challenges in building modelling.

5 References

[1] L. Zhang, D. Jeong and S. Lee, "Data Quality Management in the Internet of Things," *Sensors,* 2021.

[2] H. Veregin, "Data quality parameters," *Geographical information systems,* pp. 177-189, 1999.

[3] S. Geisler, C. Quix, S. Weber and M. Jarke, "Ontology-Based Data Quality Management for Data Streams," *Journal of Data and Information Quality,* pp. 1-34, 2016.

[4] R. Perez-Castillo, C. G. Ana, I. Caballero, M. Rodriguez, M. Piattini, A. Mate, S. Kim and D. Lee, "DAQUA-MASS: An ISO 8000-61 Based Data Quality Management Methodology for Sensor Data," *Sensors,* p. 3105, 2018.

[5] T. C. Redman, "Measuring data accuracy: A framework and review.," in *Information quality*, 2014, pp. 21-36.

[6] H. Cheng, D. Feng, X. Shi and C. Chen, "Data quality analysis and cleaning strategy for wireless sensor networks," *EURASIP Journal on Wireless Communications and Networking,* pp. 1-11, 2018.

[7] A. Klein and W. Lehner, "Representing Data Quality in Sensor Data Streaming Environments," *Journal of Data and Information Quality,* vol. 1, no. 2, pp. 1--28, September 2009.

[8] A. Karkouch, H. Mousannif, H. Al Moatassime and T. Noel, "Data quality in internet of things: A state-of-the-art survey," *Journal of Network and Computer Applications,* vol. 73, pp. 57--81, 9 August 2016.

[9] A. P. Vedurmudi, J. Neumann, M. Gruber and S. Eichstädt, "Semantic description of quality of data in sensor networks," *Sensors,* vol. 21, no. 19, p. 6462, 2021.

[10] DAMA International, Data Management Body of Knowledge - 2nd Edition, Basking Ridge, New Jersey: Technics Publications, 2017.

[11] ClearPoint Strategy, "7 Important Characteristics of Data Quality & Metrics to track," 21 February 2023. [Online]. Available: https://www.clearpointstrategy.com/blog/data-quality-metrics. [Accessed 1 November 2023].

[12] P. Cykana, A. Paul and M. Stern, "DoD Guidelines on Data Quality Management," *IQ,* pp. 154-171, 1996.

[13] S. G. Hackel, F. Härtig, J. Hornig and T. Wiedenhöfer, "The Digital Calibration Certificate," *PTB Mitteilungen,* vol. 127, no. 4, pp. 75--81, December 2017.

[14] G. Kok and P. Harris, "Quantifying Metrological Redundancy in an Industry 4.0 Environment," in *IEEE International Workshop on Metrology for Industry 4.0 & IoT*, Rome, 2020.

[15] S. De Vito, G. D'Elia, S. Ferlito, G. D. Francia, M. Davidović, D. Kleut, D. Stojanović and M. Jovaševic-Stojanović, "A global multi-unit calibration as a method for large scale IoT particulate matter monitoring systems deployments," *IEEE Transactions on Instrumentation and Measurement,* vol. 73, pp. 1--16, 2024.

[16] L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," *Data Science Journal,* vol. 14, 2015.

[17] J. Wang, Y. Liu, P. Li, Z. Lin, S. Sindakis and S. Aggarwal, "Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality," *Journal of the Knowledge Economy,* 2023.

[18] Danish Agency for Digital Government, "Fælles sprog for datakvalitet," 24 September 2021. [Online]. Available: https://arkitektur.digst.dk/metoder/begrebs-og-datametoder/faelles-sprog-datakvalitet. [Accessed 9 April 2024].

[19] *ISO 25012 - Data quality model,* 2008.

[20] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, "Quality Assessment for Linked Data: A Survey'," *Semantic Web,* vol. 7, no. 1, pp. 63-93, 2016.

[21] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems,* vol. 12, no. 4, p. 5–33, 1996.

[22] M. Zio, N. Fursova, T. Gelsema, S. Giessing, U. Guarnera, J. Petrauskienė, L. Quensel, M. Scanu, K. Bosch, M. Loo and K. Walsdorfer, "Methodology for data validation 1.0," European Comission, 2016.

[23] A. P. Vedurmudi, J. Neumann, M. Gruber and S. Eichstädt, "Semantic Description of Quality of Data in Sensor Networks," *Sensors,* vol. 21, no. 19, p. 6462, 28 September 2021.

[24] A. Miquel-Ibarz, J. Burgués and S. Marco, "Global calibration models for temperature-modulated metal oxide gas sensors: A strategy to reduce calibration costs," *Sensors and Actuators B: Chemical,* vol. 350, p. 130769, 1 January 2022.

[25] A. Solórzano, R. Rodríguez-Pérez, M. Padilla, T. Graunke, L. Fernandez, S. Marco and J. Fonollosa, "Multi-unit calibration rejects inherent device variability of chemical sensor arrays," *Sensors and Actuators B: Chemical,* vol. 265, pp. 142--154, 15 July 2018.

[26] BIPM, *The International System of Units (SI),* 2019.

[27] BIPM et al., *International vocabulary of metrology - Basic and general concepts and associated terms (VIM),* 2012.

[28] F. Delaine, B. Lebental and H. Rivano, "In Situ Calibration Algorithms for Environmental Sensor Networks: A Review," *IEEE Sens. J.,* vol. 19, no. 15, pp. 5968-5978, August 2019.

[29] M. Gruber, A. P. Vedurmudi and S. Eichstädt, "Co-Calibration in Distributed Homogeneous Sensor Networks," *Sensor and Measurement Science International, SMSI,* vol. 2023, pp. 47-48, 2023.

[30] K. Whitehouse and D. Culler, "Macro-Calibration in Sensor/Actuator Networks," *Mob. Netw. Appl.,* vol. 8, pp. 463-472, August 2003.

[31] L. Balzano and R. Nowak, "Blind Calibration of Sensor Networks," *IEEE,* pp. 79-88, 2007.

[32] M. Stankovic, S. Stankovic, K. Johansson, M. Beko and L. Camarinha-Matos, "On Consensus-based Distributed Blind Calibration of Sensor Networks," 2018.

[33] M. Gruber, "Consensus-based Online Co-Calibration for Networks of Homogeneous Sensors in IIoT Environments under Consideration of Semantic Knowledge," *Freie Universität Berlin, Berlin, Available,* 2024.

[34] A. Martins, I. Fonseca, J. T. Farinha, J. Reis and A. J. M. Cardoso, "Online Monitoring of Sensor Calibration Status to Support Condition-Based Maintenance," *Sensors,* vol. 23, no. 5, 2023.

[35] J. Grübel, T. Thrash, L. Aguilar, M. Gath-Morad, D. Hélal, R. W. Sumner, C. Hölscher and V. R. Schinazi, "Dense Indoor Sensor Networks: Towards passively sensing human presence with LoRaWAN," *Pervasive and Mobile Computing,* vol. 84, p. 101640, 2022.

[36] M. S. Stanković, S. S. Stanković and K. H. Johansson, "Asynchronous Distributed Blind Calibration of Sensor Networks Under Noisy Measurements," *IEEE Trans. Control Netw. Syst.,* vol. 5, no. 1, pp. 571-582, March 2018.

[37] M. S. Stanković, S. S. Stanković and K. H. Johansson, "Distributed Blind Calibration in Lossy Sensor Networks via Output Synchronization," *IEEE Trans. Autom. Control,* vol. 60, no. 12, pp. 3257-3262, December 2015.

[38] F. Kizel, Y. Etzion, R. Shafran-Nathan, I. Levy, B. Fishbain, A. Bartonova and D. M. Broday, "Node-to-node field calibration of wireless distributed air pollution sensor network," *Environ. Pollut.,* vol. 233, pp. 900-909, February 2018.

[39] A. Swain, E. Abdellatif, A. Mousa and P. Pong, "Sensor Technologies for Transmission and Distribution Systems: A Review of the Latest Developments," *Energies,* vol. 15, no. 19, p. 7339, 2022.

[40] M. Carnero, J. Hernández, M. Sánchez and A. Bandoni, "An Evolutionary Approach for the Design of Nonredundant Sensor Networks,," *Industrial & Engineering Chemistry Research,* vol. 40, no. 23, 23 October 2001.

[41] B. Santos, A. Soares, T.-A. Nguyen, D.-K. Min, J.-W. Lee and F.-A. Silva, "IoT Sensor Networks in Smart Buildings: A Performance Assessment Using Queuing Models," *Sensors,* vol. 21, no. 16, p. 5660, 23 August 2021.

[42] A. Forbes, "Traceable measurements using sensor networks," *Trans. Mach. Learn. Data Min.,* vol. 8, p. 77, 2015.

[43] S. Moltchanov, I. Levy, Y. Etzion, U. Lerner, D. M. Broday and B. Fishbain, "On the feasibility of measuring urban air pollution by wireless distributed sensor networks," *Sci. Total Environ.,* vol. 502, pp. 537-547, January 2015.

[44] W. Tsujita, A. Yoshino, H. Ishida and T. Moriizumi, "Gas sensor network for air-pollution monitoring," *Sens. Actuators B Chem.,* vol. 110, no. 2, pp. 304-311, October 2005.

[45] K. Klauenberg, S. Martens, A. Bošnjaković, M. G. Cox, A. M. H. v. d. Veen and C. Elster, "The GUM perspective on straight-line errors-in-variables regression," *Measurement,* vol. 187, p. 110340, January 2022.

[46] C. Lin, N. Masey, H. Wu, M. Jackson, D. Carruthers, S. Reis, R. Doherty, I. Beverland and M. Heal, "Practical Field Calibration of Portable Monitors for Mobile Measurements of Multiple Air Pollutants," *Atmosphere,* vol. 8, no. 12, December 2017.

[47] L. Sun, D. Westerdahl and Z. Ning, "Development and Evaluation of A Novel and Cost-Effective Approach for Low-Cost NO2 Sensor Drift Correction," *Sensors,* vol. 17, no. 8, August 2017.

[48] V. Bychkovskiy, S. Megerian, D. Estrin and M. Potkonjak, "A Collaborative Approach to In-Place Sensor Calibration," in *Information Processing in Sensor Networks, F. Zhao and L. Guibas, Eds., in Lecture Notes in Computer Science. Berlin*, Berlin Heidelberg, 2003.

[49] C. R. a. Z. N. a. K. A. Martin, R. R. Dickerson, X. Ren, B. N. Turpie and K. J. Weber, "Evaluation and environmental correction of ambient CO2 measurements from a low-cost NDIR sensor," *Atmospheric Meas. Tech.,* vol. 10, no. 7, pp. 2383-2395, July 2017.

[50] E. Miluzzo, N. D. Lane, A. T. Campbell and R. Olfati-Saber, "CaliBree: A Self-calibration System for Mobile Sensor Networks," in *Distributed Computing in Sensor Systems, S. E. Nikoletseas, B. S. Chlebus, D. B. Johnson, and B. Krishnamachari, Eds., in Lecture Notes in Computer Science.*, Berlin Heidelberg, 2008.

[51] D. Hasenfratz, O. Saukh and L. Thiele, "On-the-Fly calibration of low-cost gas sensors," in *presented at the Proceedings of the 9th European conference on Wireless Sensor Networks*, 2012.

[52] O. Saukh, D. Hasenfratz, C. Walser and L. Thiele, "On Rendezvous in Mobile Sensing Networks," in *Real-World Wireless Sensor Networks*, Cham, 2014.

[53] O. Saukh, D. Hasenfratz and L. Thiele, "Reducing multi-hop calibration errors in large-scale mobile sensor networks," in *presented at the Proceedings of the 14th International Conference on Information Processing in Sensor Networks, ACM, Apr. 2015*, 2015.

[54] B. Maag, Z. Zhou, O. Saukh and L. Thiele, "SCAN: Multi-Hop Calibration for Mobile Sensor Arrays," in *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2017.

[55] Airlab, "Mesures et perceptions," 2024. [Online]. Available: https://airlab.solutions/projets/mesures-et-perceptions-184. [Accessed 12 July 2024].

[56] S. Eichstädt, "Analysis of Dynamic Measurements - Evaluation of dynamic measurement uncertainty," 2012.

[57] T. J. Esward, C. Elster and J. P. Hessling, "Analysis of dynamic measurements: New challenges require new solutions," in *XIX IMEKO World Congress on Fundamental and Applied Metrology*, Lisbon, Portugal, 2009.

[58] BIPM, "JCGM, Evaluation of measurement data – Supplement 1 to the “Guide to the Expres-sion of Uncertainty in Measurement” – Propagation of Distributions using a

Monte Carlo Method.," 2008. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf.

[59] J. P. Hessling, "A novel method of evaluating dynamic measurement uncertainty utilizing digital filters," *Meas. Sci. Technol.,* vol. 20, no. 5, p. 055106, April 2009.

[60] W. Elmenreich, "Sensor fusion in time-triggered systems," 2022. [Online]. Available: https://resolver.obvsg.at/urn:nbn:at:at-ubtuw:1-11148.

[61] M. Gruber, W. Pilar von Pilchau, V. Gowtham, N.-S. Koutrakis, M. Riedl, S. Eichstädt, J. Hähner, E. Uhlmann, J. Polte and A. Willner, "Uncertainty-Aware Sensor Fusion in Sensor Networks," *presented at the Sensor and Measurement Science International, Nuremberg, Germany, May 2021,* 2021.

[62] M. Gruber, W. P. von Pilchau, V. Gowtham, N.-S. Koutrakis, N. Schönborn, S. Eichstädt, J. Hähner, M.-I. Corici, T. Magedanz, J. Polte and E. Uhlmann, "Application of Uncertainty-Aware Sensor Fusion in Physical Sensor Networks," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2022.

[63] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar and B. Qureshi, "An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques," *Sensors,* vol. 20, no. 21, 2020.

[64] H. Durrant-Whyte, "Data Fusion in Sensor Networks," in *2006 IEEE International Conference on Video and Signal Based Surveillance*, 2006.

[65] F. Alam, R. Mehmood, I. Katib, N. N. Albogami and A. Albeshri, "Data Fusion and IoT for Smart Ubiquitous Environments: A Survey," *IEEE Access,* vol. 5, pp. 9533-9554, 2017.

[66] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE,* vol. 85, no. 1, pp. 6-23, 1997.

[67] W. Li, Z. Wang, G. Wei, L. Ma, J. Hu and D. Ding, "A Survey on Multisensor Fusion and Consensus Filtering for Sensor Networks," *Discrete Dyn. Nat. Soc.,* vol. 2015, p. 683701, October 2015.

[68] R. Olfati-Saber and J. S. Shamma, "Consensus Filters for Sensor Networks and Distributed Sensor Fusion," in *Proceedings of the 44th IEEE Conference on Decision and Control, 2005*, 2005.

[69] E. Lovisari and S. Zampieri, "Performance metrics in the average consensus problem: A tutorial," *Annu. Rev. Control,* vol. 36, no. 1, pp. 26-41, 2012.

[70] D. Simon, Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches, John Wiley & Sons, 2006.

[71] S. J. J. a. J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE,* vol. 92, no. 3, pp. 401-422, 2004.

[72] BIPM, "Evaluation of Measurement Data - Guide to the Expression of Uncertainty," 2008. [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf.

[73] A. Link, A. Täubner, W. Wabinski, T. Bruns and C. Elster, "Calibration of accel-erometers: determination of amplitude and phase response upon shock excitation," *Meas. Sci. Technol.,* vol. 17, no. 7, p. 1888, June 2006.

[74] M. Jafaripanah, B. M. Al-Hashimi and N. M. White, "Application of analog adaptive filters for dynamic sensor compensation," *IEEE Trans. Instrum. Meas.,* vol. 54, no. 1, pp. 245-251, 2005.

[75] B. Saggin, S. Debei and M. Zaccariotto, "Dynamic error correction of a thermometer for atmospheric measurements," *Measurement,* vol. 30, no. 3, pp. 223-230, 2001.

[76] B. Seeger, L. Klaus and D. Nordmann, "Dynamic calibration of digital angular rate sensors," *ACTA IMEKO,* vol. 9, no. 5, p. 394, December 2020.

[77] S. Eichstädt, K. Ruhm and A. Shestakov, "Dynamic measurement and its relation to metrology, mathematical theory and signal processing: a review," *J. Phys. Conf. Ser.,* vol. 1065, no. 21, p. 212018, August 2018.

[78] T. Esward, S. Eichstädt, I. Smith, T. Bruns, P. Davis and P. Harris, "Estimating dynamic mechanical quantities and their associated uncertainties: application guidance," *Metrologia,* vol. 56, no. 1, p. 015002, February 2019.

[79] P. Ulbrich, F. Franzmann, F. Scheler and W. Schröder-Preikschat, "Design by uncertainty: Towards the use of measurement uncertainty in real-time systems," in *7th IEEE Inter-national Symposium on Industrial Embedded Systems (SIES'12), 2012*, 2012.

[80] P. D. Hale, D. F. Williams and A. Dienstfrey, "Waveform metrology: signal measurements in a modulated world," *Metrologia,* vol. 55, no. 5, p. S135, August 2018.

[81] M. Weber and V. Wilkens, "A Comparison of Different Calibration Techniques for Hydrophones Used in Medical Ultrasonic Field Measurement," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control,* vol. 68, no. 5, pp. 1919-1929, 2021.

[82] Y. Wu and Z. Huang, "Research on Dynamic Calibration Technology of Shock Accelerometer Based on Model Method," *J. Phys. Conf. Ser.,* vol. 2041, no. 1, p. 012002, October 2021.

[83] V. Massimo, A. Carbonari and A. Giretti, "Bayesian Networks for Supporting Model Based Predictive Control of Smart Buildings," 2014.

[84] S. M. E. Sepasgozar, "Differentiating digital twin from digital shadow: Elucidating a paradigm shift to expedite a smart, sustainable built environment," *Buildings,* vol. 11, 2021.

[85] A. Thelen and e. al., "A comprehensive review of digital twin---part 1: modeling and twinning enabling technologies," *Structural and Multidisciplinary Optimization,* vol. 65, p. 12, 2022.

[86] B. Yang, Z. Lv and F. Wang, "Digital twins for intelligent green buildings," *Buildings,* vol. 12, p. 6, 2022.

[87] R. Bortolini, R. Rodrigues, H. Alavi, L. Vecchia and N. Forcada, "Digital twins' applications for building energy efficiency: A review," *Energies,* vol. 15, p. 19, 2022.

[88] L. Brocca, S. Barbetta, S. Camici, L. Ciabatta, J. Dari, P. Filippucci, C. Massari, S. Modanesi, A. Tarpanelli, B. Bonaccorsi, H. Mosaffa and W. Wagner, "A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations," *Frontiers in Science,* vol. 1, 2024.

[89] H. Chen, C. Fang and X. Xiao, "Visualization of Environmental Sensing Data in the Lake-Oriented Digital Twin World: Poyang Lake as an Example," *Remote Sensing,* vol. 15, 2023.

[90] G. Tancev and F. G. Toro, "Towards a Digital Twin for Air Quality Monitoring Networks in Smart Cities," *2022 IEEE International Smart Cities Conference (ISC2),* 2022.

[91] D. Arinsyah, M. Isnan, R. Rahutomo and B. Pardamean, "Digital Twin (DT) Smart City for Air Quality Management," *Procedia Computer Science,* vol. 227, pp. 524-533.

[92] D. Topping, T. J. Bannan, H. Coe, J. Evans, C. Jay, E. Murabito and N. Robinson, "Digital Twins of Urban Air Quality: Opportunities and Challenges," *Frontiers in Sustainable Cities,* vol. 3, 2021.

[93] IEA, "Net Zero by 2050," Paris, 2021. [Online]. Available: https://www.iea.org/reports/net-zero-by-2050.

[94] IEA, "Digitalisation and Energy," Paris, 2017. [Online]. Available: https://www.iea.org/reports/digitalisation-and-energy.

[95] S. Yoon, "Building digital twinning: Data, information, and models," *Journal of Building Engineering,* vol. 76, p. 107021, 2023.

[96] M. a. M. C. C. a. K. V. R. Deng, "From BIM to digital twins: A systematic review of the evolution of intelligent building representations in the AEC-FM industry," *Journal of Information Technology in Construction,* vol. 26, 2021.

[97] I. Martínez, B. Zalba, R. Trillo-Lado, T. Blanco, D. Cambra and R. Casas, "Internet of things (Iot) as sustainable development goals (sdg) enabling technology towards smart readiness indicators (sri) for university buildings," *Sustainability,* vol. 13, no. 14, 2021.

[98] Y.-W. Lin, T. L. E. Tang and C. J. Spanos, "Hybrid approach for digital twins in the built environment.," in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, 2021.

[99] A. Zaballos, A. Briones, A. Massa, P. Centelles and V. Caballero, "A smart campus' digital twin for sustainable comfort monitoring," *Sustainability,* vol. 12, no. 21, 2020.

[100] D. Calì, T. Osterhage, R. Streblow and D. Müller, "Energy performance gap in refurbished German dwellings: Lesson learned from a field test," *Energy and buildings,* vol. 127, pp. 1146-1158, 2016.

[101] F. Mofidi and H. Akbari, "Intelligent buildings: An overview," *Energy and Buildings,* vol. 223, p. 110192, 2020.

[102] Y. Pan, M. Zhu, Y. Lv, Y. Yang, Y. Liang, R. Yin, Y. Yang, X. Jia, X. Wang, F. Zeng, S. Huang, D. Hou, L. Xu, R. Yin and X. Yuan, "Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies," *Advances in Applied Energy,* vol. 10, p. 100135, 2023.

[103] M. Wetter, "Co-simulation of building energy and control systems with the Building Controls Virtual Test Bed," *Journal of Building Performance Simulation,* vol. 4, no. 3, pp. 185-203, 2011.

[104] S. Wang and X. Xu, "Simplified building model for transient thermal performance estimation using GA-based parameter identification," *International Journal of Thermal Sciences,* vol. 45, no. 4, pp. 419-432, 2006.

[105] S. D'Oca and T. Hong, "Occupancy schedules learning process through a data mining framework," *Energy and Buildings,* vol. 88, pp. 395-408, 2014.

[106] T. Ahmad, H. Chen, Y. Guo and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review.," *Energy and Buildings,* vol. 165, pp. 301-320, 2018.

[107] J. Hu, "A deep learning approach for fault diagnosis in HVAC systems," *Automation in Construction,* vol. 104, pp. 329-344, 2019.

[108] Z. Zhang, "Reinforcement learning for energy management in buildings: A survey.," *Buildings,* vol. 10, no. 12, p. 226, 2020.

[109] M. Gouda, "Quasi-adaptive fuzzy heating control of a simulated office using a hybrid mathematical model," *Control Engineering Practice,* vol. 14, no. 7, pp. 759-768, 2006.

[110] L. Zhang, D. Jeong and S. Lee, "Data Quality Management in the Internet of Things," *Sensors,* 2021.

[111] W3C Working Group, *Data on the Web Best Practices: Data Quality Vocabulary,* 2016.

[112] L. C. Günther, E. Colangelo, H.-H. Wiendahl and C. Bauer, "Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises," *Procedia Manufacturing,* vol. 29, pp. 583-591, 2019.

[113] *ISO 8000-61 - Data Quality Management: Process reference model,* 2016.

[114] M. v. Dijk, A. P. Vedurmudi, J. A. Sousa, C. Pires, P. Harris and M. Iturrate-Garcia, "A1.2.1 - Data Quality Metrics for Sensor Networks," 2024.

[115] G. D'Aniello, M. Gaeta and T. Hong, "Effective Quality-Aware Sensor Data Management," *IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE,* vol. 2, pp. 65-77, 2018.

[116] DAMA UK Working Group, "The Six Primary Dimensions for Data Quality Assessment," 2013.

[117] J. Pearce, A. Greenan, A. Smith and C. Elliott, "Relating Composition and Thermoelectric Stability of Pt–Rh Alloy Thermocouples," *Int J Thermophys,* vol. 38:26, 2017.

[118] J. Pearce, "A validated physical model of the thermoelectric drift of Pt-Rh thermocouples above 1200 ℃," *Metrologia,* vol. 57, 2020.

[119] J. Pearce, "Some predictions of a validated physical model of Pt-Rh thermocouple drift above 1200 ℃," *Metrologia,* vol. 58, 2021.

[120] Z. Ni, P. Eriksson, Y. Liu, M. Karlsson and S. Gong, "Improving energy efficiency while preserving historic buildings with digital twins and artificial intelligence," *IOP Conference Series: Earth and Environmental Science,* vol. 863, 2021.

[121] S. Agostinelli, F. Cumo, G. Guidi and C. Tomazzoli, "Cyber-Physical Systems Improving Building Energy Management: Digital Twin and Artificial Intelligence," *Energies,* vol. 14, 2021.

[122] S. Agostinelli, F. Cumo, G. Guidi and C. Tomazzoli, "The potential of digital twin model integrated with artificial intelligence systems," in *2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 2020.

[123] M. A. Zaidan, Y. Xie, N. H. Motlagh, B. Wang, W. Nie, P. Nurmi, S. Tarkoma, T. Petäjä, A. Ding and M. Kulmala, "Dense Air Quality Sensor Networks: Validation, Analysis, and Benefits," *IEEE Sensors Journal,* vol. 22, no. 23, pp. 23507--23520, 2022.

6 Appendix A

FunSNM consortium/stakeholder level survey (A1.2.2) (VINS, FORCE**)**

Dear participants,
Thank you for participating in this consortium/stakeholder level short survey!

The survey aims to elucidate different data requirements during the lifecycle of a typical network node, as well as data requirements for a complete sensor network during the deployment (data coverage needed for metrics, methods, results reporting, physical modelling, possibility of soft sensor support etc.)

Please fill in this survey based on the current practices in sensor networks you have experience with. The goal is to capture current practices and requirements, so that the methods developed in FunSNM can be used where they fit best once they are fully developed.

There are only 5 questions in the survey, but please take your time and be as detailed as possible in your answers.

Question 1:
What is your application of sensor networks?

Question 2:
Describe the different periods/steps of lifecycle of a typical network node in your application of a sensor network.

Question 3:
Describe the data requirements such as amount of needed data, data coverage, other indicators of data quality that are expected during the lifecycle of a sensor node. Use each of the different periods/steps of a lifecycle you have described in the previous question to indicate data requirements. Please be as detailed as possible.

Question 4:
Describe the typical and/or possible use-cases of your sensor network in more detail along with the data requirements that are expected from a deployed sensor network. If applicable/appropriate comment on the physical modelling used, and possibility of soft sensor support.

Question 5:
The extent of needed data is different for different sensor networks, influencing operational costs and loss-benefits considerations. How is this manifested in your considered sensor network example, and what are some mitigation strategies?

7 Appendix B

Overview of sensor type, calibration model and features, performance metrics, training and test datasets used in the global calibration approach in low-cost sensor networks.

	Sensor type	Calibration model	Model features	Performance metrics	Training and test dataset	Highlighted ideas	Research group
$PM_{2.5}$ PM_{10}	Optical nephelometer PMS7003 (Nanchang Panteng Technology Co. Ltd. Plantower, China)	(Huber) linear regression	*LCS PM signal, relative humidity, intercept*	mean absolute error (MAE) and R^2. Both short term (same season) and long term (season to season) performance was estimated.	In field 3 x 10 units x 3 weeks in 2 seasons (winter and summer).	Ten calibration models that were trained on one device, models that were trained on two devices, models that were trained on three devices, and so on. Use of 5 or more units in global model reduces both interquartile and variance intervals in MAE and R2. However, the median of both metrics remains similar.	De Vito et al, 2023
CO	Alphasense B4 Electrochemical	Limited quadratic regression	CO, CO_2, T, T^2, RH, RH^2, CO*T, CO*RH, T*RH, intercept. Note that not all quadratic terms are in the feature set (hence the name limited quadratic regression)	mean normalized bias (MNB), coefficient of variation in the mean absolute error (CvMAE). Pearson linear correlation coefficient (R), precision, explained variance R^2, MAE (ppb), bias (ppb)	In field collocations of the LCS units with regulatory-grade monitors. 75% of LCS units used for training, remaining units for test. Up to 3-4 weeks of training data. 3-75 days testing.		Malings et al, 2019

CO	Temperature-modulated MOX sensor SB-500-12 (FIS Inc., Japan)	Orthogonalized Partial Least-Squares (O-PLS) with Repeated Stratified K-Fold cross-validation for model optimization		Limit of detection (LOD) as per IUPAC definition in low concentration range. Performance of global models built with data from 1 to 4 sensors is tested when applied to unseen sensors.	In lab 6 replicas of a temperature-modulated MOX sensor exposed to gas mixtures of carbon monoxide (range 0–20 ppm) and humid synthetic air (range 20–80% RH at 26 ± 1 ∘C) inside a laboratory controlled gas mixing station.		Miquel-Ibarz et al, 2022
CO_2	nondispersive infrared (NDIR) CO2 sensor, also measures T and RH (SST Sensing, UK)	hybrid random forest–linear regression model	Random forest with inputs from all sensors alongside T and RH, which is for high concentration replaced with linear model (with features single sensor output alongside T and RH).	mean normalized bias (MNB), coefficient of variation in the mean absolute error (CvMAE). Pearson linear correlation coefficient (R), precision, explained variance R^2, MAE (ppb), bias (ppb)	In field collocations of the LCS units with regulatory-grade monitors. 75/% of LCS units used for training, remaining units for test. Up to 3-4 weeks of training data.	If estimated concentration exceeds 90 % of the maximum concentration observed during the training, a linear model is used instead of random forest	Malings et al, 2019 supplement
NO	Electrochemical Alphasense NO-B4	Neural network model with single hidden layer of 20 neurons	inputs from all sensors alongside T and RH	mean normalized bias (MNB), coefficient of variation in the mean absolute error (CvMAE). Pearson linear correlation coefficient (R), precision,	In field collocations of the LCS units with regulatory-grade monitors. 75/% of LCS units used for training, remaining units for test. Up to 3-4 weeks of training data. 4-93 days testing period.	Overhead of using different models can be reduced by using single model architecture for all gas species, e.g. quadratic models or RF with linear model for high concentrations.	Malings et al, 2019

				explained variance R^2, MAE (ppb), bias (ppb)			
NO	Electrochemical Alphasense NO-B4 Each LCS unit consists of four electrochemical sensors: two NO2 sensors and two NO sensors, along with temperature (T) and relative humidity (RH) sensors (Sensirion STH21)	Machine learning based: Multivariate Linear Regression (MLR), Support Vector Regression (SVR), and Random Forest (RF)	Basic model has six features: voltage signals of the 4 electrochemical sensors: NO_A, NO_B, NO2_A, and NO2_B, T and RH. Additionally, improved model includes O3 obtained from nearby monitoring stations.	MAE, R^2, and RMSE	In field 4-5 months of 1h averages. Training data-set split into k-folds (k=5), 1 split used for parameter tuning, and then models were tested on secondary units.	Both models with LCS data inputs only were used, and models with additional O3 from nearby reference. LCS data was standardized using Z scoring.	Abu-Hani et al, 2024
NO_2	electrochemical Alphasense NO2-B43F	hybrid random forest–linear regression model	Random forest with inputs from all sensors alongside T and RH, which is for high concentration replaced with linear model (with features single sensor output alongside T and RH).	mean normalized bias (MNB), coefficient of variation in the mean absolute error (CvMAE). Pearson linear correlation coefficient (R), precision, explained variance R^2, MAE (ppb), bias (ppb)	In field collocations of the LCS units with regulatory-grade monitors. 75/% of LCS units used for training, remaining units for test. Up to 3-4 weeks of training data, testing 4 to 110 days.	If estimated concentration exceeds 90 % of the maximum concentration observed during the training, a linear model is used instead of random forest	Malings et al, 2019
NO_2	electrochemical Alphasense NO2-B43F	Machine learning based: Multivariate Linear	Basic model has six features: voltage signals	MAE, R^2, and RMSE	In field 4-5 months of 1h averages. Training data-set split into k-folds (k=5), 1 split used	Both models with LCS data inputs only were used, and models with additional O3 from nearby	Abu-Hani et al, 2024

	Each LCS unit consists of four electrochemical sensors: two NO_2 sensors and two NO sensors, along with temperature (T) and relative humidity (RH) sensors (Sensirion STH21)	Regression (MLR), Support Vector Regression (SVR), and Random Forest (RF)	of the 4 electrochemical sensors: NO_A, NO_B, NO2_A, and NO2_B, T and RH. Additionally, improved model includes O_3 obtained from nearby monitoring stations.		for parameter tuning, and then models were tested on secondary units.	reference. LCS data was standardized using Z scoring.	
O_3	Alphasense B4 electrochemical	hybrid random forest–linear regression model	Random forest with inputs from all sensors alongside T and RH, which is for high concentration replaced with linear model (with features single sensor output alongside T and RH).	mean normalized bias (MNB), coefficient of variation in the mean absolute error (CvMAE). Pearson linear correlation coefficient (R), precision, explained variance R^2, MAE (ppb), bias (ppb)	In field collocations of the LCS units with regulatory-grade monitors. 75/% of LCS units used for training, remaining units for test. Up to 3-4 weeks of training data. 2-76 days testing period.	If estimated concentration exceeds 90 % of the maximum concentration observed during the training, a linear model is used instead of random forest	Malings et al, 2019
VOCs	MOS gas sensor (SGP40, Sensirion AG, Stäfa, Switzerland)	Deep transfer learning model.	Deep neural network with 10 layers, input is 4 (gas sensitive layers) x1440 array (144 seconds x 10Hz sampling)	RMSE 15 – 40 ppb across various species of VOCs, 110 ppb for CO, 50 ppb for H_2.	In lab multiple unique gas mixtures (UGM) were randomly defined based on predefined concentration distributions with Latin hypercube sampling. In total, 906 UGMs were set, exposing all three SGP40 sensors		Robin et al, 2022

					simultaneously for 1440 seconds, yielding an overall calibration duration of more than 15 days.		
VOCs	Alphasense (UK) photoionization detector	NA	NA	NA	NA	NA	Malings et al, 2019

8 Appendix C

Evaluation of levels for models and digital twins as proposed by Den et al. [5]

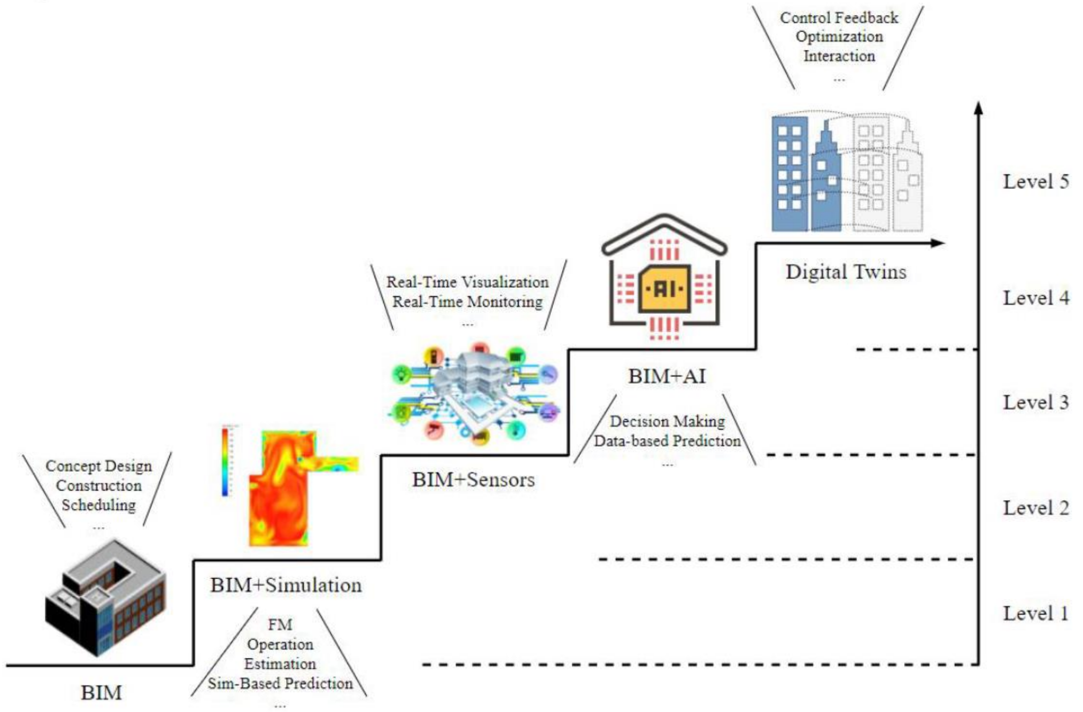


Figure 1: Evolution of BIM to digital twins. Copied from [5]

Framework for deriving a digital twin as proposed by Yoon [14]

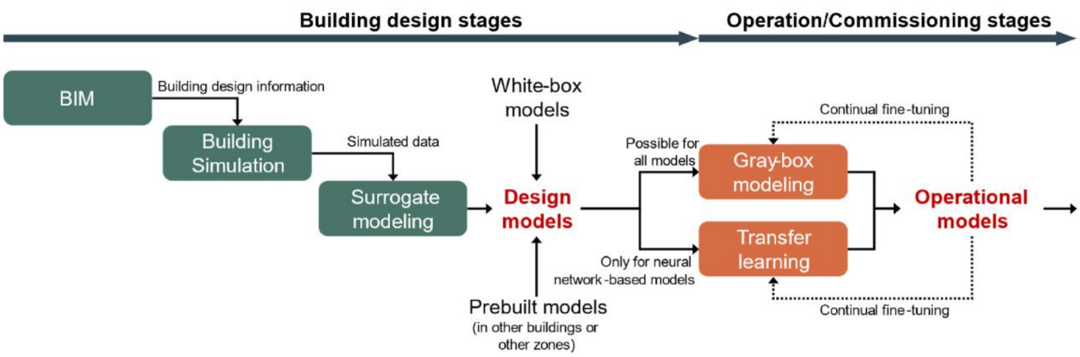


Fig. 4. Transitional process from design models to operational models using the applied techniques.

Figure 2:Copied from [14]

Tools utilized for building simulation according to [17]

Review on the applications of building simulation for operation.

Ref.	Type of buildings	Tools	Application (Case study)			Optimization Strategy
			System	Equipment	Component	
[74]	non-residential building	Modelica	●			· The compressor speed PID-controller parameters, K_p and T_I were optimized.
[75]	educational building	DesignBuilder	●			· Heating and cooling setting temperature of air conditioning were selected as decision-making parameters.
[63]	hotel	DeST	●			· The output schemes of the solar CCHP under climate change were tackled.
[57]	mosque	EnergyPlus	●			· The HVAC setpoint schedule is modified subject to the thermal-comfort threshold based on the temperature response as well as the occupancy prediction.
[76]	educational building	R		●		· Identification for flow rates of chilled water and condensing water, the supplied chilled water temperature, and the cooling tower fan speed.
[62]	metro station	TRNSYS		●		· The chiller loading was optimized by adjusting the set points of the chilled water outlet temperature.
[77]	commercial building	IES-VE		●		· The supply temperature of the AHU and the airflow of VAV are optimized independently.
[78]	–	EnergyPlus, CONTAM, and Matlab			●	· Optimal trajectories of damper angles and fan pressure were determined.
[79]	–	Matlab			●	· The pressure drops of AHU's filters due to clogging were predict.
[80]	data center	TRNSYS			●	· The operation mode (mechanical cooling, partial, free cooling, and free cooling) that can satisfy the cooling requirement and give the best performance was selected.
[81]	office building	EnergyPlus			●	· Window and ventilation supply air fans were controlled in mixed-mode buildings.

The concept of a digital twin according to [17]

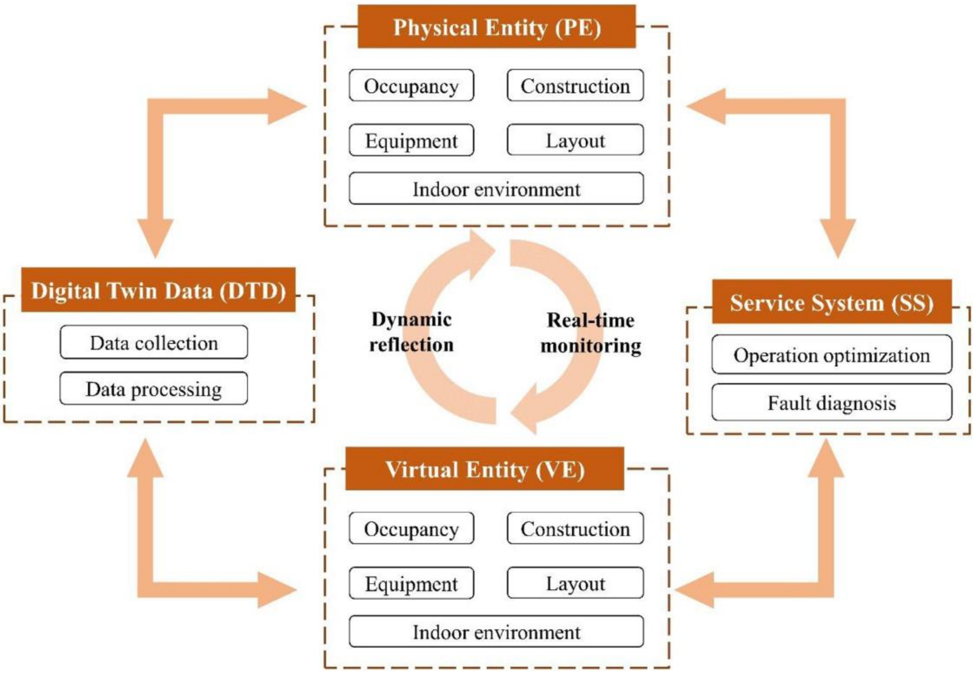
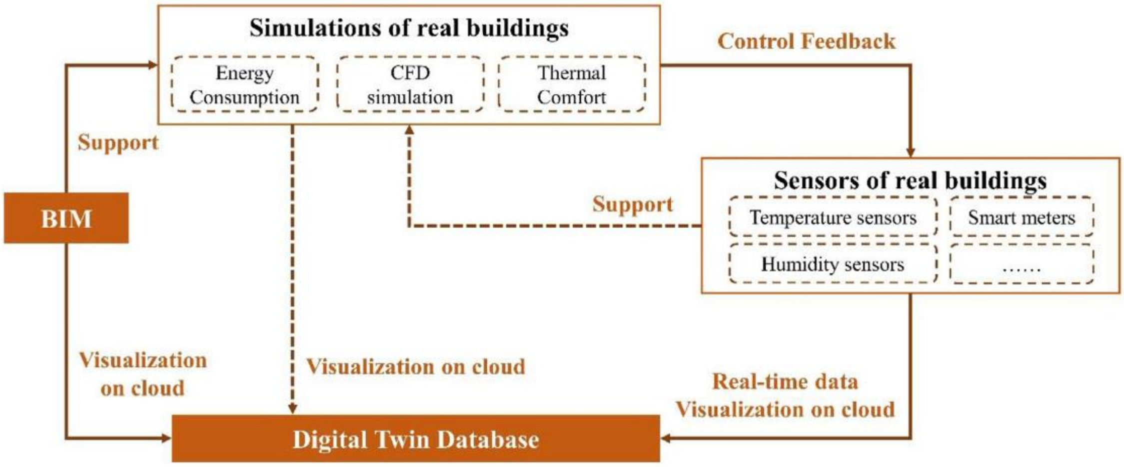


Fig. 7. Concept of digital twins.

The role of simulations in digital twins according to [17]



FunSNM

Fundamental principles of sensor network metrology
www.funsnm.eu