

<https://doi.org/10.1038/s41612-025-01161-2>

A framework for advancing independent air quality sensor measurements via transparent data generating process classification



Sebastian Diez¹✉, Thomas J. Bannan², Miriam Chacón-Mateos³, Pete M. Edwards^{4,5}, Valerio Ferracci⁶, Doğuşhan Kılıç⁷, Alastair C. Lewis⁵, Carl Malings^{8,9}, Nicholas A. Martin⁶, Olalekan Popoola¹⁰, Colleen Rosales^{11,12}, Sean Schmitz¹³, Philipp Schneider¹⁴ & Erika von Schneidemesser¹³

We propose operational definitions and a classification framework for air quality sensor-derived data, thereby aiding users in interpreting and selecting suitable data products for their applications. We focus on differentiating independent sensor measurements (ISM) from other data products, emphasizing transparency and traceability. Recommendations are provided for manufacturers, academia, and standardization bodies to adopt these definitions, fostering data product differentiation and incentivizing the development of more robust, reliable sensor hardware.

Lower-cost sensing technologies are reshaping and expanding the landscape of Earth Observations¹, opening unprecedented opportunities for atmospheric composition measurements². These innovations are relatively more affordable than reference-grade instrumentation, and therefore hold significant importance not only for Low- and Middle-Income Countries where they can enhance often limited air quality (AQ) data capabilities³, but also in underserved regions within wealthier countries that are similarly deprived of adequate AQ infrastructure^{4,5}. We use the more general term “regions” (or areas) to acknowledge that disparities in wealth concentration occur not only between countries, but also within national borders across the globe⁶. Globally, the pervasive shortage of traditional AQ monitoring amplifies the urgency for reliable data to support public health initiatives and policy making^{7–10}.

While sensor systems hold the promise of bridging data gaps related to air pollution sources, pollution distribution and population exposure¹¹—potentially democratizing information and empowering public action¹²—gaps remain between their potential and the real-world performance: sensor-derived data vary significantly in quality,

robustness, and traceability. Unless specified otherwise, we use the term “sensor” to refer to sensor systems (also commonly referred to as “sensor pods”, “sensor devices”, “sensor platforms”) ^{13,14}. Part of these challenges can be attributed to the limitations of detection hardware¹⁵ (e.g., drift, selectivity, cross-interferences, etc.). However, the process by which data outputs are constructed plays a critical but less tangible role. Concerns arise from the increasingly unclear and often untraceable methods employed to manipulate raw sensor signals to derive atmospheric concentrations^{16,17}, leading end-users to be unaware of the assumptions, hypotheses, and limitations of the data products they rely on.

This issue is exacerbated by the growing reliance on advanced data-intensive methods, including Machine Learning (ML), Big Data (BD), and Artificial Intelligence (AI), which, while enhancing processing efficiency, can also obscure the processing steps¹⁸. Additionally, commercial constraints often prevent the sharing of proprietary techniques, further clouding the transparency of how data outputs are derived. While the use of data-centric techniques may enhance performance metrics for a specific

¹Centro de Investigación en Tecnologías para la Sociedad, Universidad del Desarrollo, Santiago, CP 7550000, Chile. ²Department of Earth and Environmental Science, Centre for Atmospheric Science, University of Manchester, Manchester, M13 9PL, UK. ³German Aerospace Center (DLR), Institute of Combustion Technology, 70569 Stuttgart, Germany. ⁴Wolfson Atmospheric Chemistry Laboratories, University of York, York, YO10 5DD, UK. ⁵National Centre for Atmospheric Science, University of York, Heslington, York, YO10 5DD, UK. ⁶Air Quality and Aerosol Metrology Group, National Physical Laboratory, Teddington, TW11 0LW, UK. ⁷National Centre for Atmospheric Science, University of Manchester, Manchester, UK. ⁸GESTAR II Cooperative Agreement, Morgan State University, Baltimore, MD, USA. ⁹NASA Global Modeling & Assimilation Office, Goddard Space Flight Center, Greenbelt, MD, USA. ¹⁰Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom. ¹¹OpenAQ, Washington, DC, 20016, USA. ¹²Air Quality Research Center, University of California, Davis, CA, 95616, USA. ¹³Research, Institute for Sustainability at GFZ, Berliner Straße 130, 14467 Potsdam, Germany. ¹⁴NILU, P.O. Box 100, 2027 Kjeller, Norway. ✉e-mail: sebastian.diez@udd.cl

application (e.g., brand “A” could fairly reproduce a higher-end instrument’s output—in season X, location Y, and time of day Z—for a fraction of its purchasing cost), this improvement may be superficial, as uncertainty can still persist regarding what is actually being reported. As the “opacity” (i.e., complexity and/or lack of transparency) and the processing “weight” (i.e., levels of data manipulation) increase, output data may be transitioning from independent measurements to what is largely a predictive model, and ultimately, data integrity may also be compromised. Therefore, it is crucial to exercise caution to prevent compromising the data information content and its transferability to different use cases (e.g., brand A, but in different seasons, and/or locations).

Hagler et al.¹⁶ and Schneider¹⁷ et al. (2019) initiated discussions on sensor measurements and their transition to “ultra-processed” data products, highlighting the need for transparency in how sensor data is generated, along with its unique challenges and limitations. This concern has also been acknowledged by the CEN/TC264 WG42 Technical Specifications 17660-1:2021¹⁹ and 17660-2:2024²⁰ (hereafter referred to as CEN TSs) explicitly recognize the need for such a framework: “we have to make sure that sensor data will be an independent measurement and not a modelled value. In addition, there is a need for a common terminology and transparency in data processing”.

These discussions set the stage for further exploration and refinement. Building on these insights, this work introduces operational definitions to clarify key aspects of sensor measurements, improve the classification of sensor-derived data products, and develop a practical framework for users to interpret sensor data and select suitable products for their specific applications. If adopted by industry, research groups, and standardization bodies, this classification could enhance trust in sensor technologies, facilitate market differentiation, and support the development of more targeted sensor systems. Ultimately, these efforts aim to foster more robust and reliable sensor systems.

To ensure conceptual clarity and alignment with existing standards, we follow the terminology established in the International Vocabulary of Metrology²¹, which remains the authoritative reference for any related concepts.

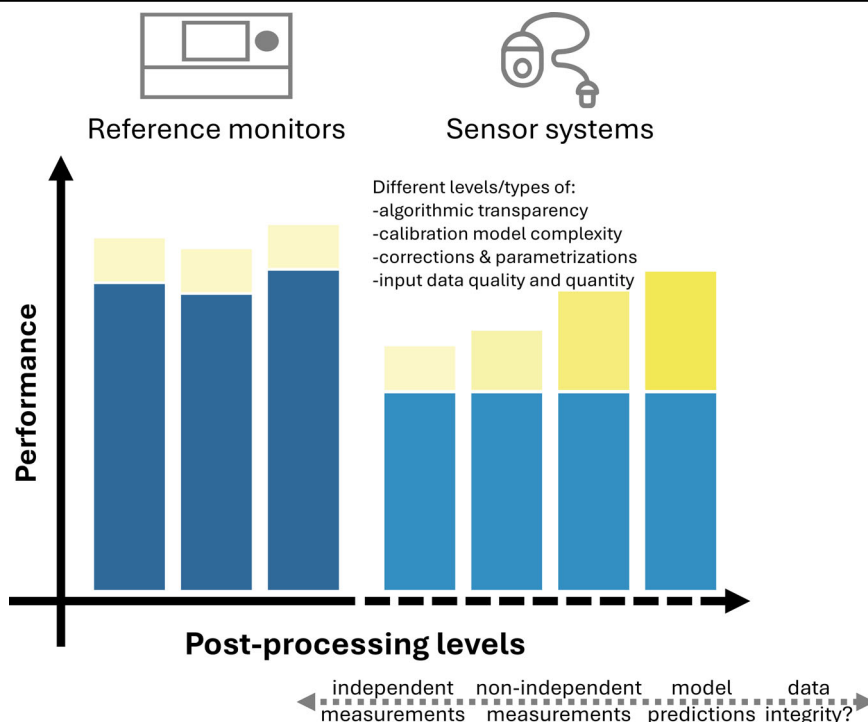
The data generating process

A Data Generating Process (DGP) represents a series of systematic procedures (e.g., statistical, mathematical, computational techniques), employed to transform specific inputs into value-added output data designed to achieve a specific objective. These processes span a continuum of mechanisms, including but not limited to modelling approaches (e.g., physical and statistical models²²), data aggregation techniques (e.g., data fusion and data assimilation²³), and measurements (e.g., direct and indirect methods²⁴). Examples of outputs from these processes include satellite products for gases or aerosols²⁵, air quality indices²⁶, early warning data products²⁷, data fusion techniques for forecasting and mapping applications^{28–31}, among others.

Among the broad spectrum of DGPs, measurements stand out for providing direct, empirical insights about the “real world”³². In the context of regulatory AQ monitoring, ground-based measurements are the only DGP recognized as able to reach the “reference” or “gold standard” status (i.e., used to calibrate, validate, or evaluate other DGP outputs) provided they demonstrate rigour and validity. Cox et al.³³ dissect the measurement process into two distinctive phases: observation and restitution. During the observation phase, a specific set of hardware components (e.g., sensor elements, optical components, sample and signal conditioning components, etc.) generate a measurand-dependent raw signal (e.g., voltage, raw counts). Subsequently, in the restitution phase, this raw signal undergoes processing through algorithmic operations via dedicated software, transforming it into an estimated measurand.

For reference AQ measurement methods, the quality of the hardware is the primary determinant of performance (see Fig. 1). Software conversions, on the other hand, though sometimes opaque to end-users, play an important yet explicit role in the restitution phase: they transform raw signals into interpretable magnitudes. Ranging from simple linear corrections to more elaborated methods, these algorithms aim to compensate for cross-interferences, provided they are thoroughly characterized. In the realm of sensors, however, it is well known that sensing elements are prone to such interferences^{34,35} and their response may change over short periods^{36,37}. Although less tangible, software issues paradoxically emerge

Fig. 1 | Qualitative representation of performance drivers for reference instruments and sensor systems. Higher-end instruments primarily rely on robust hardware (in blue) for their performance, complemented by robust, well-documented and well-justified Quality Assurance and Quality Control (QA/QC) procedures (e.g., zero and span corrections, data validation, continuous system checks, etc.). As a consequence of their lower cost, sensor hardware struggles with pollutant discrimination due to limited selectivity, sensitivity, and stability. QA/QC procedures remain a work in progress, and there is an increasing trend towards the reliance on sophisticated approaches to the restitution phase as a way to improve data quality. As reliance on post-processing algorithms increases (in different grades of yellow), the output data may diverge from a measurement, and in the worst-case scenario, data integrity may be lost. Under these circumstances, improvements in sensor performance may not reflect actual measurement quality. The dashed arrow underneath the figure is intended to illustrate a conceptual transition from independent measurements toward model-derived data products. Side note: the near-identical lengths of blue bars for sensor systems represent manufacturers’ tendency to use similar/identical mass-produced sensing elements.



from attempts to mitigate hardware limitations. Influenced by the overwhelming trend towards adopting data-driven techniques—such as BD, AI, and ML—and driven by commercial interests, the restitution phase becomes increasingly opaque, with raw-signal-to-concentration conversion approaches often resembling a “black box”. For most sensor system companies, their unique intellectual property (IP) relies heavily on software development. Enhancing software, the “peel” of sensor systems, enables rapid deployment and adaptation to market needs and can be more cost-effective than investing in the underlying sensor hardware—the “core” of measurement DGPs. As a result, most commercial sensor systems use similar, if not identical, sensing elements³⁸, potentially suppressing the development of better versions of these components. As previously addressed by Mari et al.²⁴ and Maul et al.³⁹, black-boxing data processing not only limits our understanding of its information content but also fails to capture the relevant characteristics of measurements. Building on their insights, we argue that although no system can capture all relevant features of measurement, increasing opacity can still cause the integrity of the measurement to blur or even be lost (Fig. 1).

There is a wide range of possible DGPs behind AQ sensor data, each carrying its own set of assumptions and hypotheses. Unlike most users of chemical transport models or research-grade instruments, who are usually extensively trained in and thus aware of the measurement method and limitations of the tools they employ, sensor users may not necessarily know the DGP behind their data. We must then recognize the main factors contributing to DGP opacity: (i) lack of transparency, where the process may be entirely or partially obscured, and (ii) algorithmic complexity, where—even when the process is open—the intricate details of the algorithms make it difficult to trace. Additionally, a factor that may further influence both dimensions is the nature of the data sources used (such as type, origin, and relevance), and how these are layered to produce output data. The way these elements are combined and integrated affects the visibility of the underlying mechanisms and the traceability of the computational steps, potentially increasing the DGP overall opacity.

In commercial AQ sensor implementations, the transparency dimension—arising from the proprietary nature of the data handling—serves as the main barrier preventing full disclosure of the “box”, leaving users in the dark about how data are produced, including the calibration methods employed, the variables used, and the extent of data manipulation¹⁴. Conversely, applications originating from academia tend to be less opaque, often employing “open-source” approaches that enhance transparency. In these cases, algorithmic complexity remains the main factor obscuring the box. Here, sophisticated data-driven processing is commonly used to offset hardware constraints (examples can be found in refs. 40–42), sometimes even including information external to the system to enhance data quality (such as in refs. 43–45). Regardless of whether the limiting factor is transparency or complexity, without a clear understanding of the assumptions and hypotheses at each step of the DGP, the credibility of the results cannot be ascertained⁴⁶. These issues compel sensor users to ask critical questions: Is the DGP behind my sensor’s data a measurement? Does my sensor provide independent information? Can I be sure that the integrity of my sensor data has been maintained across the process? Are these questions relevant to my application?

However, efforts to standardize AQ sensor systems have thus far focused on performance evaluation through laboratory type-testing routines and field co-location with reference-grade instruments, as reflected in current standardization efforts—such as the CEN TSs, ASTM standards (i.e., D8406-22⁴⁷ and D8559-24⁴⁸), UK PAS 4023:2023⁴⁹ specification and USEPA guidelines (i.e., EPA/600/R-20/279⁵⁰, EPA/600/R-20/280⁵¹ and EPA/600/R-23/146⁵²). Notably, none of these documents to date address the transparency and traceability of the data generation process within AQ sensor systems.

Predictions, measurements, and independent measurements

In the context of air pollution, data-driven methods (e.g., neural networks, deep learning, etc.) have served as inferential tools⁵³, providing predictive

insights⁵⁴ and revealing hidden patterns⁵⁵. Specifically, for sensor calibration schemes, these methods utilize targeted datasets to develop models aimed at producing accurate measurement outputs³⁴. Although complex statistical models can sometimes outperform linear models⁵⁶, their tendency to learn not only the underlying patterns but also the noise from the training data can severely impair their ability to generalize to new data, leading to over-fitting—when a model becomes overly adapted to its training dataset and fails to generalize to new, unseen data⁵⁷. This issue, compounded by the data context in which they are trained, can also make these models susceptible to concept drift—when the underlying relationships between inputs and outputs change over space and/or time, impairing the model’s ability to maintain performance in new conditions⁵⁸. Such issues are to be expected since these models are built to capture the relational information contained within the inputs, which they do very well. However, unless a human expert intervenes to define the relevant features to be included, these models may not necessarily have the ability to distinguish the sensor dynamics—a fundamental part of the measurement process—from the environment dynamics, essential to any predictive model.

Hagler et al.¹⁶ took a first step in the field by providing general rules for the exclusion of input parameters that can lead to a measurement-prediction transition. Input parameters should be excluded when they are demonstrably not measurement artefacts, based on unverified assumptions, or intended to achieve a desired outcome. Schneider et al.¹⁷ further advanced the discussion by proposing a unified terminology for sensor processing levels, categorizing data from raw sensor outputs (Level 0) to highly processed data products (Level 4). The table provided by the authors serves as a practical and straightforward tool to clarify processing levels, while offering general definitions and examples of the progression from measurement to predictions. However, this framework does not explicitly distinguish between data products that rely solely on internally generated signals and those that incorporate site-specific adjustments or dependencies. These initial contributions have sparked important debates, yet there is still room for deeper examination.

The role of the primary sensor signal in the DGP—particularly during the development and testing of calibration models—is sometimes underestimated. By “primary sensor”, we refer to the main hardware component (i.e., the sensing element) within the sensor system that directly detects the parameter of interest (e.g., in a system designed to measure NO₂; the primary sensor is the one that detects NO₂ concentrations, whereas other components—such as humidity or temperature sensors—may assist in adjusting the signal but play only a secondary role). It is imperative that in a calibration model for a pollutant “X”, data from its primary sensor “x” not only be included but also be the most influential signal (e.g., explaining a significant percentage of the output variance). A model where the most influential signal comes from a secondary sensor “y” (e.g. temperature, humidity, etc.), or that does not even include the sensor signal “x”, transgresses the boundaries towards a prediction.

A challenging yet essential aspect that requires further exploration is the distinction between independent and non-independent measurements. We refer to independent sensor measurements (ISM) as those where the output is primarily determined by the sensor system’s own signals—specifically, the primary sensor’s signal, with correction variables limited to demonstrated artefacts—-independent of the data infrastructure at the application site. Whether a data product incorporates external inputs or utilizes a calibration model tailored for local conditions, it will inevitably introduce a dependency on the local data infrastructure—understood here as the combination of technical resources (e.g., instrumentation, computing capabilities) and qualified personnel dedicated to producing, processing, and analysing high-quality air quality data at a given site. These factors vary significantly from one region to another.

Where external inputs—i.e., data not provided by the sensor system itself but sourced from external entities such as nearby monitors, meteorological stations, or model outputs—are employed to enhance the quality of data products (assuming their inclusion is justified), challenges may arise, especially if the end-user lacks detailed information on how the output data

were produced. For instance, comparing data from two distinct sites, which may differ in the quality or type of external data used, without a clear understanding of the processing involved, can lead to misleading comparisons. Such scenarios can sometimes be akin to comparing apples to oranges, where superficial similarities mask underlying differences.

Complex calibration algorithms used to adjust sensor data towards a “gold standard” reference values under specific conditions are influenced not only by the site-specific meteorological conditions of the training period, but also by the local data infrastructure. As with the inclusion of external variables, making sense of comparisons between two sites can be challenging if the user is not aware of the data product’s limitations. In contrast, models developed independently of the site, though not perfect, strive to be universally applicable⁵⁹. These aim to minimize individual adjustments and enhance adaptability, offering a more traceable approach by ensuring that the same model is applied to the same hardware (e.g., identical units from the same company). This broader application helps mitigate some of the limitations found in site-dependent models.

The ISM concept emphasizes the sensor system’s hardware capabilities over the software used to optimize the output, thereby illuminating the DGP and enhancing its traceability. As processing levels vary significantly between ISM and non-ISM, users should be well-informed about these differences. Ultimately, the choice of which data product to utilize should be carefully informed by the user’s specific requirements.

Towards an operative definition of independent sensor measurements

In regions where air quality measurements are scarce and sensor technology could provide a first glimpse into local pollution dynamics, there remains an urgent need to further refine the concept behind AQ ISM and establish clear criteria that distinguish these from other sensor-derived data products. Current regulatory frameworks, guidelines, and technical standards (such as CEN TSs, ASTM standards and USEPA guidelines) do not address this distinction, highlighting a gap that must be filled to enhance the transparency and reliability of sensor data for specific applications. Building upon the initial definition, we propose the following minimum criteria for a measurement to qualify as an ISM:

1. The concentration output for the target species is mainly determined by the primarily sensor’s raw signal;
2. Only corrections that are relevant to the measurement principle (as outlined by ref. 16) are allowed;
3. Corrections must also be contemporaneous (i.e., on-line) with the measurement (extending Hagler’s point);
4. Data must originate exclusively from the sensor system itself;
5. The calibration model has been developed independently of locally collected environmental data.

It should be noted that ISM criteria does not preclude the possibility of users at the application site making subsequent adjustments to the sensor output data products—such as corrections, calibrations, or aggregations. These are typically transparent post hoc operations (e.g., part of evaluation or reporting workflows) and are generally aligned with standard QA/QC practices conducted on-site, with the responsibility for their implementation resting on the end-users. These adjustments are distinct from the internal data generation pipeline that the ISM framework seeks to characterize.

As discussed in the previous section, the first criterion states that the output concentration must be mainly determined by the raw signal from the primary sensor (e.g., 70% of NO₂ output concentration explained by the NO₂ raw signal), shifting reliance from algorithmic adjustments to hardware quality. By clarifying the basis of reported values, this restriction enhances DGP transparency and incentivizes the use and/or development of more robust hardware (e.g., higher-quality sensing elements), thereby reducing the need for extensive post-processing. Establishing an explainability threshold (e.g., percentage of explained variance), is complex and beyond the scope of this work, yet it remains a critical issue that should be addressed through relevant standards or guidelines.

The second criterion, incorporating unjustifiable parameters into a calibration model, as noted by ref. 16, can result in transitioning from a measurement to a prediction. Furthermore, including more variables not only increases the complexity of the model but also elevates the risk of shifting to an empirically modelled value. Therefore, limiting corrections—both in number and type—to those directly relevant to the specific analyte being measured is key to preserving the integrity of the measurements. The third point builds on the previous one by restricting the use of input variables to those contemporaneous with the measurement. For example, any correction applied to a PM_{2.5} measurement must use signals (such as RH) generated simultaneously with the main sensor signal, to produce the final output. From here on, we use the term “on-line” to refer specifically to this type of contemporaneous data input. The exclusion of retrospective data helps prevent reliance on past trends or patterns (e.g., autocorrelation), which could otherwise shift the DGP focus toward a prediction model.

The fourth criterion requires that data input into the calibration model be generated exclusively by the system itself. Incorporating external data—such as meteorological data, readings from nearby monitoring stations and sensors, or outputs from air quality models—compromises the integrity of measurements by creating dependencies on the quality and quantity of these sources. Ensuring that the sensor DGP relies solely on system-generated data preserves the independence and enhances their comparability across diverse contexts.

Building or re-training a calibration model based on local-to-the-application data (e.g., using reference data from the surroundings where the sensor will be deployed) can bias sensor outputs toward agreement with the used data and create dependency on the quality and quantity of this data, thereby compromising their traceability and transferability. Therefore, the final condition requires that model architectures be constructed independently of the data infrastructure of the application site to prioritize generalizability across different contexts. For instance, two identical systems from the same brand (e.g., sensor systems A1 and A2), each using a site-specific calibration model (e.g., model M1 built in site S1 and model M2 built in site S2), could produce inconsistent results when co-located at a third site (e.g., S3), due to the dependence on training data. To prevent this, calibration models for an ISM should be based on static, well-documented architectures that ensure consistent responses to the same input variables across diverse applications.

Linear adjustments (e.g., zero and span corrections), designed to only correct systematic errors, remain compatible with this approach. These adjustments are typically: (i) external to the data generation process (once the sensor system has produced the output), (ii) transparent (end-users are aware of their application), (iii) reversible (i.e., adjustments do not fundamentally alter the underlying data structure/distribution), and (iv) standardized (e.g., part of QA/QC protocols in regulatory and research contexts). When applied externally and clearly documented as part of QA/QC procedures, such adjustments do not invalidate ISM classification. However, if these corrections are embedded within the sensor system—thus falling outside user control and compromising traceability—they would violate ISM criteria. Conversely, complex (typically nonlinear) corrections applied by users to the already processed out-of-the-box concentration output are discouraged. The risk of overfitting is not the only concern; such adjustments can also undermine the independence of the measurements.

Table 1 introduces a classification system adapted and expanded from¹⁷. Building into this framework, we have incorporated the criteria outlined for ISM, aiming to improve clarity in the distinctions between different sensor data processing levels. The table includes seven columns. The first, “Level”, labels the data processing stage, ranging from raw signals to more processed outputs. The last, “Definition”, provides a brief description of each processing level to clarify the nature of the output. Columns two to five cover the previously described criteria (i.e., main signal, corrections, signal provenance, and model architecture). The sixth column marks whether the data qualifies as an ISM. Each tier in Table 1 represents a range of data processing steps required to produce a sensor data product, spanning from the “raw signal” (level 0) and independent measurements

Table 1 | AQ sensor system data processing classification (adapted from Schneider et al.¹⁷), differentiating independent and non-independent measurements, as well as predictions

Level ^a	Main signal	Corrections ^b	Signals provenance	Model architecture ^c	ISM?	Definition ^d
0	raw signal	-----	-----	-----	-----	Output signal produced by the sensing element (e.g., voltage, current, resistance, etc.), representing the raw, unprocessed data.
1	intermediate geophysical quantities	Pollutant of interest (on-line only)	No compensation	Independent of local data	Yes	Independent measurements derived from Level 0 data, without compensation schemes, mostly using – but not constrained to – basic physical principles or simple calibration equations to obtain the measurand.
2A	standard geophysical quantities	Pollutant of interest (on-line only)	Demonstrated artefact (on-line only)	Independent of local data	Yes	Independent measurements derived from Level 0 or Level 1, using on-board, on-line parameters for demonstrated artefact correction, and basic physical principles or simple calibration equations. More complex calibration models can be used, however on-site model building and/or re-training with local environmental data is not allowed.
Independent / Non-Independent measurement boundary						
2B	standard geophysical quantities (extended B)	Pollutant of interest (on-line only)	Demonstrated artefact (on-line only)	Requires model building/re-training using local data	No	Non-independent measurements derived from Level 0, combined with on-board on-line parameters for artefact correction. Level-1 or level-2 data may also be used as an input. These measurements rely on mostly complex algorithms developed for a specific context and/ or requiring on-site re-training for final application (i.e., relying on local co-located data for calibration).
2C	standard geophysical quantities (extended C)	Pollutant of interest (on-line only)	Demonstrated artefact (on-line only)	Not constrained	No	Non-independent measurements generated using one or more external, on-line data sources for demonstrated artefact correction. These rely on simple or complex algorithms that may have been developed under various data contexts but do not necessarily require on-site re-training. Since these measurements depend on external-to-the-system signals, they remain non-independent.
Measurement / Prediction boundary						
3	advanced geophysical quantities	Not constrained	Unrelated to the measurement principle	Not constrained	No	Prediction generated using parameters (internal or external, on-line or off-line) unrelated to the measurement principle. Data products at this level may use simple or complex algorithms developed in various data contexts, which may or may not require on-site retraining.
4	spatially continuous geophysical quantities	Not constrained	Not constrained	Not constrained	No	Similar to Level 3 but providing a spatially continuous map. This data product may be derived from sensor networks alone or in combination with satellite data and/or model outputs (e.g., chemical transport models).

^a Most processing levels were retained from Schneider et al.¹⁷. These authors refer to Level 0 as “raw measurement”, however, we have labelled this column “raw signal” as measurement not only involves producing the signal, but also transforming the signal into a specific magnitude.

^b Adjustment: parameters are defined as per Hagler et al.¹⁶.

^c Local linear corrections (such as zero and span adjustments) can be applied at any data processing level.

^d With the exception of the main signal, which must be generated internally within the system.

^e We use “on-line” to refer to variables or corrections that are contemporaneous with the measurement, as opposed to those based on retrospective inputs.

(levels 1 and 2) to increasingly complex data products. At higher levels, such as those involving advanced architectures with additional internal or external variables, the outputs transition to non-independent measurements (levels 2B and 2C) and predictions (level 3). Importantly, this classification is concerned solely with the extent of data processing and is not indicative of the data quality itself. The assessment of data quality provided by a sensor system should be conducted through a separate, independent process (e.g., via evaluation tests as specified in the CEN TS's). Ultimately, the choice of processing level and data quality will depend on the user's specific measurement objectives. The proposed ISM framework is not intended to challenge or replace existing regulatory procedures, but to complement them by increasing transparency in how sensor-derived data products are generated.

Recommendations and path forward

In the context of regulatory AQ monitoring, ground-based measurements are the sole DGP currently recognized as capable of reaching the “gold standard” status, as they yield direct, empirical insights into the real world. Although sensor data may never attain such high status, they still hold the potential. And “potential” might represent “hope” for underserved regions. Although this perspective may seem overly optimistic, such aspirations could serve as a driving force toward improved sensor technologies. It then becomes paramount for the global AQ community to strongly advocate for, and actively support the development and refinement of affordable measurement solutions, while contributing to building technical capabilities where they are most needed.

Although current sensor technologies—both hardware and software—have room for improvement, they can still provide valuable insights when used correctly and when questions posed to their data are formulated appropriately⁶⁰. However, the opacity of the DGP in these devices, whether commercial or not, presents significant challenges, often compromising the integrity and reliability of the data they produce. Constrained by either the IP and/or the complexity of the algorithmic methods employed, sensor users should be informed of at least the basic building elements behind their data and must be aware of the data product's scope and associated limitations, such as potential hidden uncertainties and the possible degradation of measurement independence. Current practices tend to fall short in addressing these issues.

In the short-term, manufacturers are encouraged to identify their DGP as accurately as possible and provide alternatives to black box models. Similarly, scientists who develop their own sensor systems should clearly specify the DGP behind these systems and continue to advocate for greater openness by offering more explainable methods and promoting their wide dissemination. Both manufacturers and the scientific community are urged to describe their data products in greater detail, for instance, according to Table 1. Implementing a labelling system such as the one outlined in Table 1 would represent a significant step toward enhancing data transparency. Categorizing sensor data products by their processing level and explicitly indicating whether they are independent or not enables users to interpret data more effectively while preserving manufacturers' IP. Recognizing the elements that constitute an ISM allows users to better understand the applications for which a sensor system is suited, thereby boosting confidence in the provided information. Explicitly stating these characteristics can also allow sensor vendors to differentiate their products in today's crowded marketplace. Moreover, this approach may motivate manufacturers to upgrade their products to meet ISM criteria, potentially leading to technological innovation and spurring competition among them. Additionally, this framework could act as a market driver, incentivizing the development of products that rely less on extensive data processing.

In the next few years, the focus should intensify on enhancing sensor hardware technology, aiming to reduce reliance on software-driven adjustments and promote independent measurements at a mid-range price point. Investment from a broad range of actors—including philanthropic initiatives, government agencies, the private sector, and international development funds—could significantly catalyze innovations in this

area. Coordination of efforts across sectors should be facilitated through regional and international initiatives—such as FAIRMODE WG6 (<https://fairmode.jrc.ec.europa.eu/activity/ct6>), Allin-Wayra (<https://igacproject.org/activities/allin-wayra-small-sensors-atmospheric-science>), among others—via workshops, conferences, and technical working groups. Furthermore, implementing such investment-driven initiatives could have a profound impact in underserved markets, particularly in Low- and Middle-Income Countries, where unique challenges such as limited infrastructure and poorly characterized pollutant profiles require tailored sensor solutions for a rapidly growing market. In these settings, measurement instruments are often required to operate autonomously, making the provision of fit-for-purpose, independent measurements a significant advantage. By offering mid-tier hardware solutions, manufacturers can meet these needs with more affordable alternatives to traditional analytical instruments, while also reducing reliance on complex, data-intensive processing methods.

Compliance with the framework presented here can be demonstrated either voluntarily or through third-party evaluations, with the possibility of certification by agencies if it is incorporated into regulatory requirements. This proposal is another step forward in the ongoing effort to enhance the value of sensor data. Successfully implementing such a system will require a consensus among key stakeholders, including manufacturers, regulatory agencies, standardization bodies (e.g., CEN/TC264 WG42), scientists, and users. Ideally, discussions about this framework will lead to the adoption of an industry standard that balances adaptability with the evolving demands of end-users. Cross-sector collaboration will be key to transforming this proposal into a practical and widely accepted framework.

Data availability

No datasets were generated or analysed during the current study.

Received: 28 March 2025; Accepted: 30 June 2025;

Published online: 29 July 2025

References

- Stephens, G. et al. The emerging technological revolution in earth observations. *Bull. Am. Meteorol. Soc.* **101**, E274–E285 (2020).
- Rai, A. C. et al. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci. Total Environ.* **607–608**, 691–705 (2017).
- Pinder, R. W. et al. Opportunities and challenges for filling the air quality data gap in low- and middle-income countries. *Atmos. Environ.* **215**, 116794 (2019).
- Apte, J. S. et al. High-resolution air pollution mapping with google street view cars: exploiting big data. *Environ. Sci. Technol.* **51**, 6999–7008 (2017).
- Martin, R. V. et al. No one knows which city has the highest concentration of fine particulate matter. *Atmos. Environ. X* **3**, 100040 (2019).
- Pfeffer, F. T. & Waitkus, N. The wealth inequality of nations. *Am. Sociol. Rev.* **86**, 567–602 (2021).
- Brunt, H. et al. Air pollution, deprivation and health: understanding relationships to add value to local air quality management policy and practice in Wales, UK. *J. Public Health* **39**, 485–497 (2017).
- Ferguson, L. et al. Exposure to indoor air pollution across socio-economic groups in high-income countries: a scoping review of the literature and a modelling methodology. *Environ. Int.* **143**, 105748 (2020).
- Terrell, K. A. & Julien, G. S. Air pollution is linked to higher cancer rates among black or impoverished communities in Louisiana. *Environ. Res. Lett.* **17**, 014033 (2022).
- Wodtke, G. T., Ard, K., Bullock, C., White, K. & Priem, B. Concentrated poverty, ambient air pollution, and child cognitive development. *Sci. Adv.* **8**, eadd0285 (2022).
- Gulia, S., Khanna, I., Shukla, K. & Khare, M. Ambient air pollutant monitoring and analysis protocol for low and middle income

- countries: an element of comprehensive urban air quality management framework. *Atmos. Environ.* **222**, 117120 (2020).
12. Stevens, K. A., Bryer, T. A. & Yu, H. Air quality enhancement districts: democratizing data to improve respiratory health. *J. Environ. Stud. Sci.* **11**, 702–707 (2021).
 13. Diez, S. et al. Long-term evaluation of commercial air quality sensors: an overview from the QUANT (Quantification of utility of atmospheric network technologies) study. *Atmos. Meas. Tech.* **17**, 3809–3827 (2024).
 14. Karagulian, F. et al. Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere* **10**, 506 (2019).
 15. Williams, D. E. Electrochemical sensors for environmental gas analysis. *Curr. Opin. Electrochem.* **22**, 145–153 (2020).
 16. Hagler, G. S. W., Williams, R., Papapostolou, V. & Polidori, A. Air quality sensors and data adjustment algorithms: when is it no longer a measurement? *Environ. Sci. Technol.* **52**, 5530–5531 (2018).
 17. Schneider, P. et al. Toward a unified terminology of processing levels for low-cost air-quality sensors. *Environ. Sci. Technol.* **53**, 8485–8487 (2019).
 18. Kar, A. K. & Dwivedi, Y. K. Theory building with big data-driven research – moving away from the “what” towards the “Why”. *Int. J. Inf. Manag.* **54**, 102205 (2020).
 19. CEN. *CEN/TS 17660-1 Air Quality—Performance Evaluation Of Air Quality Sensor Systems - Part 1: Gaseous Pollutants In Ambient Air*. <https://standards.iteh.ai> (2021).
 20. CEN. *CEN/TS 17660-2 Air Quality—Performance Evaluation Of Air Quality Sensor Systems—Part 2: Particulate Matter In Ambient Air*. <https://standards.iteh.ai> (2024).
 21. JCGM. *The International Vocabulary Of Metrology—basic And General Concepts And Associated Terms (VIM)*. https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf (2012).
 22. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* 2nd edn, Vol. 1232 (Wiley, 2016).
 23. Gettelman, A. et al. The future of Earth system prediction: Advances in model-data fusion. *Sci. Adv.* **8**, eabn3488 (2022).
 24. Mari, L., Wilson, M. & Maul, A. Modeling measurement and its quality. In *Measurement Across the Sciences: Developing a Shared Concept System for Measurement* (eds. Mari, L., Wilson, M. & Maul, A.) 213–264 (Springer International Publishing, Cham, 2023).
 25. van Donkelaar, A., Martin, R. V., Brauer, M. & Boys, B. L. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* **123**, 135–143 (2015).
 26. Tan, X. et al. A review of current air quality indexes and improvements under the multi-contaminant air pollution exposure. *J. Environ. Manag.* **279**, 111681 (2021).
 27. Mo, X., Zhang, L., Li, H. & Qu, Z. A novel air quality early-warning system based on artificial intelligence. *Int. J. Environ. Res. Public Health* **16**, 3505 (2019).
 28. Gressent, A., Malherbe, L., Colette, A., Rollin, H. & Scimia, R. Data fusion for air quality mapping using low-cost sensor observations: feasibility and added-value. *Environ. Int.* **143**, 105965 (2020).
 29. Lin, Y.-C., Chi, W.-J. & Lin, Y.-Q. The improvement of spatial-temporal resolution of PM_{2.5} estimation based on micro-air quality sensors by using data fusion technique. *Environ. Int.* **134**, 105305 (2020).
 30. Schneider, P. et al. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environ. Int.* **106**, 234–247 (2017).
 31. Weissert, L. F. et al. Low-cost sensors and microscale land use regression: data fusion to resolve air quality variations with high spatial and temporal resolution. *Atmos. Environ.* **213**, 285–295 (2019).
 32. Finkelstein, L. Widely-defined measurement – An analysis of challenges. *Measurement* **42**, 1270–1277 (2009).
 33. Cox, M. G., Rossi, G. B., Harris, P. M. & Forbes, A. A probabilistic approach to the analysis of measurement processes. *Metrologia* **45**, 493 (2008).
 34. Liang, L. Calibrating low-cost sensors for ambient air monitoring: techniques, trends, and challenges. *Environ. Res.* **197**, 111163 (2021).
 35. Pang, X., Shaw, M. D., Gillot, S. & Lewis, A. C. The impacts of water vapour and co-pollutants on the performance of electrochemical gas sensors used for air quality monitoring. *Sens. Actuators B Chem.* **266**, 674–684 (2018).
 36. Malings, C. et al. Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation. *Aerosol Sci. Technol.* **54**, 160–174 (2020).
 37. Sayahi, T., Butterfield, A. & Kelly, K. E. Long-term field evaluation of the plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **245**, 932–940 (2019).
 38. Diez, S., Lacy, S., Urquiza, J. & Edwards, P. QUANT: a long-term multi-city commercial air sensor dataset for performance evaluation. *Sci. Data* **11**, 904 (2024).
 39. Maul, A., Mari, L., Torres Iribarra, D. & Wilson, M. The quality of measurement results in terms of the structural features of the measurement process. *Measurement* **116**, 611–620 (2018).
 40. Malings, C. et al. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmos. Meas. Tech.* **12**, 903–920 (2019).
 41. Zimmerman, N. et al. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* **11**, 291–313 (2018).
 42. van Zoest, V., Osei, F. B., Stein, A. & Hoek, G. Calibration of low-cost NO₂ sensors in an urban air quality network. *Atmos. Environ.* **210**, 66–75 (2019).
 43. Abu-Hani, A., Chen, J., Balamurugan, V., Wenzel, A. & Bigi, A. Transferability of machine-learning-based global calibration models for NO₂ and NO low-cost sensors. *Atmos. Meas. Tech.* **17**, 3917–3931 (2024).
 44. Munir, S., Mayfield, M., Coca, D., Jubba, S. A. & Osammor, O. Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities – a case study in sheffield. *Environ. Monit. Assess.* **191**, 94 (2019).
 45. Okafor, N. U., Alghorani, Y. & Delaney, D. T. Improving data quality of low-cost iot sensors in environmental monitoring networks using data fusion and machine learning approach. *ICT Express* **6**, 220–228 (2020).
 46. Williams, D. E. Low cost sensor networks: how do we know the data are reliable?. *ACS Sens* **4**, 2558–2565 (2019).
 47. ASTM. *ASTM D8406-22: Standard Practice for Performance Evaluation of Ambient Outdoor Air Quality Sensors and Sensor-based Instruments for Portable and Fixed-point Measurement*. <https://store.astm.org/d8406-22.html> (2022).
 48. ASTM. *ASTM D8559-24: Standard Specification for Ambient Outdoor Air Quality Sensors and Sensor-based Instruments for Portable and Fixed-Point Measurement*. <https://store.astm.org/d8406-22.html> (2024).
 49. PAS 4023. *Selection, Deployment, And Quality Control Of Low-cost Air Quality Sensor Systems In Outdoor Ambient Air—Code Of Practice*. <https://www.en-standard.eu/> (2023).
 50. Duvall et al. *Performance Testing Protocols, Metrics, and Target Values for Ozone Air Sensors: Use in Ambient, Outdoor, Fixed Site, Non-Regulatory and Informational Monitoring Applications*. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=CEMM&dirEntryID=350784 (2021).
 51. Duvall et al. *Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors: Use in Ambient, Outdoor, Fixed Site, Non-Regulatory Supplemental and Informational Monitoring Applications*. https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=350785&Lab=CEMM (2021).
 52. Duvall, R. et al. *NO₂, CO, and SO₂ Supplement to the 2021 Report on Performance Testing Protocols, Metrics, and Target Values for Ozone Air Sensors*. https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=360578&Lab=CEMM (2024).

53. Liu, X., Lu, D., Zhang, A., Liu, Q. & Jiang, G. Data-driven machine learning in environmental pollution: gains and problems. *Environ. Sci. Technol.* **56**, 2124–2133 (2022).
54. Concas, F. et al. Low-cost outdoor air quality monitoring and sensor calibration: a survey and critical analysis. *ACM Trans. Sens. Netw.* **17**, 1–44 (2021).
55. Kumar, K. & Pande, B. P. Air pollution prediction with machine learning: a case study of Indian cities. *Int. J. Environ. Sci. Technol.* **20**, 5333–5348 (2023).
56. Topalović, D. B. et al. In search of an optimal in-field calibration method of low-cost gas sensors for ambient air pollutants: comparison of linear, multilinear and artificial neural network approaches. *Atmos. Environ.* **213**, 640–658 (2019).
57. deSouza, P. et al. Calibrating networks of low-cost air quality sensors. *Atmos. Meas. Tech.* **15**, 6309–6328 (2022).
58. De Vito, S., Esposito, E., Castell, N., Schneider, P. & Bartonova, A. On the robustness of field calibration for smart air quality monitors. *Sens. Actuators B Chem.* **310**, 127869 (2020).
59. Barkjohn, K. K., Gantt, B. & Clements, A. L. Development and application of a United States-wide correction for PM_{2.5} data collected with the PurpleAir sensor. *Atmos. Meas. Tech.* **14**, 4617–4637 (2021).
60. Diez, S. et al. Air pollution measurement errors: is your data fit for purpose?. *Atmos. Meas. Tech.* **15**, 4091–4105 (2022).

Acknowledgements

This work was funded by the National Agency for Research and Development (ANID), through the FONDEQUIP Mayor Fund, under Grant No. EQY200021. The authors would like to express their sincere gratitude to Priscilla Adong, Alexandre Caseiro, Gayle Hagler, David Harrison, Stuart Lacy, Gustavo Olivares, Pallavi Pant, Jorge Saturno, Saumya Singh, and Brian Stacey for their valuable contributions and constructive feedback during the development of this work.

Author contributions

S.D. conceived the study and led the manuscript writing. S.D. wrote the initial draft and handled all iterations of text refinement. S.D. developed the

classification scheme, with feedback and suggestions from P.S., C.M., and M.C.-M. Visualizations were prepared by S.D. Definitions were refined by S.D., M.C.-M., C.M., O.P., P.S., and E.v.S. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Sebastian Diez.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025