sensors

MDPI

Article

Fuzzy Fault Detection and Band-Conditioned Data Quality Assessment for Sensor Streams: A Continuous Glucose Monitoring Case Study

Juan F. Gaitán-Guerrero [1,*], Carmen Martínez-Cruz [2], Macarena Espinilla [1] and José L. López [1]

[1] Department of Computer Science, University of Jaén, Campus Las Lagunillas, 23071 Jaén, Spain; mestevez@ujaen.es (M.E.); llopez@ujaen.es (J.L.L.)
[2] Department of Languages and Computer Systems, University of Granada, 18071 Granada, Spain; cmcruz@ugr.es
* Correspondence: jgaitan@ujaen.es

Abstract

Sensor faults can propagate beyond raw measurements and distort the aggregates used to interpret a monitored phenomenon. We propose a domain-agnostic, five-layer fuzzy pipeline that characterizes the reliability of a sensor stream jointly over time and value bands and carries it through to every derived aggregate: (L1) fault tagging of each sample against five fuzzy failure primitives; (L2) fuzzy windowing; (L3) band reliability; (L4) routed repair with propagated reconstruction uncertainty; and (L5) veracity aggregation with linguistic protoforms. An end-to-end acquisition architecture is also specified. The pipeline was instantiated for continuous glucose monitoring (CGM) with clinical knowledge of glucose dynamics and applied to a 14-month recording of an adult with type 1 diabetes (FreeStyle Libre 3, 5-min cadence), of which 294 days contained at least 120 readings. Clean reference days remained essentially unchanged (mean sample quality 0.98–0.99). On the most corrupted day, the estimated time below range fell from 16% to 3%, with a veracity of only 0.14. Over the recording, 38% of the days were too unreliable to phenotype and 5% changed phenotype. Joint time- and band-conditioned reliability assessment thus prevents corrupted sensor data from producing falsely precise clinical conclusions.

Keywords: data reliability; sensor monitoring; fuzzy logic; linguistic protoforms; uncertainty-aware aggregation; continuous glucose monitoring

Check for updates

Academic Editors: Zaib Ullah and Muddasar Naeem

Received: 7 September 2026
Revised: 6 October 2026
Accepted: 8 October 2026
Published: 12 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Diabetes management increasingly relies on continuous glucose monitoring (CGM). Its minute-by-minute readings are condensed into a small set of aggregates that comprise time in range (TIR), time below and above range (TBR, TAR), mean glucose and the glucose management indicator, which now constitute standard clinical targets and drive therapeutic decisions [1]. Yet, CGM sensors are imperfect: they leave gaps during warm-up, disconnection or compression; they saturate against the lower bound of the assay; they stick at a constant value; and they emit physiologically impossible excursions [2]. These faults are rarely confined to the raw signal. Once undetected, they flow through cleaning, aggregation and reporting and silently distort the very metrics on which treatment rests: a few spurious low readings inflate the TBR and can fabricate an episode of hypoglycaemia, while an unnoticed gap leaves the TIR simply unknowable for part of the day; the metrics

themselves, moreover, are known to vary with the sensor model and to carry a systematic bias [3,4].

A central observation motivates this work: the reliability that results from these faults is not a single number attached to a series or a day, but a field defined over two axes: time and glucose range. The same day of data can be perfectly trustworthy for detecting hyperglycemia while being unusable for hypoglycaemia, simply because the corrupting fault falls in the low range during a particular interval. Current practice ignores this structure: the international consensus on CGM interpretation gates analysis on a single crisp criterion—at least 70% of sensor-active time over 14 days [1]—which neither localizes unreliability in time nor conditions it on the glucose range it affects.

Neither these failure modes nor this blind spot, however, are specific to glucose: missing values, anomalous readings, and the silent way they contaminate downstream aggregates recur across environmental sensing [5,6], structural health monitoring [7], wearable physiology and the industrial internet of things (IoT), where sensor faults degrade the service quality of deployed networks [8] and dropouts corrupt the monitoring of offshore installations [9]. Each of these fields has confronted a piece of the same problem under its own name and in isolation from the others.

In biomedical signal processing, the dominant response has been to clean the stream itself: a smart sensor cascade of denoising, enhancement and prediction placed after a commercial CGM device [2,10], driven by a Kalman filter that adapts to a signal-to-noise ratio varying across sensors and individuals [11], with later variants tackling coloured, non-stationary noise or reconstructing corrupted stretches outright [12,13]. A parallel effort detects faults rather than removing them, modeling transient disconnections and compression artifacts [14] and flagging sensor failure through statistical or clustering criteria [15–17]; in wearable monitoring, the same idea takes the form of signal-quality indices that grade each segment and route the processing that follows [18,19]. Such methods work at the sample level and serve real-time use, i.e., alarms, dosing, and control, but they do not carry a judgment of reliability forward to the aggregates that summarize a day, nor put it in a form a clinician can read.

A largely separate tradition formalizes data quality in the abstract, as a set of dimensions such as completeness, accuracy, consistency, timeliness and, more recently, as a context-dependent notion of "fitness for use" [20], with consolidated metric sets and frameworks that add reliability flags and interpolation over the raw stream across the IoT, wearables, environmental sensing and manufacturing [21–25]. This offers a useful vocabulary, yet stops short of a characterization that is resolved in time, conditioned on the value range, and tied to the specific quantity being computed. Where data are missing or noisy, the classical remedies such as interpolation, Kalman or Gaussian-process smoothing, multiple imputation, and their fuzzy-rough variants [26] remain in active use across monitoring domains, from multiple interpolation of long environmental records [27] to generative reconstruction of multivariate sensor streams [28], and have lately been refined to return a variance alongside each reconstructed value, making explicit how much was inferred [29–31]. Closest to our own concern, Hosseini and Moeini [32] derive a reliability index and use it to complete gaps in a sensor record; the index, however, is consumed by the reconstruction and does not survive into the result. These uncertainty-aware reconstructions are natural building blocks for what we propose, but on their own, they neither localize reliability by value band nor communicate it, which is found to be fundamental to better understand the evolution of a monitored phenomena.

The means to communicate it come from a further line of work on fuzzy linguistic summarization, in which data is described through protoforms (quantified natural-language statements carrying a degree of truth [33]), subsequently refined with measures of appropriateness and focus [34,35], temporal variants [36] and composite forms [37]. Closest to our own machinery, fuzzy temporal windows and protoforms have been used to summarize health and activity streams in fog and edge settings [38–41]; in all of it, the subject of the summary is the monitored phenomenon: a trend, a recognized activity or a heart-rate stream. However, we keep the same procedure and turn it on a different subject, the veracity of the data and of the aggregates directly derived from the phenomena. The formal grounding for attaching such a judgment to an aggregate is not provided in temporal or linguistic monitoring, or oriented to sensor monitoring (see Figure 1).

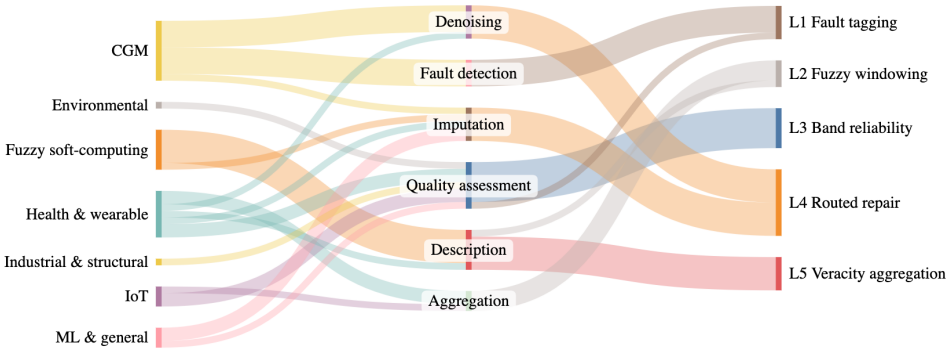


Figure 1. Map of the reviewed literature structuring the domain landscape into sensing application domains, core research objectives, and the layers of the proposed reliability pipeline. Note: All diagrams are original contributions created by the authors specifically for this manuscript.

The objective of this work is to characterize the reliability of a sensor stream jointly over the time and value range, and to carry that characterization through to the aggregates derived from the stream, so that each of them is delivered with an explicit degree of veracity. Its novelty lies in combining three elements that, to our knowledge, no existing approach brings together: reliability is computed on the raw signal as a fuzzy quantity conditioned on both time and value band; it routes fault-specific repair whose uncertainty is propagated rather than discarded; and it is reported for every aggregate as a veracity degree and a linguistic protoform. Methodologically, we formulate a five-layer fuzzy pipeline parameterized by a sensor's range, maximum plausible rate of change and sampling cadence, and instantiate it for CGM with clinical knowledge of glucose dynamics and of sensor failure modes. Fuzzy logic is the natural formalism, because whether a sample is faulty, an interval adequately covered or a value band trustworthy are matters of degree, and forcing them into binary verdicts is what makes conventional quality criteria discard good data and accept bad. For CGM, the expected relevance is clinical: TBR, TIR and TAR drive therapeutic decisions, and knowing which of them can be trusted, and for which periods, is a prerequisite for interpreting them.

To systematically position this work within the current state of the art, we mapped the 30 reviewed studies that inform the proposed pipeline (Table 1). Each study was assigned by the authors to a single application domain (CGM, health & wearable, IoT, environmental, industrial & structural, fuzzy soft-computing, machine learning (ML) & general), to its primary research objective (*Denoising*, Fault detection, hlImputation, Description, Aggregation or Quality assessment) and to the layer of the proposed pipeline it informs (L1 Fault tagging, L2 Fuzzy windowing, L3 Band reliability, L4 Routed repair or L5 Veracity aggregation). Figure 1 visualizes this mapping as a Sankey diagram in which each study flows from its domain to its objective and then to its layer; the width of each flow

is proportional to the number of studies on that path, and every flow can be traced to its citations through Table 1. While previous biomedical and industrial monitoring approaches have focused predominantly on sample-level denoising and imputation (feeding L1 and L4), or on abstract data quality frameworks (L3), none of the reviewed studies provides a unified, fuzzy time- and value-band conditioned pipeline that delivers aggregates annotated with linguistic veracity protoforms (L5).

Table 1. Traceable mapping of the 30 reviewed studies. Each row groups the studies that share a research objective and a pipeline layer (the right-hand flows of Figure 1).

Pipeline Layer	Objective	Domains (Studies)	References
L1 Fault tagging	Fault detection	CGM (4)	[14–17]
	Quality assessment	Health & wearable (1)	[19]
L2 Fuzzy windowing	Aggregation	Health & wearable (2), IoT (1)	[39–41]
	Description	Health & wearable (1)	[38]
L3 Band reliability	Quality assessment	Environmental (1), Health & wearable (1), Industrial & structural (1), IoT (2), ML & general (1)	[20–25]
L4 Routed repair	Denoising	CGM (4), Health & wearable (1)	[2,10–12,18]
	Imputation	CGM (1), Fuzzy soft-computing (1), Health & wearable (1), ML & general (2)	[13,26,29–31]
L5 Veracity aggregation	Description	Fuzzy soft-computing (5)	[33–37]

Accordingly, this study aims to develop and evaluate a domain-agnostic framework that models sensor reliability as a fuzzy field over time and value band and propagates it from raw signal data to reliability-aware aggregates, veracity degrees, and linguistic protoforms. The framework is implemented through a five-layer processing pipeline and evaluated on a 14-month real-world CGM recording.

2. Materials and Methods

2.1. Study Participant and Dataset

The pipeline is evaluated on a long real-world recording [42] from a single adult participant with type 1 diabetes mellitus, monitored under free-living ambulatory conditions. Demographic and clinical characteristics of the participant (such as age, sex, diabetes duration and treatment) are withheld to protect the participant's privacy, as required by the ethics committee that approved the study (protocol code DIC.21/10.PRY, 2022).

Interstitial glucose was measured with factory-calibrated FreeStyle Libre 3 sensors and retrieved through the architecture described in Section 2.2. For this study, the architecture was configured with a single monitored patient, who authorized a single follower account held by the acquisition tier; the acquisition tier polled the manufacturer's cloud every 60 s and stored the series as described in Section 2.2.4. The series retrieved for this study contains one reading every five minutes, so it has a nominal cadence of $\Delta = 5$ min; the median interval between consecutive stored readings is 6.0 min and 75% of the intervals lie between 4.5 and 6.5 min. Readings are reported in mg/dL; the analyzed data contain 1159 readings at the assay floor (39 mg/dL), 376 below it and 9 above the nominal ceiling of 400 mg/dL, all of which are retained and handled by the saturation primitive (Section 2.4).

The recording spans approximately 14 months (435 calendar days), of which 404 contain at least one reading. Individual days are referred to by their position in the recording (Day 1 being the first recorded day), rather than by calendar date. A day is eligible when it contains at least 120 readings, i.e., 41.7% of the $N_\Delta = 288$ readings expected at $\Delta = 5$ min. The resulting 294 eligible days (58,898 readings) form the dataset analyzed in this work. The threshold is deliberately lower than the 70% active-time criterion of the in-

ternational consensus [1]. That criterion is a crisp gate that decides whether a 14-day report is valid, whereas the purpose here is to grade incomplete days rather than discard them: coverage enters the analysis continuously through the global veracity of Equation (15), so that days with between 42% and 70% coverage are retained but penalized (159 of the 294 eligible days would also satisfy a 70% criterion). The 120-reading floor only ensures that every day contains enough readings for the fuzzy windows and band memberships to be defined.

2.2. Data Acquisition Architecture

This section specifies the architecture through which the glucose stream analyzed in this work is acquired, and delimits the function, the protocol and the configuration parameters of each of its components. It is organized in five tiers (Figure 2):

1. A wearable sensor;
2. A gateway hosted on the patient's smartphone;
3. A cloud service operated by the device manufacturer;
4. An acquisition tier under institutional control;
5. The analysis tier.

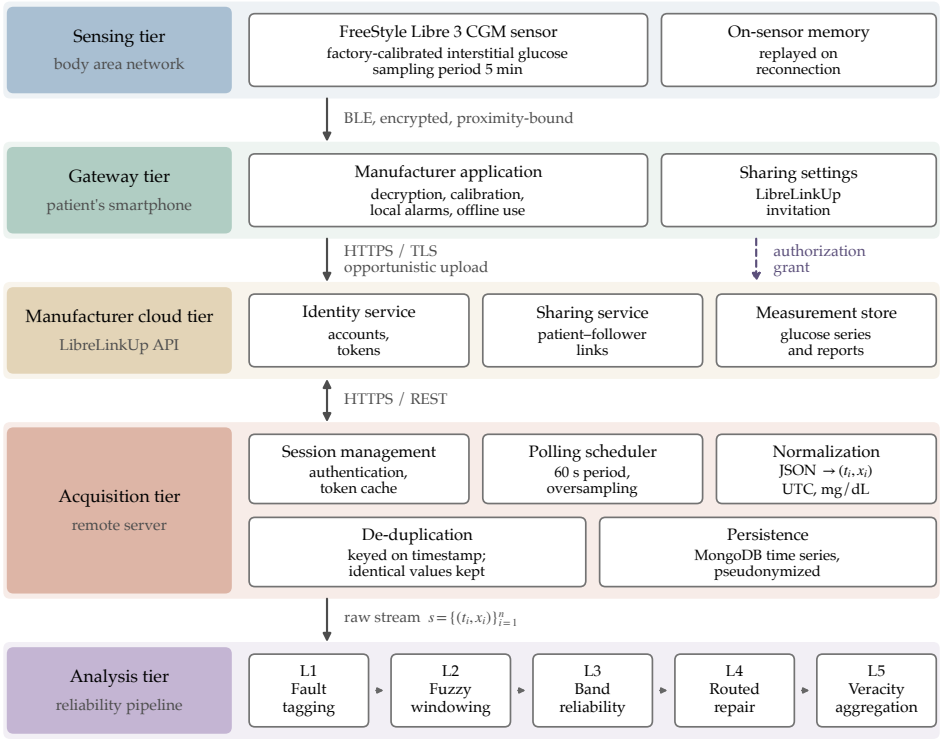


Figure 2. Tiered acquisition architecture for CGM. Solid arrows denote data flow; the dashed arrow is the patient's authorization grant. The Analysis tier hosts the five layers of the reliability pipeline (L1 fault tagging, L2 fuzzy windowing, L3 band reliability, L4 routed repair and L5 veracity aggregation; Section 2.3). BLE: Bluetooth Low Energy; HTTPS: Hypertext Transfer Protocol Secure; TLS: Transport Layer Security; REST: Representational State Transfer; JSON: JavaScript Object Notation; UTC: Coordinated Universal Time.

2.2.1. Sensing Tier

The sensing element is a factory-calibrated FreeStyle Libre 3 sensor (Abbott FreeStyle Libre 3: https://www.freestyle.abbott/us-en/products/freestyle-libre-3.html (accessed on 1 July 2026)), which was the device used throughout the recording analyzed in this study (Section 2.1). Like the other members of the FreeStyle Libre family, it measures interstitial glucose through a subcutaneous filament and requires no capillary calibration by

the user. The characteristics described below are those of the FreeStyle Libre 3: the sensor has a nominal wear time of 14 days, transmits each reading over an encrypted Bluetooth Low Energy (BLE) link, and the series retrieved for this study contains one reading every five minutes ($\Delta = 5$ min; Section 2.1). The link operates over a range of a few meters and requires no network connectivity, so that acquisition continues while the gateway is unreachable. Samples produced during such an interval are retained in the internal memory of the sensor and retransmitted upon reconnection. Acquisition time and arrival time at the gateway are consequently decoupled by an interval bounded by the capacity of that memory.

2.2.2. Gateway Tier

The gateway is the manufacturer's application executing on the patient's smartphone. It decrypts the sensor payload, applies the calibration model that converts the raw signal into a glucose concentration, evaluates the alarm thresholds and presents the reading to the patient; none of these functions depends on network availability. The gateway additionally uploads the calibrated series to the manufacturer's cloud over Hypertext Transfer Protocol Secure (HTTPS). Upload is opportunistic and asynchronous with respect to acquisition: an interval without connectivity postpones transmission rather than discarding samples, so the cloud record of a given interval may be completed after that interval has elapsed. The sharing settings by which the patient authorizes a third party to read the series are configured from this tier.

2.2.3. Manufacturer Cloud Tier

The third tier is operated by the device manufacturer and is reached through the interface that serves its companion application for followers. It provides an identity service that maintains accounts and issues bearer tokens, a sharing service that maintains patient–follower associations, and a measurement store that persists the glucose series and derives the standard clinical reports. Access is region-specific: an authentication request issued against the default endpoint returns, for an account held elsewhere, the region code identifying the endpoint against which the request must be reissued. This tier holds the only authoritative copy of the series; neither the patient nor the acquisition tier retains an independent original, so no reference is available against which a repaired value could be verified.

The authorization model is disjointed from the data plane. The patient issues a sharing invitation to the account under which the acquisition tier authenticates, and the association becomes effective only once that invitation has been accepted, at which point the patient is enumerated among the connections of the account. The grant is revocable, and revocation is not reported as an error condition: the patient ceases to be enumerated.

2.2.4. Acquisition Tier

The acquisition tier is the only component of the architecture under institutional control. It comprises a session manager, a polling scheduler, a normalizer, a de-duplication stage and a persistence store.

Retrieval follows a fixed sequence at every cycle and is mediated by a third-party client library for the manufacturer's sharing service [43]. The session manager authenticates with the credentials of the authorized account, obtains a bearer token and, when the account is not held in the default region, resolves the regional endpoint against which the request must be reissued. The patients who have authorized that account are then enumerated, which yields the internal identifier of the monitored subject together with the metadata of the active sensor: serial number, activation time, configured target range and alarm thresholds. The reading is finally requested for that identifier. Every request subsequent

to authentication carries the bearer token and an account header derived from the user identifier. The interface is the one serving the manufacturer's companion application rather than a formally specified public application programming interface (API), so its contract is not guaranteed to be stable; the tier is implemented such that a change of contract degrades into an interruption of retrieval rather than into malformed records. Failed requests are not logged as separate events: retrieval is simply reattempted at the next polling cycle, and every record is written by its timestamp, overwriting any previous version of the same reading. Acquisition interruptions therefore appear in the stored series as missing timestamps, indistinguishable from sensor-side gaps, and are handled downstream by the gap primitive and the coverage measures of the pipeline (Sections 2.3.1 and 2.3.2).

The scheduler polls at a period of 60 s, below the refresh period of the source. The resulting oversampling bounds the latency with which a new reading is observed, which is necessary because the interface exposes no notification channel, and has the consequence that most responses repeat the preceding value. Repetition in the retrieved responses is therefore attributable to the polling policy before it is attributable to the sensor. The de-duplication stage uses the sample timestamp reported by the cloud rather than the value: a reading already stored is discarded when a subsequent poll returns it again, whereas two distinct readings that happen to carry the same value are both retained. Runs of identical values are thus preserved for the stagnation primitive of Section 2.3.1, which would otherwise be unable to fire on any stored series. The normalizer maps each accepted response to a record comprising a Coordinated Universal Time (UTC) timestamp, a value in mg/dL, the trend indicator and the sensor serial number, and the persistence store holds these records as a time series in which the patient identifier is replaced by a study code.

2.2.5. Analysis Tier

The analysis tier receives the stored series as the raw stream s and is the only tier that issues a judgment about the data rather than transporting or persisting it. It hosts the reliability pipeline of Section 2.3, organized in five layers (L1–L5).

2.3. A Reliability Pipeline for Monitoring

The pipeline turns a raw sensor stream into aggregates that each carry a degree of veracity, together with a linguistic statement of how far they can be trusted. It is organized in five layers, L1–L5 (Figure 3), and is parameterized, for a given sensor, by three quantities: the admissible value range $[x_{\min}, x_{\max}]$, the maximum physiologically plausible rate of change ρ, and the nominal sampling period Δ. Throughout, we write a stream as $s = \{(t_i, x_i)\}_{i=1}^{n}$. A guiding principle keeps apart two concerns that are usually conflated: faults are detected on the raw signal (L1), whereas any repair is a distinct branch (L4) whose every output carries the uncertainty it introduces, so that cleaning can never silently erase the evidence of its own necessity. Every quantity that the layers compute (sample quality, effective coverage, band reliability and veracity) is expressed in $[0, 1]$ rather than as a verdict.

2.3.1. L1: Fault Tagging

Fault detection is performed per sample on the raw stream. Each sample is scored by five parameterized primitives, one per failure mode recurrent across monitoring domains, all built from the piecewise-linear ramps

$$R^{\uparrow}_{a,b}(z) = \min\left(1, \max\left(0, \frac{z-a}{b-a}\right)\right), \qquad R^{\downarrow}_{a,b}(z) = 1 - R^{\uparrow}_{a,b}(z), \tag{1}$$

where a is the point at which doubt begins and b is the point of full membership. For sample i, the fault degrees are

$$\mu_i^{\text{sat}} = \max\left(R^{\downarrow}_{x_{\min},\, x_{\min}+\delta_{\text{lo}}}(x_i),\ R^{\uparrow}_{x_{\max}-\delta_{\text{hi}},\, x_{\max}}(x_i)\right), \tag{2}$$

$$\mu_i^{\text{rate}} = R^{\uparrow}_{\rho,\, 2\rho}(\dot{x}_i), \qquad \dot{x}_i = \min\left(\frac{|x_i - x_{i-1}|}{t_i - t_{i-1}},\ \frac{|x_{i+1} - x_i|}{t_{i+1} - t_i}\right), \tag{3}$$

$$\mu_i^{\text{stag}} = R^{\uparrow}_{\ell_a,\, \ell_b}(\ell_i), \tag{4}$$

$$\mu_i^{\text{gap}} = R^{\uparrow}_{1.5\Delta,\, 6\Delta}(g_i), \qquad g_i = \min\left(t_i - t_{i-1},\ t_{i+1} - t_i\right), \tag{5}$$

$$\mu_i^{\text{osc}} = R^{\uparrow}_{s_a,\, s_b}(s_i), \qquad s_i = \text{sd}\left\{\nabla^2 x_j : |j - i| \le 2\right\}, \quad \nabla^2 x_j = x_j - 2x_{j-1} + x_{j-2}, \tag{6}$$

where ℓ_i is the length, in samples, of the maximal run of identical consecutive values that contains sample i, and boundary samples use a virtual neighbor at distance Δ with the same value. Saturation (Equation (2)) measures proximity to the physical bounds; implausible rate (Equation (3)) compares the rate of change with the slew limit ρ, taking the smaller of the rates to the two neighbors so that an isolated spike scores high while a monotone trend or a single step does not; stagnation (Equation (4)) grows with the length of a run of identical values; missingness (Equation (5)) compares the distance to the nearest neighbor with the expected cadence; and oscillation (Equation (6)) responds to local roughness, measured as the dispersion of second differences. The fault degrees are combined into a per-sample quality through the complement of their probabilistic sum,

$$q_i = \prod_{f \in \mathcal{F}} \left(1 - \mu_i^f\right), \qquad \mathcal{F} = \{\text{sat}, \text{rate}, \text{stag}, \text{gap}, \text{osc}\}, \tag{7}$$

so that several mild faults accumulate rather than being masked by the largest one, and into a dominant fault label,

$$f_i = \arg\max_{f \in \mathcal{F}}\ \mu_i^f, \tag{8}$$

which routes repair in L4. The saturation, rate and gap primitives are defined only through $[x_{\min}, x_{\max}]$, ρ and Δ; stagnation and oscillation additionally require two shape parameters each (ℓ_a, ℓ_b and s_a, s_b), which must be set for every sensor (Section 2.4).

2.3.2. L2: Fuzzy Windowing

L2 places overlapping windows W_j, centered at instants c_j spaced h_c apart, whose membership decays smoothly from the center,

$$\mu_{W_j}(t) = \exp\left(-\frac{(t - c_j)^2}{2s_W^2}\right), \tag{9}$$

with width s_W. What a window carries is not the number of samples that fall in it but its quality-weighted effective coverage,

$$c_W = \frac{\sum_i \mu_W(t_i)\, q_i}{\sum_{t \in G_W} \mu_W(t)}, \tag{10}$$

where G_W is the grid of expected sampling instants under Δ over the analyzed period (for one day, the $N_\Delta = 1440/\Delta$ instants $0, \Delta, 2\Delta, \ldots$); the numerator runs over the samples that are actually present. A gap therefore lowers c_W continuously instead of producing a discontinuity between neighboring bins, and a sample near a window edge contributes partially to both windows.



Figure 3. The five-layer reliability pipeline, shown for a generic sensor (left) and instantiated in glucose (right). The left column lists illustrative estimators admitted by each layer; the right column shows the procedures that were actually implemented and evaluated in this study (Section 2.4). The output values are illustrative.

2.3.3. L3: Band Reliability

Reliability is resolved not only in time but along the value axis. The range is covered by linguistic bands $\{B_k\}$ with memberships $\mu_{B_k}(x)$. For a window W and band B_k,

$$r_{W,k} = \frac{\sum_i \mu_W(t_i)\, \mu_{B_k}(x_i)\, q_i}{\sum_i \mu_W(t_i)\, \mu_{B_k}(x_i)} \tag{11}$$

weights sample quality by how strongly each reading belongs to the band. A window may thus be reliable in one band and untrustworthy in another: the situation that a single global figure cannot express. When the denominator vanishes (below 10^{-6}, i.e., no reading of the window belongs to the band), $r_{W,k}$ is undefined and reported as missing.

2.3.4. L4: Routed Repair

Repair is applied selectively, according to the fault label f_i, and only when the dominant fault degree reaches a routing threshold θ_r; samples below it pass through unchanged. Every routed estimator must return a repaired value $\hat{x}_i$ together with a reconstruction standard deviation σ_i. The routing admits any estimator with this property: for instance, a Hampel or median filter for spikes and oscillations or a Kalman or Gaussian-process smoother for gaps. The estimators actually implemented and evaluated in this study are specified in Section 2.4: a running median for rate and oscillation faults, and flagging without reconstruction for saturation and stagnation, since a sensor pinned at its bound carries no recoverable information; missing stretches are *not* imputed. The reconstruction

variance σ_i^2 is propagated forward rather than discarded (L5). This branch never feeds back into L1: the fault scores remain those of the raw signal.

2.3.5. L5: Veracity Aggregation

Reconstruction uncertainty enters every aggregate through a reliability weight that attenuates the quality of each sample by the uncertainty of its repaired value,

$$\kappa_i = \min\left(1,\ \frac{\sigma_{\text{ref}}^2 + \sigma_0^2}{\sigma_{\text{ref}}^2 + \sigma_i^2}\right), \qquad \omega_i = q_i\,\kappa_i, \tag{12}$$

where σ_0 is the standard deviation assigned to an unrepaired reading, so that $\kappa_i = 1$ when no repair took place, and σ_{ref} is the scale, in the units of x, at which reconstruction uncertainty becomes comparable to the resolution of the value bands. For a band B_k and a period containing n readings, the *naïve* time in band (TIB) takes every present reading at face value, whereas the *reliability-aware* TIB is computed over the repaired stream and weighted by ω_i:

$$\text{TIB}_k^{\text{naïve}} = \frac{1}{n}\sum_i \mu_{B_k}(x_i), \qquad \text{TIB}_k^{\text{aware}} = \frac{\sum_i \mu_{B_k}(\hat{x}_i)\,\omega_i}{\sum_i \omega_i}. \tag{13}$$

Each aware estimate is delivered with a band veracity and the profile as a whole with a global veracity,

$$v_k = \frac{\sum_i \mu_{B_k}(\hat{x}_i)\,\omega_i}{\sum_i \mu_{B_k}(\hat{x}_i)}, \tag{14}$$

$$V = C \cdot \frac{1}{n}\sum_i \omega_i, \qquad C = \min\left(1, \frac{n}{N}\right), \tag{15}$$

where N is the number of readings expected in the period under Δ, so that V decreases both with low sample reliability and with missing data. The same quantities feed a linguistic summary through a quantifier-driven aggregation in the sense of Kacprzyk et al. [33]. Writing

$$p_k = \frac{\sum_i \mu_{B_k}(\hat{x}_i)\,(1-\omega_i)}{\sum_i \mu_{B_k}(\hat{x}_i)} \tag{16}$$

for the proportion of unreliable readings in band B_k, the system emits the protoform "during the $\{P\}$, Q^* of the $\{B_k\}$ readings are unreliable" whenever $p_k > \theta_p$ and the band contains a minimum mass $\sum_i \mu_{B_k}(\hat{x}_i) \geq m_{\min}$. Given a set of fuzzy quantifiers $\{Q_j\}$ defined on the proportion p_k, the quantifier is the one with the highest truth degree, which is also the truth degree of the following statement:

$$Q^* = \arg\max_j\ \mu_{Q_j}(p_k), \qquad \mathcal{T} = \mu_{Q^*}(p_k). \tag{17}$$

The period P is the fuzzy period of the day with the highest membership at the unreliability-weighted mean time $\bar{t}_k = \sum_i t_i\,\mu_{B_k}(\hat{x}_i)(1-\omega_i) / \sum_i \mu_{B_k}(\hat{x}_i)(1-\omega_i)$, and composite protoforms [37] are formed when several such statements must be combined. The output is therefore not a bare number, but an aggregate, its veracity, and a sentence that a practitioner can read.

2.4. Instantiation in CGM

The general pipeline is instantiated for CGM by fixing the three sensor parameters and the shape of each fuzzy primitive. We use a nominal cadence $\Delta = 5$ min, a value range $[x_{\min}, x_{\max}] = [39, 400]$ mg/dL (the reported floor and ceiling of the assay) and a

physiological rate limit derived below. Table 2 lists every parameter of the instantiation together with its basis.

Table 2. Parameters of the CGM instantiation. Ramps (Equation (1)) are given from "doubt begins" (a) to "full membership" (b). Basis: sensor, fixed by the device specification; physiological, fixed by clinical knowledge of glucose dynamics; clinical, centered on consensus thresholds; expert rule, set a priori from expert rules on CGM failure signatures (Section 2.4); design, set by the authors.

Layer	Parameter	Value	Basis
L1	Cadence Δ	5 min	sensor
L1	Range $[x_{\min}, x_{\max}]$	$[39, 400]$ mg/dL	sensor
L1	Saturation, floor/ceiling	$48 \to 39/380 \to 400$ mg/dL	sensor
L1	Rate $\rho \to 2\rho$	$2 \to 4$ mg/dL·min^{-1}	physiological
L1	Stagnation $\ell_a \to \ell_b$	$4 \to 10$ readings	expert rule
L1	Gap $1.5\Delta \to 6\Delta$	$7.5 \to 30$ min	sensor (cadence)
L1	Oscillation $s_a \to s_b$	$8 \to 25$ mg/dL	expert rule
L2	Window spacing h_c/width s_W	30/25 min	design
L3	Hypoglycemic band	$75 \to 65$ mg/dL	clinical
L3	Hyperglycemic band	$170 \to 190$ mg/dL	clinical
L4	Routing threshold θ_r	0.3	design
L4	Running-median window	5 readings	design
L4	Reconstruction s.d. rules	Equation (18)	design
L5	$\sigma_0/\sigma_{\text{ref}}$	2/10 mg/dL	design
L5	Expected readings per day N	288	sensor (cadence)
L5	Protoform threshold θ_p/mass $m_{\min}$	0.35/3	design
L5	Periods of the day	Equation (21)	design
L5	Quantifiers	Equation (22)	design

Rate of change (physiological)

The best-justified primitive is the implausible-rate detector. Interstitial glucose cannot change arbitrarily fast: two consecutive readings differ by at most about 10 mg/dL, i.e., a rate of roughly 2 mg/dL·min^{-1} at the 5-min cadence. We therefore let μ^{rate} rise from 0 at $\rho = 2$ to 1 at $2\rho = 4$ mg/dL·min^{-1}; expressing the criterion as a rate rather than a fixed per-sample step keeps the bound correct across gaps. A fall from 240 to 100 mg/dL within a single 5-min interval (28 mg/dL·min^{-1}), for instance, lies far beyond full membership.

Saturation, stagnation, gap and oscillation

Saturation membership is 1 at or below the assay floor (39 mg/dL) and decays to 0 by 48 mg/dL ($\delta_{\text{lo}} = 9$ mg/dL), and rises from 380 to 400 mg/dL at the ceiling ($\delta_{\text{hi}} = 20$ mg/dL), so that a reading pinned at, or beyond, a bound is treated as an artifact rather than a true extreme. Stagnation starts at runs of $\ell_a = 4$ identical readings (20 min) and is full at $\ell_b = 10$ (50 min): because readings are reported as integers, short repeats are normal during stable periods, and only long runs are suspicious; the de-duplication of the acquisition tier preserves such runs (Section 2.2.4). The gap primitive ramps from $1.5\Delta = 7.5$ to $6\Delta = 30$ min, so a single missed reading is not penalized. Because g_i uses the nearest neighbor, it marks readings that are isolated on both sides, as in sparse, intermittent stretches; the absence of readings itself is captured by the effective coverage of Equation (10) and by the coverage factor of Equation (15), not by any sample score. Oscillation ramps from $s_a = 8$ to $s_b = 25$ mg/dL of standard deviation of the second differences within five consecutive readings (about 25 min). Unlike the rate limit, these thresholds are not fixed by physiology: they encode expert rules on the failure signatures of CGM (how long a reported value may remain identical, how rough a physiological trace may be), and were set a priori from those rules rather than tuned on the days used for evaluation (Section 2.5).

Value bands

For L3, the glucose axis is covered by three linguistic bands with soft boundaries: hypoglycemic ($\mu = 1$ below 65, 0 above 75 mg/dL), in-range, and hyperglycemic ($\mu = 0$ below 170, 1 above 190 mg/dL), bracketing the clinical 70–180 mg/dL target range. The in-range membership is $\mu_{\text{in}} = \max(0, 1 - \mu_{\text{hypo}} - \mu_{\text{hyper}})$, so the three memberships sum to one (Equation (20)). The boundaries are centered on the consensus thresholds of 70 and 180 mg/dL [1]; their half-widths of 5 and 10 mg/dL (about 7% and 6% of the threshold) are commensurate with the accuracy of factory-calibrated sensors, so that a reading within measurement error of a clinical threshold belongs partially to both adjacent bands.

Fuzzy windows

L2 uses windows centered every $h_c = 30$ min from 00:00 to 24:00 (49 windows per day), with width $s_W = 25$ min. Adjacent windows overlap substantially: each center has membership of 0.49 in its neighbors' windows.

Routed repair (L4)

Samples whose dominant degree is below $\theta_r = 0.3$ pass through unchanged. Otherwise, with m_i, the running median of five consecutive readings centered on i, and $\text{MAD}_i = 1.4826 \cdot \text{median}_{|j-i|\le 2} |x_j - m_j|$, the corresponding median absolute deviation (MAD), the repaired value and its standard deviation (in mg/dL) are

$$(\hat{x}_i,\ \sigma_i) = \begin{cases} (x_i,\ \sigma_0) & \text{if } \max_f \mu_i^f < \theta_r, \\ \left(m_i,\ \mu_i^{f_i} \max(\text{MAD}_i, |x_i - m_i|) + 3\right) & \text{if } f_i \in \{\text{rate}, \text{osc}\}, \\ \left(m_i,\ 50\,\mu_i^{f_i} + 5\right),\ \text{flagged} & \text{if } f_i \in \{\text{sat}, \text{stag}\}, \\ \left(x_i,\ 20\,\mu_i^{f_i} + 3\right) & \text{if } f_i = \text{gap}, \end{cases} \tag{18}$$

with $\sigma_0 = 2$ mg/dL. Readings dominated by the implausible-rate or oscillation memberships are thus replaced by a local robust estimate whose uncertainty grows with both the local scatter and the fault degree. Readings pinned at the assay floor or stuck at a constant value are *flagged and left unreconstructed*: the median is a mere placeholder and its standard deviation of up to 55 mg/dL spans more than one band, so that such readings carry essentially no band information ($\kappa_i \approx 0.03$ in Equation (12)). Readings isolated by gaps keep their value with an inflated uncertainty. No value is generated for missing timestamps: missing stretches are not imputed or interpolated, and enter the analysis only through c_W and the coverage factor C.

Veracity aggregation and protoforms (L5)

The clinical profile is the triple TBR, TIR and TAR: that is, the TIB of each glucose band, computed with Equation (13) over each day. We set $\sigma_{\text{ref}} = 10$ mg/dL, the width of the transition of the hypoglycemic band, so that a reconstruction uncertainty comparable to that width roughly halves the weight of a reading ($\kappa_i = 0.52$ at $\sigma_i = 10$ mg/dL), and $N = 288$ readings per day in Equation (15). Protoforms are emitted for $p_k > \theta_p = 0.35$ with $m_{\min} = 3$, using the quantifiers and periods of the day of Equations (21) and (22); for instance, the system states *"during the afternoon, almost all of the low readings are unreliable"*, with truth degree $\mathcal{T} = \mu_{\text{almost all}}(p_{\text{low}})$.

Linguistic terms

All linguistic terms of L3 and L5 are trapezoidal fuzzy sets,

$$\text{Tr}_{a,b,c,d}(z) = \max\left(0,\ \min\left(\frac{z-a}{b-a},\ 1,\ \frac{d-z}{d-c}\right)\right), \tag{19}$$

where a side with $a = b$ (or $c = d$) is taken as constant and equal to one, so that ramps are obtained as particular cases. The glucose bands (in mg/dL), the periods of the day (in hours) and the quantifiers (on the proportion p_k) are

$$\mu_{\text{hypo}} = \text{Tr}_{0,0,65,75}, \quad \mu_{\text{hyper}} = \text{Tr}_{170,190,x_{\max},x_{\max}}, \quad \mu_{\text{in}} = \max\Big(0, 1 - \mu_{\text{hypo}} - \mu_{\text{hyper}}\Big), \tag{20}$$

$$\mu_{\text{night}} = \text{Tr}_{0,0,5.5,6.5}, \quad \mu_{\text{morning}} = \text{Tr}_{5.5,6.5,11.5,12.5},$$

$$\mu_{\text{afternoon}} = \text{Tr}_{11.5,12.5,17.5,18.5}, \quad \mu_{\text{evening}} = \text{Tr}_{17.5,18.5,24,24}, \tag{21}$$

$$\mu_{\text{about half}} = \text{Tr}_{0.30,0.40,0.50,0.65}, \quad \mu_{\text{most}} = \text{Tr}_{0.50,0.65,0.80,0.90}, \quad \mu_{\text{almost all}} = \text{Tr}_{0.80,0.90,1,1}, \tag{22}$$

The three families are fuzzy partitions: the memberships of adjacent terms sum to one in every transition. For the periods, the transitions last one hour and are centered at 06:00, 12:00 and 18:00. For the quantifiers, the partition guarantees that the selected quantifier of Equation (17) always has a truth degree of at least 0.5 for $p_k > \theta_p$.

2.5. Evaluation Procedure

Reference days

Ten days of the recording were selected as reference cases on the basis of expert knowledge of CGM failure modes, informed by feedback from the participant and from the participant's physician on the incidents and anomalous readings observed during the recording: two clean control days and eight days exhibiting characteristic faults (missing stretches, anomalous readings in the low or high band, oscillation, and clamping at the assay floor), each labeled with its dominant fault character (Table 3). Two of the artifact days (Days 90 and 184) are dominated by missing data and contain fewer than 120 readings, so they fall outside the eligible set; the remaining six are marked in the phenotyping analysis (Section 2.6). Because these labels and the fault rules of L1 derive from the same body of clinical knowledge, agreement on the reference days is a consistency check of the implementation rather than an independent validation of the detector. Nevertheless, the participant and the physician provided their feedback without access to the pipeline outputs: the aim of the method is to model their knowledge of when the data can be trusted, not to adjust that knowledge to the method, so that no "false" knowledge introduced by imputation or smoothing could leak into the reference.

Naïve versus reliability-aware profiles

On the reference days, the profile obtained by taking every reading at face value (the naïve estimate of Equation (13)) is contrasted with the reliability-aware profile of four quantities: the three times in band and the global veracity V.

Sensitivity to the reconstruction uncertainty

To isolate the contribution of the reconstruction uncertainty propagated from L4, every reliability-aware analysis (reference days and phenotype clustering) is repeated with $\kappa_i = 1$ in Equation (12), i.e., weighting each reading by its quality q_i alone.

2.6. Phenotype Clustering

To assess whether reliability-aware processing changes how a patient is characterized over time, or only on the odd corrupted day, we profiled every eligible day of the recording (Section 2.1) and clustered the days in the time-in-band space, once from the naïve profiles and once from the reliability-aware ones. For each day, the features are the times below and above range (TBR, TAR) of Equation (13), standardized to zero mean and unit variance across days; TIR is omitted because TIR $= 100 -$ TBR $-$ TAR carries no additional information. Each set of profiles is partitioned with k-means (Euclidean distance, 10 random initializations, random seed 0; scikit-learn). The number of clusters was chosen among $k = 2, \ldots, 8$ by two criteria evaluated on both the naïve and the reliability-aware

profiles: the mean silhouette coefficient, and the stability of the partition, measured as the mean adjusted Rand index (ARI) between the full-data partition and 100 partitions fitted on random subsamples of 80% of the days. $k = 3$ maximized both criteria (naïve profiles: silhouette 0.46, ARI 0.97; aware profiles of the profilable days: silhouette 0.49, ARI 0.99, with 5th-percentile ARI above 0.94 in both cases). The two partitions are fitted independently: the reliability-aware clustering is a separately trained model, not a relabeling of the naïve one. Consistently with the meaning of the global veracity, days with $V < 0.5$ (Equation (15)) are considered to be too unreliable to profile and receive no reliability-aware phenotype: the aware partition is fitted only on the remaining days, with k re-selected on them by the same criteria, whereas the naïve partition, which ignores reliability, assigns a phenotype to every eligible day. Each day is therefore classified as keeping its naïve phenotype, changing it, or becoming non-profilable. To compare the two partitions, clusters are matched one-to-one by maximizing their overlap (Hungarian algorithm on the $k \times k$ contingency table), using the profilable days, and a profilable day is said to change phenotype when its matched labels differ. To compare the veracity-based exclusion with the crisp consensus rule, the profilable days are also contrasted with the days that meet the 70% active-time criterion [1].

Table 3. Per-sample quality on the ten reference days, ordered by mean quality. Character: dominant fault character according to the feedback of the participant and the physician (Section 2.5): *clean control*, no known incident; *missing*, long stretches without readings; *anomaly*, readings reported as anomalous (in the indicated band, when known); *oscillation*, physiologically impossible alternation of values; *floor*, readings clamped at the assay floor. n: number of readings of the day. Mean q: mean per-sample quality, $\frac{1}{n}\sum_i q_i$ (Equation (7)). $q < 0.5$: number of readings with quality below 0.5, and percentage of the day's readings. Days 90 and 184 contain fewer than 120 readings and lie outside of the eligible set (Section 2.1).

Day	Character	n	Mean q	$q < 0.5$, n (%)
Day 9	clean control	275	0.99	0 (0.0)
Day 72	clean control	200	0.98	5 (2.5)
Day 52	missing	181	0.87	24 (13.3)
Day 161	anomaly (high band)	240	0.79	38 (15.8)
Day 158	anomaly (low band)	244	0.78	53 (21.7)
Day 352	oscillation + floor	239	0.68	72 (30.1)
Day 192	anomaly	254	0.66	80 (31.5)
Day 90	missing	108	0.54	49 (45.4)
Day 245	oscillation	246	0.54	110 (44.7)
Day 184	anomaly	104	0.51	51 (49.0)

3. Results

3.1. Per-Sample Quality on the Studied Days

With the memberships of Table 2, the clean control days retain a mean quality of 0.98–0.99, with 2.5% of their readings below $q = 0.5$ at most, whereas every artifact day drops to 0.51–0.87, with 13–49% of its readings below 0.5 (Table 3). The fault rules thus mark the days that the participant and the physician identified as problematic while leaving the clean days almost untouched, which confirms that the knowledge encoded in L1 is implemented consistently (Section 2.5).

3.2. Fault Detection on Representative Days

Figures 4–6 run the instantiated pipeline (L1–L3) on three contrasting days. On a clean control day (Day 9; Figure 4), quality stays at $q \approx 0.99$, effective coverage sits near one, and every band reliability is close to unity: the method is silent when there is nothing to flag. On a day dominated by physiologically impossible oscillation (Day 245; Figure 5), with the

sensor alternating between roughly 90 and 200 mg/dL within minutes, the implausible-rate and oscillation memberships fire across the affected interval, the effective coverage c_W collapses towards zero there, and the in-range, hypoglycemic and hyperglycemic assessments are all marked untrustworthy for that stretch, while the calm morning and late evening remain reliable. On a day with long missing stretches (Day 90; Figure 6), the gap membership dominates instead: effective coverage falls wherever the sensor is absent and the band reliabilities are low or undefined there, so the day's aggregates are honestly reported as incomplete rather than silently interpolated.

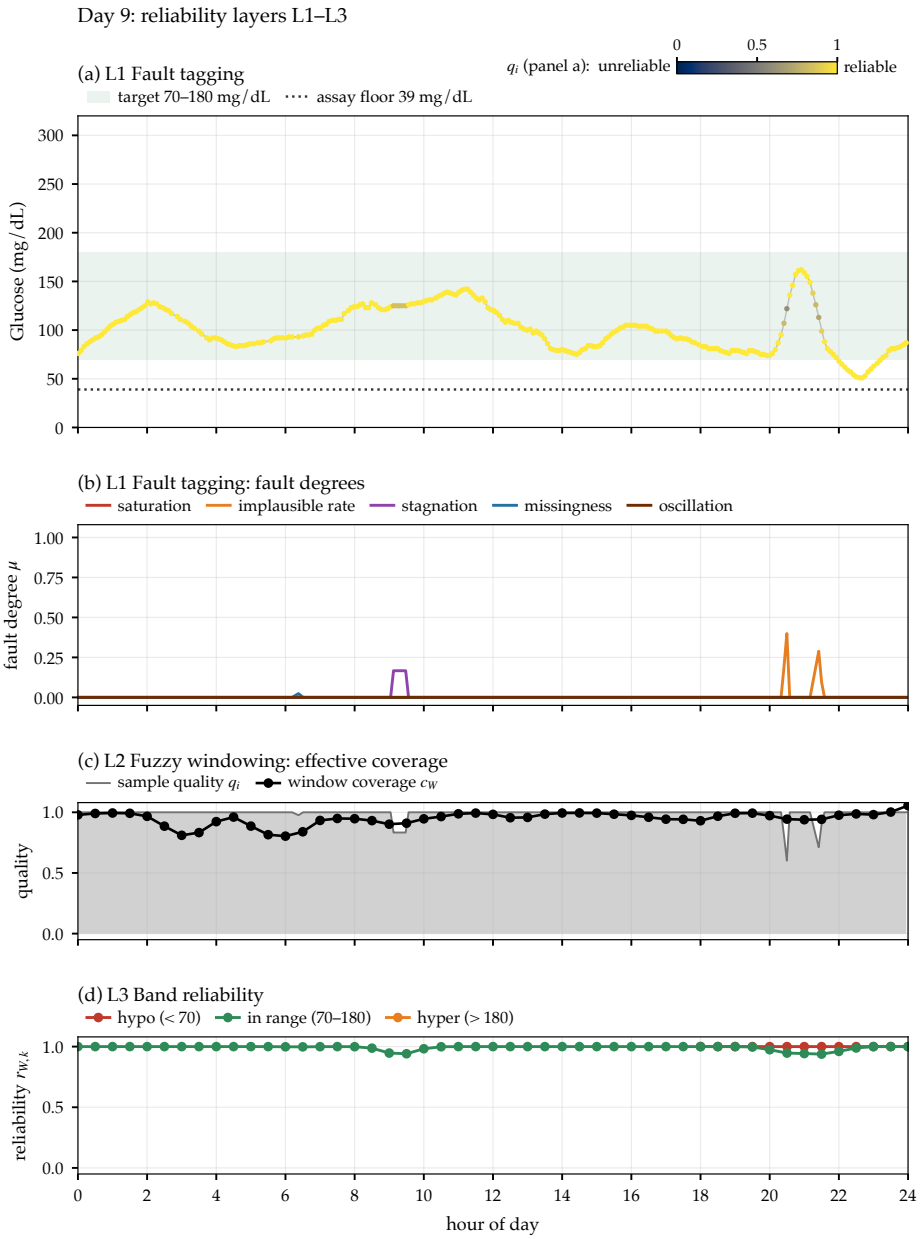


Figure 4. Reliability layers L1–L3 on a *clean control* day (Day 9 of the recording). (**a**) L1 fault tagging: each glucose reading coloured by its quality q_i (the colour bar applies to this panel only: 0 = unreliable, 1 = reliable); the shaded band marks the 70–180 mg/dL target range, the dotted line the assay floor (39 mg/dL), and hatched areas intervals without readings for more than 30 min. (**b**) Fault degree μ_i^f of each of the five primitives (Equations (2)–(6)). (**c**) L2 fuzzy windowing: sample quality q_i (grey area) and quality-weighted effective coverage c_W of each window (black dots; Equation (10)). (**d**) L3 band reliability $r_{W,k}$ of the hypoglycemic, in-range and hyperglycemic bands (Equation (11)); gaps mark windows in which no reading belongs to the band. Quality and coverage stay near one all day.

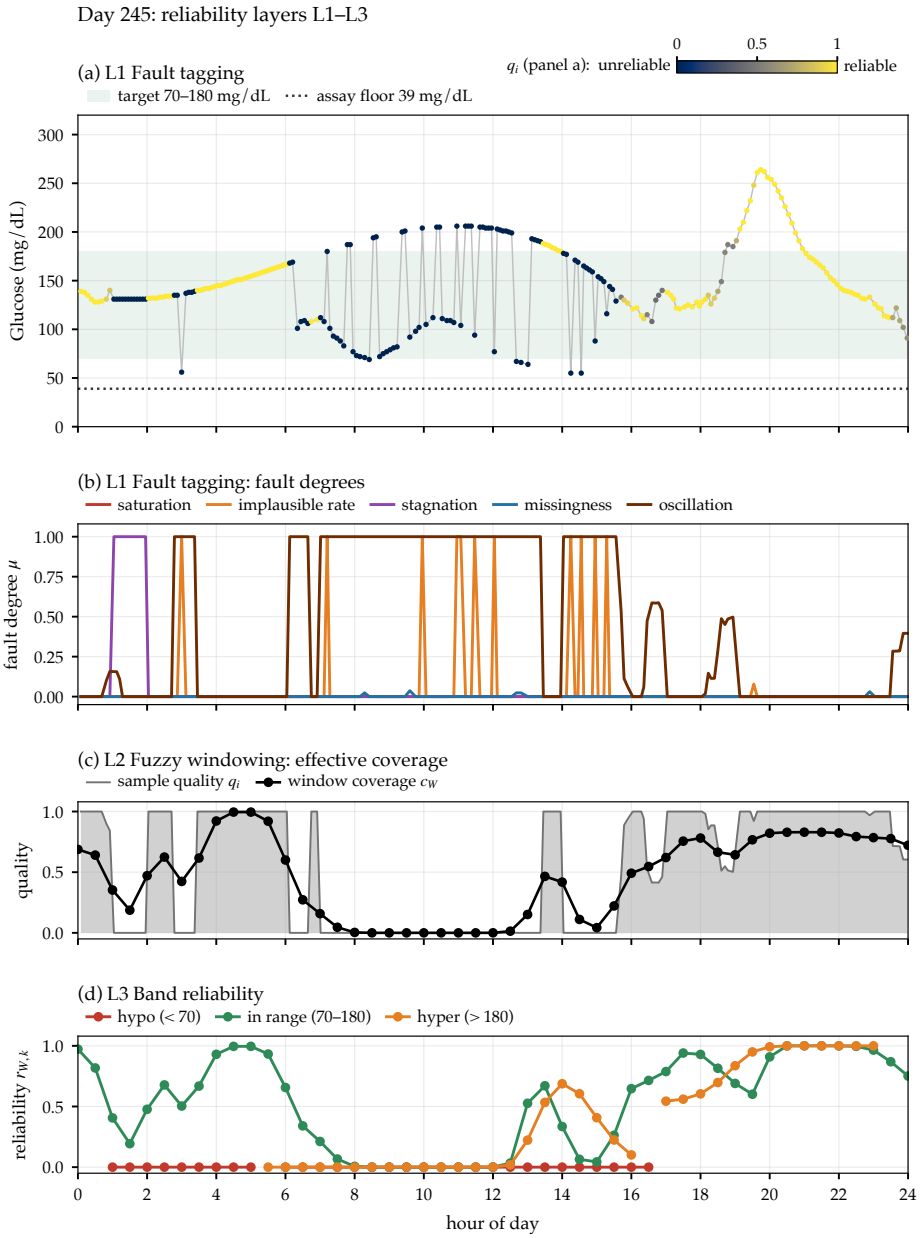


Figure 5. Reliability layers L1–L3 on a day of physiologically impossible *oscillation* (Day 245), with the layout of Figure 4. Panels: (**a**) readings coloured by q_i (colour bar for panel (**a**) only); (**b**) fault degrees; (**c**) q_i (grey area) and window coverage c_W (dots); (**d**) band reliability $r_{W,k}$; hatched areas, no readings for more than 30 min. The oscillation between about 6 and 16 h drives quality, coverage and every band reliability down, whereas the calm intervals stay reliable.

3.3. Repair and Reliability-Aware Profiling of the Most Corrupted Day

Figure 7 illustrates the performance of L4 and L5 on Day 352, selected as the most corrupted day according to the multi-fault compound criterion. Although some days in Table 3 exhibit a lower global mean quality score ($\bar{q}$), mainly due to extended missing gaps or isolated noise spikes, Day 352 uniquely combines persistent assay floor-clamping (saturation at $\leq$39 mg/dL) with high-amplitude physiological oscillations (>100 mg/dL within minutes). This combination produces a substantial distortion in the estimated TBR: the naïve profile reports $\mathrm{TBR}_{\mathrm{naive}} = 16\%$, whereas the reliability-aware profile reports $\mathrm{TBR}_{\mathrm{aware}} = 3\%$, a reduction of about 80%. More importantly, the reliability-aware profile assigns a veracity of only 0.14 to the low-range estimate and generates the protoforms *"during the afternoon, almost all of the low readings are unreliable"* ($\mathcal{T} = 0.63$) and *"during the*

evening, most of the high readings are unreliable" ($\mathcal{T} = 1.00$). In contrast, on a clean control day (Day 9), the two profiles remain essentially unchanged (5% versus 6% below range), with all band veracities close to one. Thus, the method remains effectively silent when the underlying data do not provide evidence of relevant quality problems.

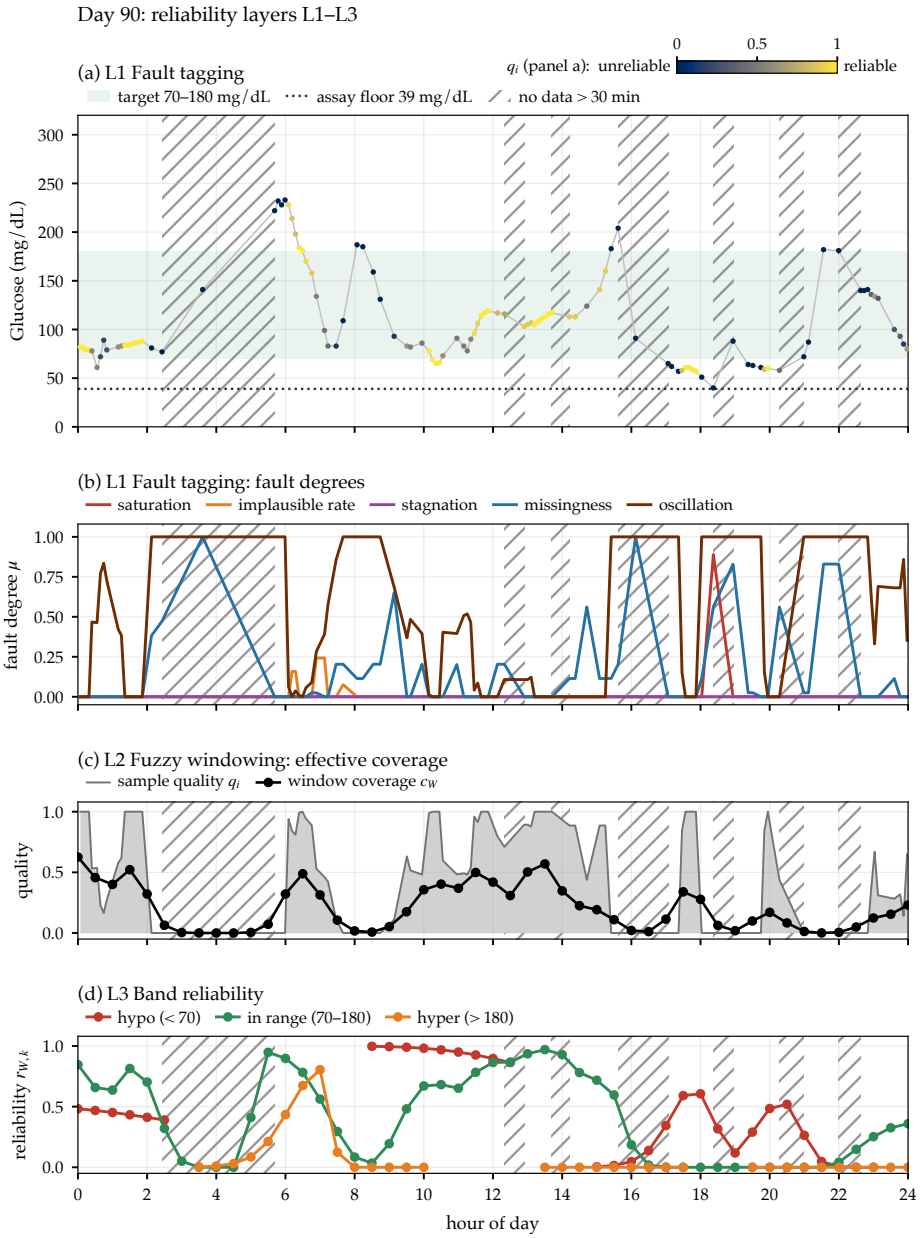


Figure 6. Reliability layers L1–L3 on a day with long *missing* stretches (Day 90), with the layout of Figure 4. Panels: (**a**) readings coloured by q_i (colour bar for panel (**a**) only); (**b**) fault degrees; (**c**) q_i (grey area) and window coverage c_W (dots); (**d**) band reliability $r_{W,k}$; hatched areas, no readings for more than 30 min. Effective coverage falls where the sensor is absent, so the aggregates are marked incomplete rather than interpolated.

3.4. Naïve Versus Reliability-Aware Profiles

Figure 8 contrasts the naïve and reliability-aware profiles of the ten reference days, with each metric on its own scale. On the clean control days, the two profiles coincide and the global veracity is high: there is nothing to correct and the method says so. On the corrupted days, the profiles separate, chiefly in TBR and TAR, where floor-clamped, oscillating or anomalous readings are discounted, and the global veracity drops; on three of them (Days 90, 245 and 184), it falls below $V = 0.5$ so that no profile would be drawn at

all. On Day 352, a naïve TBR of 16% (enough to prompt a change of therapy) shrinks to 3%, and the low-range estimate carries a veracity of 0.14. Disregarding the reconstruction uncertainty ($\kappa_i = 1$; Section 2.5) changes these profiles by at most 1.4 percentage points and the global veracity by at most 0.02.

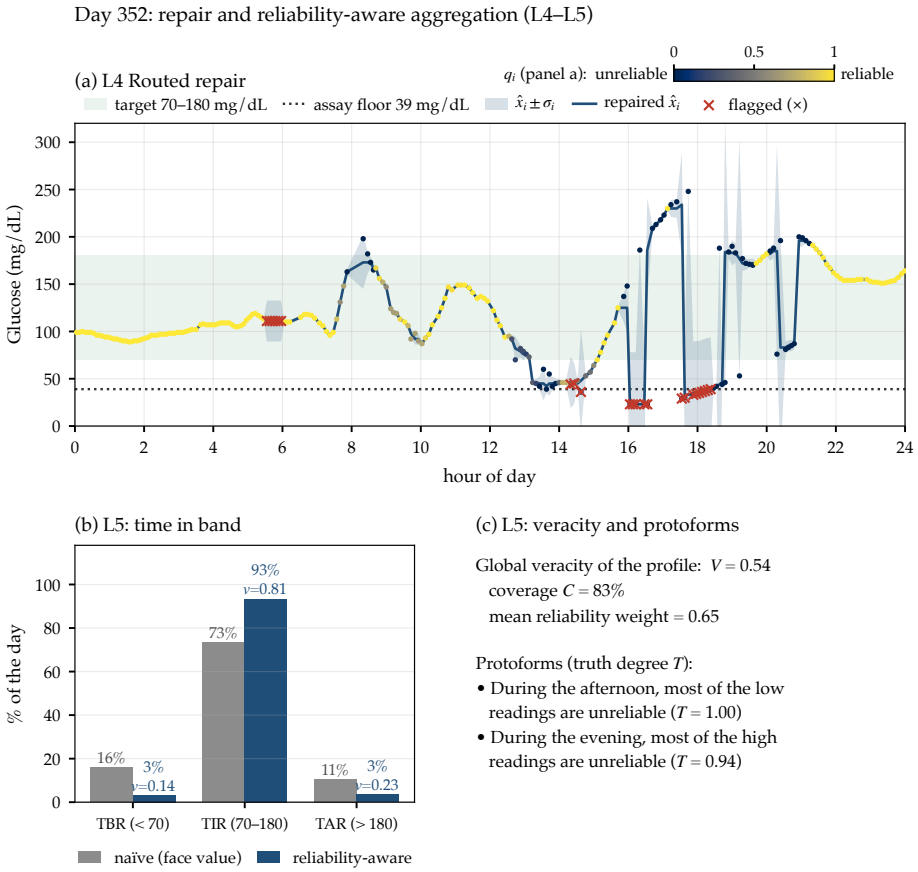


Figure 7. Repair and reliability-aware aggregation (L4–L5) on the most corrupted reference day (Day 352). (**a**) L4 routed repair: raw readings coloured by their quality q_i (the colour bar applies to this panel only), repaired values $\hat{x}_i$ (line) with their $\pm\sigma_i$ reconstruction band (shaded), and readings at the assay floor or stuck at a constant value, flagged (×) (Equation (18)); the green band marks the 70–180 mg/dL target range and the dotted line the assay floor. (**b**) TBR, TIR and TAR from the naïve profile, which takes every reading at face value (grey), and from the reliability-aware profile (blue), annotated with the band veracity v_k (Equations (13) and (14)). (**c**) Global veracity V (Equation (15)), with the coverage C and the mean reliability weight, and the protoforms emitted with their truth degrees (Equation (17)).

Because the profile veracity folds the day's coverage into its quality (it is the mean quality scaled by the fraction of the day that is actually recorded), it also flags days that are clean but sparsely worn: the second control day (Day 72), recorded for only 69% of its hours, scores a veracity of 0.68 despite pristine samples.

3.5. Impact on Phenotyping over the Whole Recording

Figure 9 shows the daily phenotypes over the whole recording (Section 2.6). The naïve profiles separate into three phenotypes: an in-range phenotype (145 days; mean TBR 14.5%, TIR 82.2%, TAR 3.3%), a hyper-dominant one (50 days; 14.7%, 63.1%, 22.1%) and a hypo-dominant one (99 days; 42.3%, 56.0%, 1.7%). Once reliability is taken into account, 112 of the 294 days (38%) have $V < 0.5$ and receive no phenotype; the remaining 182 days again form three phenotypes with similar centroids (in-range: 99 days, 13.0%, 84.4%, 2.6%; hyper-dominant: 30 days, 14.8%, 64.0%, 21.2%; hypo-dominant: 53 days, 37.4%, 61.5%, 1.2%).

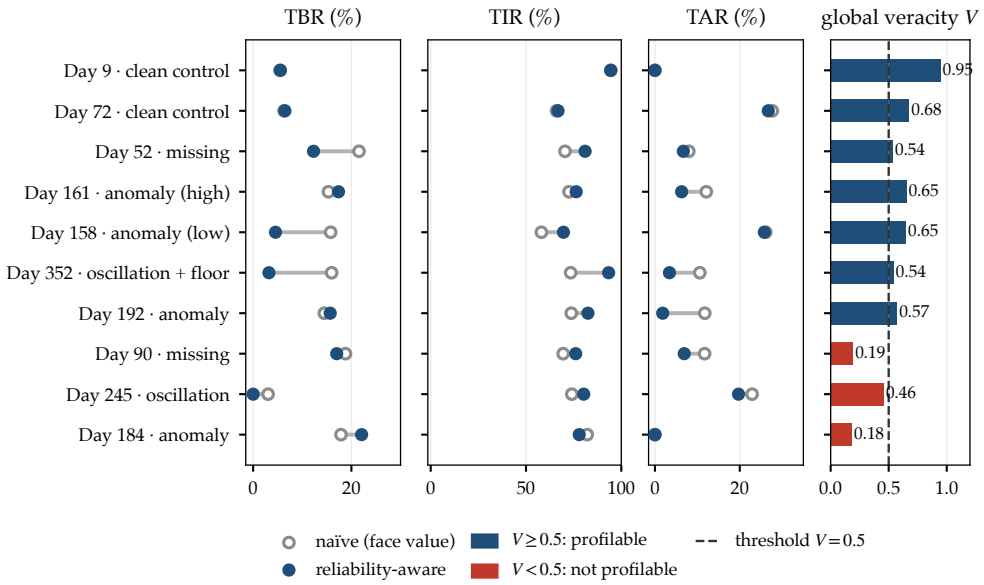


Figure 8. Naïve versus reliability-aware profiles of the ten reference days (Table 3). Rows: day of the recording and its fault character. First three panels: TBR, TIR, TAR, Equation (13), each on its own percentage scale; open grey circles, naïve profile (every reading at face value); filled blue circles, reliability-aware profile; the grey segment joins the two estimates of the same day. Right panel: global veracity V of the reliability-aware profile (Equation (15)); days below the dashed threshold $V = 0.5$ (red) are considered too unreliable to profile.

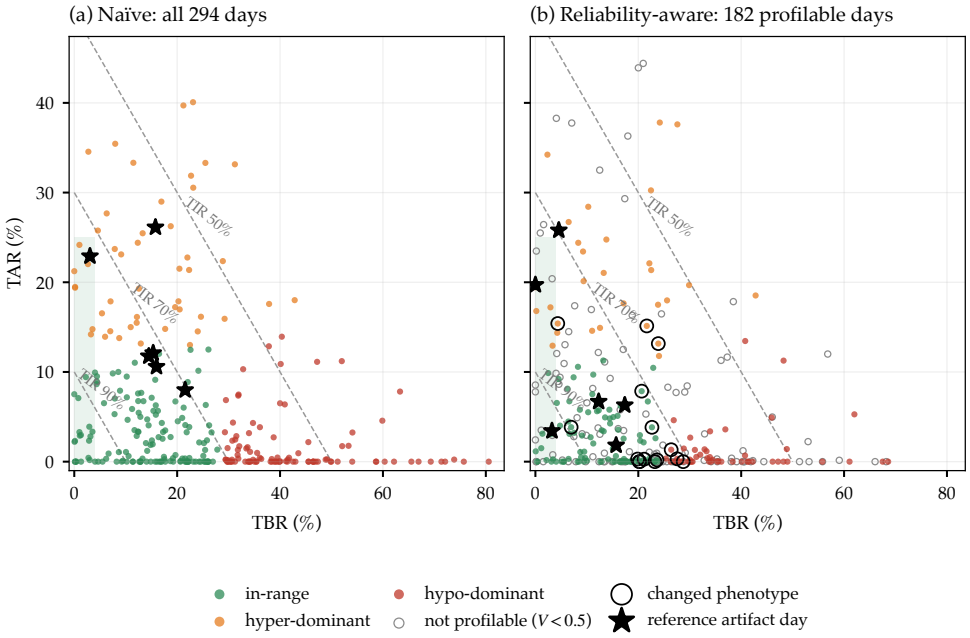


Figure 9. Daily glycemic phenotypes over the 294 eligible days, in the plane of TBR and TAR; TIR is implicit (dashed iso-TIR lines), and the shaded corner marks the consensus targets TBR $< 4\%$ and TAR $< 25\%$. Colours: phenotypes obtained by k-means ($k = 3$) and matched between panels (Section 2.6); the number of days in each phenotype is given in Section 3.5. (**a**) Naïve profiles: every eligible day is phenotyped. (**b**) Reliability-aware profiles: days with global veracity $V < 0.5$ (open grey circles) receive no phenotype; black rings mark profilable days whose phenotype changed with respect to (**a**). Stars: reference artifact days (Table 3).

Of the 294 days, 167 (57%) keep their naïve phenotype, 15 (5%) change it and 112 (38%) become non-profilable, so that 43% of the days receive a different characterization than in the naïve analysis. Phenotype changes occur at the boundary between the in-range and

hypo-dominant phenotypes (6 days leave the in-range phenotype and 9 the hypo-dominant one), whereas non-profilable days come from all three phenotypes (34%, 46% and 40% of the in-range, hyper-dominant and hypo-dominant days, respectively). Of the six eligible reference artifact days, five keep their phenotype and one (Day 245) becomes non-profilable. Of the 112 non-profilable days, 60 would be excluded by their coverage alone ($C < 0.5$), whereas 10 meet the 70% active-time criterion of the consensus [1] but have low sample reliability; conversely, 33 profilable days have between 42% and 70% coverage and would be discarded by that criterion. Disregarding the reconstruction uncertainty ($\kappa_i = 1$) yields 174, 16 and 104 days that keep, change and lose their phenotype, respectively.

4. Discussion

4.1. Principal Findings

The results show that reliability-aware processing alters the knowledge derived from CGM at two scales. At the scale of a single day, it leaves clean days essentially untouched (Days 9 and 72; Figure 8) while correcting metrics distorted by artifacts: on Day 352, a naïve TBR of 16% falls to 3% once floor-clamped and oscillating readings are discounted, and the low-range estimate is explicitly marked as untrustworthy ($v = 0.14$). At the scale of the whole recording, the main effect is not that days change phenotype (only 5% do, all at the boundary between the in-range and hypo-dominant phenotypes) but that 38% of the eligible days carry too little reliable information to be phenotyped at all; in total, 43% of the days receive a different characterization from that of the naïve analysis. The phenotypes themselves are stable, since their centroids barely move once unreliable days are excluded: the patient's glycemic profile is not an artifact of corrupted data, but a substantial part of the evidence upon which a naïve analysis rests is.

4.2. Graded Versus Crisp Sufficiency Criteria

The global veracity behaves as a graded counterpart of the 70% active-time criterion of the international consensus [1]. The two criteria largely agree, but they disagree in an informative way: 33 days with 42–70% coverage but reliable readings would be discarded by the crisp rule and are retained by the veracity, whereas 10 days that satisfy the rule are excluded because their readings are unreliable. Coverage alone, which is what the consensus assesses, is therefore neither sufficient nor necessary to decide whether a day can be interpreted: the quality of what was recorded matters as much as how much was recorded.

4.3. Comparison with CGM Signal Processing

These results are consistent with, and extend, previous work on CGM data processing. Denoising approaches, from adaptive Kalman filtering [11] and smart-sensor cascades [2,10] to the Bayesian treatment of coloured noise [12], improve the signal but return a cleaned series without stating how much of it was inferred; deep imputation of corrupted stretches [13] likewise reconstructs values without carrying their uncertainty into the clinical metrics. Fault-detection methods for CGM, based on statistical criteria [15], model-based detection for closed-loop control [16], clustering [17] or explicit models of disconnections and compression [14], share our premise that faults must be identified on the raw signal, but they produce sample-level flags aimed at real-time use. The present pipeline differs in what it delivers rather than in its ability to detect a given fault: graded fault degrees are carried forward to the aggregates that clinicians actually read, so that the effect of a fault on TBR, TIR or TAR is quantified and stated (Figure 7). The observation that a stretch of floor-clamped readings can inflate the TBR of a day several-fold also complements reports that CGM-derived metrics depend on the sensor and carry a systematic

bias [3,4]: besides device-level bias, day-level faults distort the same metrics, and they can be detected from the data themselves.

4.4. Comparison with Data-Quality and Uncertainty-Aware Approaches

Outside glucose, signal-quality indices for wearable photoplethysmography grade each segment and route the subsequent processing [18,19], an idea close to our fault-routed repair; data-quality frameworks for the IoT, wearables and industrial monitoring attach reliability flags to the stream [21–23,25]; and uncertainty-aware imputation returns a variance with each reconstructed value [29–31]. In these approaches, however, quality is a scalar per segment or per stream, and the uncertainty is consumed by the reconstruction, as with the reliability index of Hosseini and Moeini [32], rather than conditioned on the value range and attached to the result. The band-conditioned veracity addresses precisely this gap: Days 158 and 352 are reliable in range but unreliable in the low band (Figure 7), a situation that a single quality score per day cannot express. As for the propagated reconstruction uncertainty, the sensitivity analysis shows that its contribution is small in this recording: at most, 1.4 percentage points in the reference profiles and eight non-profilable days over the recording. The readings that receive a large reconstruction variance are precisely those that are already discounted by L1, so the sample quality dominates the result. The propagation therefore acts as a safeguard that becomes relevant when repair is applied to moderately faulty readings or when more aggressive estimators, such as imputers of long gaps, are used, rather than as the main source of the observed effects. Finally, the protoforms reorient fuzzy linguistic summarization [33,37] from the monitored phenomenon, the subject of previous summaries of activity and physiological streams [38,39,41], to the trustworthiness of its description, in statements short enough to accompany a clinical report.

4.5. Clinical Significance

The most consequential distortions concern the low range. A TBR of 16%, as reported naïvely for Day 352, would exceed the consensus target of less than 4% [1] four-fold and could prompt a reduction in insulin, whereas the reliability-aware estimate of 3% lies within target. More generally, a report that pools days of very different reliability conveys a precision that the data do not support; identifying the days that should not be profiled, and stating which bands and periods are unreliable, would allow a clinician to discount them or to request a better-worn period before adjusting therapy. These implications must nonetheless be read with caution: they derive from a single participant, the effect of the corrected metrics on clinical decisions has not been assessed, and whether the proportion of unreliable days generalizes to other patients, sensors and wearing habits remains to be established.

4.6. Generalizability

The pipeline is domain-agnostic in its structure, but its transfer to another sensor rests on explicit assumptions: (i) the stream is a timestamped scalar series with a known nominal cadence Δ; (ii) the sensor has a bounded measurement range at which readings saturate; (iii) a maximum plausible rate of change ρ can be stated from the physics or physiology of the measured quantity; and (iv) the value axis can be partitioned into bands that are meaningful to the end user. Under these assumptions, the saturation, rate and missingness primitives transfer by re-parameterization, whereas stagnation and oscillation require two shape parameters each that encode knowledge of the sensor's failure signatures, and the bands require domain-specific thresholds. Failure modes that preserve physiological plausibility, such as compression artifacts [14] or slow drift, are not captured by the five primitives and would require additional ones. Since the method has been evaluated only on

CGM data, its applicability to other sensors is a hypothesis supported by its construction, not a demonstrated result.

4.7. Limitations

The study has further limitations. First, it rests on a single long recording; although 294 days cover a broad variety of daily conditions, the results cannot be generalized to other patients without a multi-subject study. Second, no independent reference standard was available: capillary or laboratory measurements were not collected, and the manufacturer's cloud holds the only copy of the series. The method therefore assesses the plausibility and completeness of the data, not the accuracy of the sensor against blood glucose, which rests on the clinical validation of the factory-calibrated device and on the participant's routine verification of its correct functioning. Third, the reference days were labeled with the same clinical knowledge that the fault rules encode, so agreement on them is a consistency check rather than an independent validation. Fourth, the fault thresholds, band boundaries and protoform templates are fixed a priori from expert rules; their shapes are reasonable but not unique. Fifth, acquisition interruptions are not logged separately and cannot be distinguished from sensor-side gaps. Finally, the phenotyping relies on k-means over two time-in-band features, a deliberately simple choice that captures coarse phenotypes; richer representations could reveal changes that the present clustering does not.

5. Conclusions

This work has shown that, in continuous monitoring, the reliability of the data is better treated as a field over the time and value range than as a single figure, and that doing so changes the knowledge derived from a sensor. In light of the results, its contributions are threefold. First, formulating reliability as a fuzzy, time- and band-conditioned quantity computed on the raw signal made it possible to establish that a day can be reliable in one glucose band and unreliable in another, such as on Days 158 and 352. Second, the five-layer pipeline, together with the acquisition architecture that feeds it, carried this judgment through to the clinical aggregates: clean days were left essentially untouched, while artifact-inflated metrics were corrected (a TBR of 16% reduced to 3% on the most corrupted day) and annotated with veracity degrees and protoforms stating which bands and periods cannot be believed. Third, over 294 days of a 14-month recording, the global veracity acted as a graded counterpart of the consensus 70% active-time rule: 38% of the days were judged to be too unreliable to profile, 5% changed phenotype, and the graded criterion retained or rejected days that the crisp rule would have treated otherwise. Rather than a single, falsely precise number, the method returns an aggregate, its veracity, and a sentence a practitioner can read. Future work will validate the pipeline in a multi-subject cohort, ideally with reference blood-glucose measurements, extend the set of primitives to plausibility-preserving faults such as compression artifacts, and test its transfer to sensing domains beyond glucose. Since the acquisition tier already enumerates every patient who authorizes its account, extending the study to a cohort is a matter of scale rather than of redesign.

Author Contributions: Conceptualization, J.F.G.-G., C.M.-C. and J.L.L.; Methodology, J.F.G.-G., C.M.-C. and J.L.L.; Software, J.F.G.-G.; Formal analysis, J.F.G.-G.; Validation, C.M.-C. and J.L.L.; Writing—original draft, J.F.G.-G.; Writing—review and editing, C.M.-C. and J.L.L.; Supervision, M.E. and J.L.L.; Funding acquisition, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially supported by grant PID2024-156412OB-I00, funded by MICIU/AEI/10.13039/501100011033, and by "ERDF/EU" and grant M.2 PDC_000756_UJA23,

funded by Consejería de Universidad, Investigación e Innovación and the ERDF Andalusia Program 2021–2027.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the corresponding ethics committee (protocol code DIC.21/10.PRY, approved in 2022).

Informed Consent Statement: Informed consent was obtained from the participant involved in the study, as required by the ethics committee approval.

Data Availability Statement: The glucose data analyzed in this study are openly available in the Zenodo (T1GDUJA dataset [42]). The code used to process the data and generate the results is available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
ARI	Adjusted Rand Index
BLE	Bluetooth Low Energy
CGM	Continuous Glucose Monitoring
HTTPS	Hypertext Transfer Protocol Secure
IoT	Internet of Things
JSON	JavaScript Object Notation
MAD	Median Absolute Deviation
ML	Machine Learning
REST	Representational State Transfer
TAR	Time Above Range
TBR	Time Below Range
TIB	Time in Band
TIR	Time in Range
TLS	Transport Layer Security
UTC	Coordinated Universal Time

References

1. Battelino, T.; Danne, T.; Bergenstal, R.M.; Amiel, S.A.; Beck, R.; Biester, T.; Bosi, E.; Buckingham, B.A.; Cefalu, W.T.; Close, K.L.; et al. Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range. *Diabetes Care* **2019**, *42*, 1593–1603. [CrossRef] [PubMed]
2. Facchinetti, A.; Sparacino, G.; Cobelli, C. Signal processing algorithms implementing the "smart sensor" concept to improve continuous glucose monitoring in diabetes. *J. Diabetes Sci. Technol.* **2013**, *7*, 1308–1318. [CrossRef] [PubMed]
3. Pleus, S.; Kamecke, U.; Waldenmaier, D.; Link, M.; Zschornack, E.; Jendrike, N.; Haug, C.; Freckmann, G. Time in specific glucose ranges, glucose management indicator, and glycemic variability: Impact of continuous glucose monitoring (CGM) system model and sensor on CGM metrics. *J. Diabetes Sci. Technol.* **2021**, *15*, 1104–1110. [CrossRef] [PubMed]
4. Hutchins, K.M.; Betts, J.A.; Thompson, D.; Hengist, A.; Gonzalez, J.T. Continuous glucose monitor overestimates glycemia, with the magnitude of bias varying by postprandial test and individual–a randomized crossover trial. *Am. J. Clin. Nutr.* **2025**, *121*, 1025–1034. [CrossRef] [PubMed]
5. Poupry, S.; Medjaher, K.; Béler, C. Data reliability and fault diagnostic for air quality monitoring station based on low cost sensors and active redundancy. *Measurement* **2023**, *223*, 113800. [CrossRef]
6. Pian, Y.; Wang, T.; Zhang, S.; Xu, R.; Liu, Y. Tackling Incomplete Data in Air Quality Prediction: A Bayesian Deep Learning Framework for Uncertainty Quantification. *arXiv* **2025**, arXiv:2511.02175.
7. Zhang, J.; Huang, M.; Wan, N.; Deng, Z.; He, Z.; Luo, J. Missing measurement data recovery methods in structural health monitoring: The state, challenges and case study. *Measurement* **2024**, *231*, 114528. [CrossRef]
8. Jihani, N.; Kabbaj, M.N.; Benbrahim, M. Kalman filter based sensor fault detection in wireless sensor network for smart irrigation. *Results Eng.* **2023**, *20*, 101395. [CrossRef]

9. Hosamo, H.; Hosamo, M.; Dai, J.; Jiang, Z. Efficient monitoring and prediction of the motion responses of compliant floating photovoltaic arrays: Case study of a hydrodynamic model test with erroneous sensor data. *Results Eng.* **2025**, *28*, 107625. [CrossRef]
10. Facchinetti, A.; Sparacino, G.; Cobelli, C. CGM filtering and denoising techniques. In *Glucose Monitoring Devices*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 203–218.
11. Facchinetti, A.; Sparacino, G.; Cobelli, C. Online denoising method to handle intraindividual variability of signal-to-noise ratio in continuous glucose monitoring. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2664–2671. [CrossRef] [PubMed]
12. Camerlingo, N.; Siviero, I.; Vettoretti, M.; Sparacino, G.; Del Favero, S.; Facchinetti, A. Bayesian denoising algorithm dealing with colored, non-stationary noise in continuous glucose monitoring timeseries. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1280233. [CrossRef] [PubMed]
13. Sanchez, O.D.; Mendez-Palos, E.; Pascoe, D.A.; Hernandez, H.M.; Alvarez, J.G.; Alanis, A.Y. Online Imputation of Corrupted Glucose Sensor Data Using Deep Neural Networks and Physiological Inputs. *Algorithms* **2025**, *18*, 688. [CrossRef]
14. Facchinetti, A.; Del Favero, S.; Sparacino, G.; Cobelli, C. Modeling transient disconnections and compression artifacts of continuous glucose sensors. *Diabetes Technol. Ther.* **2016**, *18*, 264–272. [CrossRef] [PubMed]
15. Zhao, C.; Fu, Y. Statistical analysis based online sensor failure detection for continuous glucose monitoring in type I diabetes. *Chemom. Intell. Lab. Syst.* **2015**, *144*, 128–137. [CrossRef]
16. Yu, X.; Rashid, M.; Feng, J.; Hobbs, N.; Hajizadeh, I.; Samadi, S.; Sevil, M.; Lazaro, C.; Maloney, Z.; Cinar, A. Fault detection in continuous glucose monitoring sensors for artificial pancreas systems. *IFAC-PapersOnLine* **2018**, *51*, 714–719. [CrossRef]
17. Yu, X.; Sun, X.; Zhao, Y.; Liu, J.; Li, H. Fault detection of continuous glucose measurements based on modified k-medoids clustering algorithm. *Neural Comput. Appl.* **2025**, *37*, 29089–29098. [CrossRef]
18. Beh, W.K.; Yang, Y.C.; Wu, A.Y. Quality-aware signal processing mechanism of PPG signal for long-term heart rate monitoring. *Sensors* **2024**, *24*, 3901. [CrossRef] [PubMed]
19. Basile, A.; Garozzo, U.; Andronaco, S.; Castellano, M.; Grasso, A.D. A Signal Quality Assessment Algorithm for Photoplethysmographic Sensors: Extended Version. *Chips* **2026**, *5*, 17. [CrossRef]
20. Fadlallah, H.; Kilany, R.; Dhayne, H.; El Haddad, R.; Haque, R.; Taher, Y.; Jaber, A. Context-aware big data quality assessment: A scoping review. *ACM J. Data Inf. Qual.* **2023**, *15*, 25. [CrossRef]
21. Kuemper, D.; Iggena, T.; Toenjes, R.; Pulvermueller, E. Valid. IoT: A framework for sensor data quality analysis and interpolation. In *Proceedings of the 9th ACM Multimedia Systems Conference*; ACM: New York, NY, USA, 2018; pp. 294–303.
22. Klima, M.; Rechtberger, V.; Bures, M.; Bellekens, X.; Hindy, H.; Ahmed, B.S. Quality and reliability metrics for IoT systems: A consolidated view. In *International Summit Smart City 360°*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 635–650.
23. Böttcher, S.; Vieluf, S.; Bruno, E.; Joseph, B.; Epitashvili, N.; Biondi, A.; Zabler, N.; Glasstetter, M.; Dümpelmann, M.; Van Laerhoven, K.; et al. Data quality evaluation in wearable monitoring. *Sci. Rep.* **2022**, *12*, 21412. [CrossRef] [PubMed]
24. Buelvas, J.; Múnera, D.; Tobón V, D.P.; Aguirre, J.; Gaviria, N. Data quality in IoT-based air quality monitoring systems: A systematic mapping study. *Water Air Soil Pollut.* **2023**, *234*, 248. [CrossRef]
25. Peixoto, T.; Oliveira, B.; Oliveira, Ó.; Ribeiro, F. Data quality assessment in smart manufacturing: A review. *Systems* **2025**, *13*, 243. [CrossRef]
26. Amiri, M.; Jensen, R. Missing data imputation using fuzzy-rough methods. *Neurocomputing* **2016**, *205*, 152–164. [CrossRef]
27. Liu, X.; Wang, B.; Jin, S.; Song, Z. Missing value interpolation algorithm for long-term temperature observation data based on data augmentation multiple interpolation method. *Results Eng.* **2025**, *27*, 106211. [CrossRef]
28. Velayudhan, N.K.; Devidas, A.R.; Savić, D. Generative AI for spatio-temporal multivariate imputation and demand prediction in water distribution systems. *Results Eng.* **2025**, *27*, 106178. . Erratum in *Results Eng.* **2026**, *31*, 111685. [CrossRef]
29. Mulyadi, A.W.; Jun, E.; Suk, H.I. Uncertainty-aware variational-recurrent imputation network for clinical time series. *IEEE Trans. Cybern.* **2021**, *52*, 9684–9694. [CrossRef] [PubMed]
30. Septiandri, A.A.; Ming, D.; DiazDelaO, F.A.; Jendoubi, T.; Ray, S. Integrative analysis and imputation of multiple data streams via deep Gaussian processes. *Bioinform. Adv.* **2026**, *6*, vbaf305. [CrossRef] [PubMed]
31. Weatherhead, A.; Goldenberg, A. Impute With Confidence: A Framework for Uncertainty Aware Multivariate Time Series Imputation. *arXiv* **2025**, arXiv:2507.09353.
32. Hosseini, S.H.; Moeini, R. Correlation-based reliability index equipped with machine learning methods to complete the groundwater level gaps. *Results Eng.* **2025**, *25*, 104146. [CrossRef]
33. Kacprzyk, J.; Wilbik, A.; Zadrożny, S. Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets Syst.* **2008**, *159*, 1485–1499. [CrossRef]
34. Kacprzyk, J.; Wilbik, A. Linguistic summaries of time series using a degree of appropriateness as a measure of interestingness. In *Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications*; IEEE: Piscataway, NJ, USA, 2009; pp. 385–390.

35. Kacprzyk, J.; Wilbik, A. Towards an efficient generation of linguistic summaries of time series using a degree of focus. In *Proceedings of the NAFIPS 2009–2009 Annual Meeting of the North American Fuzzy Information Processing Society*; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6.
36. Kacprzyk, J.; Wilbik, A. Temporal linguistic summaries of time series using fuzzy logic. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 436–445.
37. Ramos-Soto, A.; Martin-Rodilla, P. Enriching linguistic descriptions of data: A framework for composite protoforms. *Fuzzy Sets Syst.* **2021**, *407*, 1–26. [CrossRef]
38. Peláez-Aguilera, M.D.; Espinilla, M.; Fernandez Olmo, M.R.; Medina, J. Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease. *Complexity* **2019**, *2019*, 2694126. [CrossRef]
39. López-Medina, M.; Espinilla, M.; Cleland, I.; Nugent, C.; Medina, J. Fuzzy cloud-fog computing approach application for human activity recognition in smart homes. *J. Intell. Fuzzy Syst.* **2020**, *38*, 709–721. [CrossRef]
40. Medina, J.; Espinilla, M.; Zafra, D.; Martínez, L.; Nugent, C. Fuzzy fog computing: A linguistic approach for knowledge inference in wearable devices. In *Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 473–485.
41. Albín-Rodríguez, A.P.; Ricoy-Cano, A.J.; de-la Fuente-Robles, Y.M.; Espinilla-Estévez, M. Fuzzy protoform for hyperactive behaviour detection based on commercial devices. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6752. [CrossRef] [PubMed]
42. Gaitán Guerrero, J.F.; López Ruiz, J.L.; Martínez Cruz, C.; Espinilla Estévez, M. T1GDUJA: Glucose dataset of a patient with type 1 diabetes mellitus. In *IEEE Internet of Things Journal*; Zenodo: Geneva, Switzerland, 2024. [CrossRef]
43. Berwick, R. PyLibreLinkUp: A Python Client for the LibreLinkUp Service. Software Library. 2024. Available online: https://github.com/robberwick/pylibrelinkup (accessed on 24 June 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.