

ORIGINAL RESEARCH

Open access

A Conceptual Physics-Guided Generative Framework for Synthetic Smart-Meter Data Reconstruction Under Grid-Edge Privacy and Demand Response Verification Constraints

Luca Ferraro^{1*}, Matteo Ricci¹, Giulia Moretti²

Abstract

Smart-meter data often contain missing intervals that hamper analytics, billing, and demand response verification. Privacy regulations and consumer confidentiality expectations further restrict access to raw grid-edge data, even though accurate reconstruction remains essential for operational decision-making. Existing imputation methods may restore missing values but often ignore the physical constraints of electricity demand. At the same time, methods that rely on real customer records can create privacy risks when individual consumption patterns are exposed or reconstructed too precisely. This article proposes a conceptual physics-guided generative framework for reconstructing missing smart-meter data using synthetic load profiles. The framework is designed to avoid dependence on real consumer records while supporting demand response verification. The proposed framework could combine generative modelling with physical and operational constraints such as non-negative demand, ramp-rate limits, and daily periodic load patterns. It would use synthetic load profiles to represent realistic demand behaviour while preserving privacy. Benchmark approaches are discussed qualitatively rather than experimentally. The expected behaviour of the framework is described in terms of reconstruction fidelity, privacy preservation, physical consistency, and demand response verification usefulness. Four original tables summarise missing-data impacts, privacy risks, qualitative benchmark comparisons, and proposed evaluation metrics. The proposed framework offers a privacy-preserving and physics-aware pathway for smart-meter data reconstruction. It is intended as a conceptual foundation for future validation using synthetic, public, or utility-approved datasets.

Keywords Smart-meter data reconstruction, Physics-guided generative model, Synthetic data, Grid-edge privacy, Demand response verification

*Correspondence:

Luca Ferraro

luca.ferraro@gmail.com

¹ Department of Smart Grid Systems, Faculty of Engineering, University of Bologna, Bologna, Italy

² Department of Energy Management and Optimization, Faculty of Engineering, Polytechnic University of Turin, Turin, Italy

Introduction

Smart-meter data have become foundational to modern grid analytics because they enable higher-resolution visibility into electricity consumption, distribution-system operation, and customer-side demand behaviour. Reviews of smart-meter analytics emphasise their role in forecasting, anomaly detection, demand response, customer segmentation, and distribution-grid planning [1]. However, the same temporal granularity that makes smart-meter data analytically valuable

also makes missing or corrupted intervals especially disruptive. This article therefore treats missing-data reconstruction not as a narrow preprocessing task, but as a methodological requirement for trustworthy grid-edge analytics.

Missing smart-meter data can arise from communication interruptions, device failures, data-transfer errors, cyber incidents, or downstream database inconsistencies. Prior work on electricity-consumption data quality has shown that incomplete and unreliable data can undermine both operational analytics and customer-facing applications [2]. Imputation studies further indicate that reconstruction quality depends on the duration, pattern, and context of missing intervals rather than on a single universal method [3]. A conceptual framework for reconstruction must therefore consider the temporal structure of load data and the operational consequences of each missing segment.

Privacy constraints complicate reconstruction because smart-meter traces may reveal occupancy, appliance use, behavioural routines, or other sensitive household information. Occupancy-detection studies demonstrate that even ordinary consumption records can support inferences about residential presence and activity [4]. Differential privacy and privacy-preserving metering research show that protecting customer information requires more than simply removing identifiers from the dataset [5]. For this reason, the proposed framework is designed around synthetic load profiles so that reconstruction logic can be developed without direct access to real customer records.

Demand response verification adds a further requirement because missing values can distort baseline estimation and event-period consumption comparison. Baseline-load estimation is central to measuring demand response performance, and review work has shown that baseline choice strongly affects estimated impact [6]. Customer-baseline-load methods and virtual control-group approaches likewise depend on complete and credible consumption histories [7]. This article proposes a conceptual physics-guided generative framework that would reconstruct missing intervals while preserving load realism, privacy, and demand response verification value.

Missing Smart-Meter Data Problem

Causes of missing data

Missing smart-meter readings may be caused by communication network failures, meter malfunction, clock synchronisation problems, storage errors, or disruptions in head-end and meter-data-management systems. Studies on smart-grid data quality identify missingness as one of several defects that can degrade the usability of consumption records for forecasting, billing, and operational analytics [2]. In a conceptual reconstruction framework, these causes matter because random single-point gaps, repeated communication outages, and event-period losses have different analytical implications. The framework therefore treats the cause of missingness as contextual information that could guide reconstruction assumptions.

Temporal patterns of missing intervals

Missing intervals can occur as isolated points, short contiguous gaps, long outages, recurring daily gaps, or missing blocks around peak and event periods. Autoencoder-based and practical imputation studies suggest that reconstruction difficulty increases when missingness removes the local temporal context needed to infer load transitions [3, 8]. In smart-meter records, the timing of a missing segment can be as important as its length because morning ramps, evening peaks, and overnight low-load periods have different structural regularities. A physics-guided generative approach would therefore be expected to condition reconstruction on both neighbouring observations and expected daily load-shape behaviour.

Impacts on load forecasting and billing

Missing smart-meter values can bias feature extraction, distort customer baselines, and weaken load forecasting models that depend on recent consumption histories. Mixture factor analysis and shape-modelling approaches to electric-load

imputation show that missing segments can alter the inferred structure of consumption profiles if they are not reconstructed in a context-aware manner [9, 10]. Billing applications may also be affected when missing values coincide with tariff periods, peak windows, or settlement intervals. For these reasons, the proposed framework would prioritise physically plausible reconstructions rather than merely smooth numerical substitutions.

Impacts on demand response verification

Demand response verification is especially sensitive to missing data because the comparison between baseline and event-period consumption depends on accurate temporal alignment. Customer baseline research shows that estimated savings can change when reference profiles or control groups do not adequately represent expected consumption [6, 7]. If missing intervals occur during the pre-event baseline window or the event itself, reconstruction error could be mistaken for load reduction or non-performance. These issues are summarized in Table 1, which links missing-data causes to their typical grid-analytics and verification consequences.

Table 1. Causes and Impacts of Missing Smart-Meter Data

Cause or pattern of missing data	Typical location in smart-meter workflow	Likely analytical impact	Demand response verification implication	Conceptual reconstruction response
Communication outage	Meter-to-head-end transmission	Consecutive missing blocks and loss of local continuity	Event-period demand may appear lower or unavailable	Condition generation on pre-gap and post-gap observations
Meter malfunction	Device-level measurement	Repeated missing or implausible values	Customer baseline may become unreliable	Apply physical feasibility checks before reconstruction
Clock or timestamp error	Data ingestion and alignment	Shifted load profiles and distorted daily shape	Event window may be misaligned with true consumption	Enforce temporal consistency and daily periodicity
Database or transfer failure	Meter-data-management system	Random missing intervals across customers or feeders	Aggregated event performance may be biased	Use synthetic profile priors and uncertainty flags
Cyber or access disruption	Grid-edge communication layer	Missingness concentrated around operationally important periods	Verification may be contested if event-period evidence is incomplete	Preserve auditability of reconstructed intervals
Privacy-driven suppression	Data release or sharing stage	Masked or aggregated intervals	Baseline studies may lack customer-level detail	Use synthetic profiles instead of raw customer traces

Synthetic Smart-Meter Data Assumptions

Rationale for using synthetic load profiles

Synthetic load profiles offer a way to design and evaluate reconstruction logic without exposing real customer consumption records. Generative energy-data studies show that recurrent generative adversarial networks and related models can be used to create artificial consumption sequences that resemble energy-demand patterns [11, 12]. In the

proposed framework, synthetic data would not be treated as a perfect substitute for utility records, but as a privacy-preserving design environment. This assumption allows the article to remain conceptual while still specifying how reconstruction, constraints, and verification logic could interact.

Types of synthetic profiles considered

The framework could use synthetic profiles generated from stochastic models, wavelet-decomposed smart-meter patterns, generative adversarial networks, or inpainting-oriented demand models. Recent work on stochastic residential load generation, multiscale GAN-based profiles, and synthetic demand inpainting demonstrates multiple pathways for creating artificial yet structured consumption sequences [13–15]. These profiles could represent residential, commercial, or mixed grid-edge behaviours in a controlled conceptual setting. The important requirement is not that the synthetic profiles reproduce any specific customer, but that they preserve realistic temporal features relevant to reconstruction and verification.

Limitations of synthetic data

Synthetic load profiles can reduce privacy risks, but they may also underrepresent rare behaviours, unusual appliance patterns, extreme weather responses, or atypical demand response participation. Data-driven stochastic load-profile generation and GAN-based synthetic-data studies indicate that artificial profiles depend strongly on modelling assumptions and training design [16, 17]. A framework trained or reasoned entirely on synthetic profiles would therefore need explicit limits on interpretation, especially when transferring to utility operations. In this article, synthetic profiles are assumed to support conceptual design and benchmark planning rather than to establish real-world performance claims.

Grid-Edge Privacy Constraints

Privacy risks in raw smart-meter data

Raw smart-meter data can reveal sensitive household information because electricity demand reflects occupancy, routines, appliance operation, and behavioural timing. Privacy reviews of smart-meter systems emphasise that high-resolution metering can create risks beyond ordinary billing exposure [18]. Studies of occupancy detection from smart-meter data show that consumption traces may support inference about whether a dwelling is occupied [4]. Therefore, reconstruction should be designed so that it does not amplify the disclosure of individual behaviour.

Regulatory and societal expectations

Grid-edge data analytics must align not only with technical requirements but also with societal expectations of confidentiality, proportionality, and responsible data use. Differential privacy research on real smart-metering data illustrates how formal privacy mechanisms can limit the information disclosed through energy datasets [5]. At the same time, privacy-preserving reconstruction must remain useful enough for operational and verification tasks. The proposed framework addresses this tension by separating conceptual model development from raw customer data and by relying on synthetic profiles where possible.

How reconstruction can exacerbate privacy risks

Reconstruction can create new privacy risks if it fills missing intervals with values that make individual routines more complete and therefore easier to infer. Research on the privacy cost of fine-grained electrical consumption data shows that high-resolution traces can expose sensitive behavioural information even when direct identifiers are absent [19]. A generative model that restores suppressed or missing intervals too confidently could therefore increase the inferential value of a record. In the proposed framework, reconstruction outputs would be accompanied by uncertainty indicators and privacy-aware restrictions on individual-level disclosure.

Mitigation strategies in the framework

The framework proposes several conceptual mitigation strategies, including synthetic-profile training, privacy-preserving evaluation, restricted release of reconstructed records, and uncertainty-aware reporting. Secure and privacy-conscious smart-meter analytics can benefit from techniques such as differential privacy, aggregation, de-identification, and synthetic data generation, although each involves trade-offs [5, 18]. The reconstruction process would be designed to support demand response verification without recreating unnecessarily detailed behavioural traces. These risks and safeguards are organized in **Table 2**, which clarifies how privacy-aware reconstruction would operate at the grid edge.

Table 2. Privacy Risks and Mitigation Strategies in Smart-Meter Data

Privacy risk	How the risk arises in smart-meter reconstruction	Potential consequence	Mitigation strategy in the conceptual framework	Residual concern
Occupancy inference	Reconstructed intervals complete presence-related load patterns	Household presence may be inferred	Use synthetic profiles and limit individual-level disclosure	Aggregate trends may still reveal broad routines
Appliance-use inference	Fine-grained load shapes reveal device signatures	Sensitive appliance or activity patterns may be exposed	Apply temporal smoothing only where compatible with verification	Excessive smoothing may weaken event analysis
Re-identification	Synthetic or reconstructed traces resemble known customers	Supposedly anonymous records may be linked	Avoid direct use of raw customer data during conceptual design	External auxiliary data could still increase risk
Event participation disclosure	Demand response records reveal behavioural response	Customer choices or constraints may become visible	Report verification results at suitable aggregation levels	Program rules may require some customer-level auditability
Overconfident reconstruction	Missing values are filled without uncertainty context	Users may treat inferred values as observed facts	Attach reconstruction flags and uncertainty descriptions	Operational users may still overinterpret filled values
Model memorisation	Generative model reproduces training examples	Sensitive consumption sequences may leak	Prefer synthetic training and privacy-aware model governance	Future real-data validation would require safeguards

Demand Response Verification Need

Basics of demand response verification

Demand response verification estimates whether participating customers reduced or shifted electricity consumption during an event relative to an expected baseline. Reviews of baseline estimation methods show that verification depends on defining a credible counterfactual load profile for the same customer, group, or operating context [6]. In practice, this process requires temporally consistent smart-meter data before, during, and sometimes after the event. Missing data therefore affects not only analytics quality but also the fairness and credibility of program settlement.

Role of accurate baseline estimation

Baseline estimation is central because it represents what demand would likely have been without a demand response event. Virtual control-group methods have been proposed to improve residential baseline calculation by identifying comparison patterns that better represent expected consumption [7]. Graph-based aggregated baseline estimation further suggests that relationships among customers or loads can support demand response assessment at higher aggregation levels [20]. The proposed reconstruction framework would therefore seek to preserve baseline-relevant load shapes rather than optimising only pointwise imputation accuracy.

Impact of missing data on verification

Missing data can affect verification by removing pre-event reference intervals, masking event-period consumption, or distorting post-event rebound analysis. Battery energy-management imputation work illustrates that missing data can influence downstream control and evaluation tasks when reconstructed values are used as analytical inputs [21]. In demand response, a missing event-period interval could make a participant appear to reduce demand, fail to reduce demand, or become unverifiable depending on the reconstruction rule. A physics-guided generative method would be expected to reduce these ambiguities by generating plausible but clearly flagged estimates.

How reconstruction enables reliable verification

Reconstruction can support reliable verification when it restores missing intervals in a way that is temporally coherent, physically feasible, and compatible with baseline methodology. Big-data analytics research for future electricity grids highlights the importance of trustworthy data pipelines for advanced grid applications [22]. Edge-computing work also suggests that some analytics may increasingly occur closer to meters and distributed devices, making privacy-aware reconstruction relevant at the grid edge [23]. In the proposed framework, reconstructed intervals would support baseline estimation and event-period comparison while remaining distinguishable from directly observed smart-meter readings.

Physics-Guided Generative Model Design

Core generative architecture

The proposed framework could be implemented as a conditional generative model that receives an incomplete smart-meter sequence, a missingness mask, temporal indicators, and synthetic load-profile context as inputs. Variational autoencoder, denoising autoencoder, GAN, and diffusion-inspired designs would each provide a possible architecture for generating plausible missing intervals, as prior energy-data studies have shown that generative models can represent structured consumption sequences [11, 24]. Unlike a purely data-driven generator, the proposed architecture would include a physics-guided module that evaluates whether the reconstructed trajectory is operationally plausible. This design would make the generator responsible not only for filling gaps, but also for preserving the shape and feasibility of electricity demand.

Physical constraints on electricity demand

The framework proposes that reconstructed values should satisfy basic physical and operational constraints, including non-negative demand, bounded ramp rates, temporal continuity, and daily cyclic structure. Physics-informed machine learning has shown that neural models can be guided by physical laws, constraints, and residual penalties rather than trained only on observed samples [25]. In smart-meter reconstruction, these constraints would not require full power-system equations, but they would encode practical expectations about how customer load can change over time. Such constraints would be especially useful when neighbouring observations are sparse or when missing intervals coincide with peak and event periods.

Incorporating physics-guided penalties

Physics-guided penalties could be incorporated into the model objective so that implausible reconstructions receive higher loss even when they appear statistically likely. Broader physics-guided machine-learning research suggests that scientific and engineering priors can improve model consistency when data are noisy, incomplete, or insufficient [26, 27]. In this framework, one penalty could discourage negative load, another could penalise unrealistic ramps, and another could encourage alignment with daily or weekly profile structure. These penalties would be conceptual design elements rather than experimentally tuned parameters in this article.

Training with synthetic load profiles

Training would be expected to use synthetic load profiles with artificially masked intervals so that the model learns reconstruction behaviour without requiring direct access to raw customer data. Synthetic demand-generation studies using GANs, stochastic models, and inpainting approaches show that artificial consumption sequences can be constructed to represent load-shape variability while reducing dependence on confidential records [12, 15, 16]. The framework would therefore simulate missingness patterns over synthetic profiles and train the generator to recover masked segments under physical constraints. This training strategy would preserve the article's privacy-first orientation while creating a conceptual pathway for later validation with approved datasets.

Figure 1 illustrates the proposed physics-guided generative reconstruction architecture, showing how incomplete smart-meter sequences, missingness masks, synthetic load-profile context, and physical feasibility constraints are integrated to produce privacy-aware reconstructed intervals with uncertainty flags.

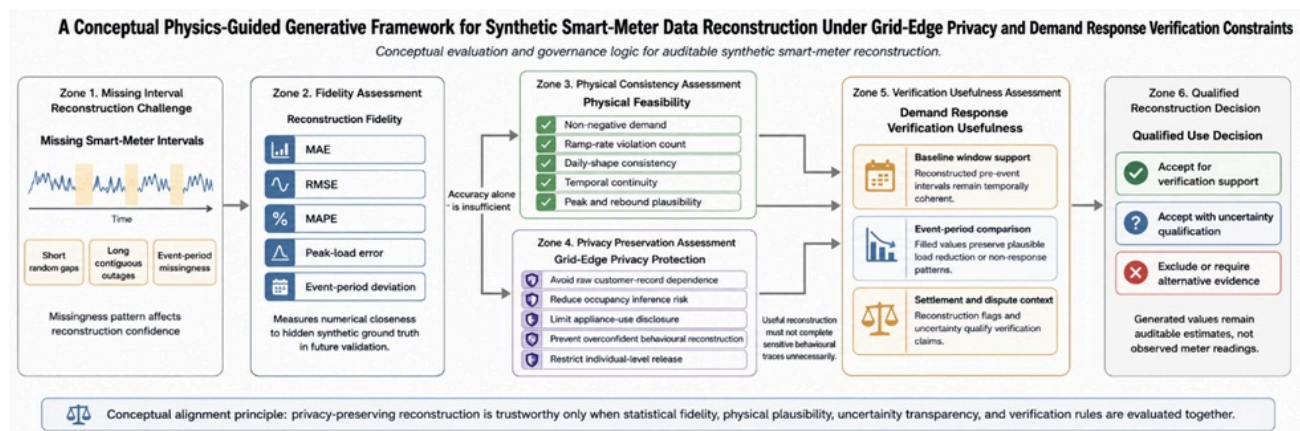


Figure 1. Physics-Guided Generative Architecture for Privacy-Preserving Smart-Meter Data Reconstruction

Generative Reconstruction Method

Conditioning on observed data

The reconstruction method would condition generation on observed readings before and after the missing interval, as well as on time-of-day, day-type, and event-status indicators. Denoising and masked-autoencoder approaches to electric-load imputation illustrate how observed context can guide reconstruction when parts of the sequence are hidden [3, 24]. In the proposed framework, the missingness mask would be treated as an explicit input rather than an accidental data defect. This would allow the model to distinguish between directly observed values and generated values throughout the reconstruction process.

Leveraging neighbouring observations

Neighbouring observations would provide local continuity information, helping the model infer whether a gap belongs to a stable overnight segment, a morning ramp, an evening peak, or a demand response event window. Mixture factor analysis and shape-modelling studies indicate that electric-load imputation benefits from using profile structure rather than isolated point estimation [9, 10]. The proposed method would therefore use pre-gap and post-gap values as anchors while allowing the generative component to infer the most plausible internal trajectory. This approach would be expected to reduce discontinuities at the boundaries of reconstructed intervals.

Load-shape priors and weather sensitivities

The framework could incorporate load-shape priors that reflect daily periodicity, weekday-weekend differences, seasonal tendencies, and possible weather sensitivity. Smart-meter analytics reviews show that consumption patterns are influenced by behavioural and contextual factors that affect forecasting, segmentation, and grid applications [1]. Although this article does not use real weather or customer data, the conceptual model could include synthetic weather-sensitivity labels or profile classes to test whether reconstruction remains coherent under different demand regimes. These priors would help the model avoid generic smoothing when the missing segment should contain a peak, ramp, or event-related reduction.

Output generation and uncertainty estimates

The output of the reconstruction method would include both a generated load trajectory and an uncertainty description that indicates how confidently the missing interval has been reconstructed. Privacy and verification concerns make this distinction important because reconstructed values should not be mistaken for directly observed meter readings [18, 19]. A diffusion-style or probabilistic generative model could produce multiple plausible trajectories, while an autoencoder-style model could be extended with uncertainty bands or reconstruction-quality flags. In this framework, uncertainty would support auditability, privacy-aware release, and demand response verification judgement.

Benchmark Models

Classical imputation methods

Classical benchmark methods would include mean substitution, last observation carried forward, linear interpolation, spline interpolation, and simple seasonal averaging. Practical imputation work for smart-meter data shows that such approaches can be attractive because they are transparent, easy to implement, and computationally inexpensive [8]. However, classical methods may struggle when gaps are long, when missingness occurs during peaks, or when event-period behaviour departs from normal daily patterns. In the proposed conceptual comparison, these methods would serve as baseline approaches against which privacy-aware and physics-guided reconstruction could be qualitatively assessed.

Matrix and tensor completion

Matrix and tensor completion benchmarks would represent approaches that exploit correlations across time, customers, feeders, or profile classes. Load-imputation studies based on latent structure suggest that missing values can be estimated by learning relationships among consumption profiles and temporal dimensions [9]. In a privacy-first framework, however, cross-customer completion may raise concerns if it requires access to identifiable customer matrices or reveals similarity relationships. Therefore, matrix and tensor completion would be discussed as useful conceptual baselines but less naturally aligned with synthetic, grid-edge privacy constraints.

Autoencoders and variational autoencoders

Autoencoders and variational autoencoders provide a middle ground between classical imputation and fully adversarial or diffusion-based generation. Denoising autoencoder and masked autoencoder studies show how neural models can learn compressed representations of electric-load sequences and reconstruct corrupted or missing regions [3, 24]. A variational

Generative adversarial networks and diffusion models

Table 3. Qualitative Comparison of Reconstruction Methods

Reconstruction method	Expected reconstruction fidelity	Privacy preservation	Physical consistency	Interpretability	Suitability for long gaps	Demand response verification support
Linear interpolation	Potentially reasonable for short smooth gaps	High if applied locally without sharing data	Limited because ramps may be unrealistic	High	Low	Limited when event periods contain peaks or reductions
Seasonal mean or profile averaging	Moderate when repeated daily patterns dominate	Moderate if based on aggregate profiles	Moderate daily-shape consistency	High	Moderate	Useful for baseline context but weak for event-specific variation
Matrix or tensor completion	Potentially strong when correlated profiles are available	Lower if customer-level matrices are required	Indirect and not guaranteed	Moderate	Moderate to high	Useful for aggregate baselines but privacy-sensitive
Autoencoders	Potentially strong for learned load shapes	Moderate if trained on synthetic or protected data	Limited unless constraints are added	Moderate	Moderate	Useful if reconstruction flags and uncertainty are provided
Generative adversarial networks	Potentially strong for realistic synthetic trajectories	Moderate to high when trained on synthetic profiles	Variable without explicit penalties	Low to moderate	Moderate	Useful for plausible scenarios but needs auditability

Diffusion models	Potentially strong for multiple plausible reconstructions	Moderate to high with synthetic training and sampling controls	Variable unless guided by constraints	Low to moderate	Moderate to high	Useful for uncertainty-aware verification
Physics-guided generative model	Expected to balance fidelity and feasibility	High under synthetic-profile training and privacy-aware release	High through non-negativity, ramp, and periodicity penalties	Moderate	Moderate to high	Strong because baseline and event-period plausibility are jointly considered

Expected Results and Evaluation Metrics

Expected reconstruction performance

The proposed framework would be expected to perform best when missing intervals retain sufficient surrounding context and when synthetic training profiles adequately represent the target load-shape regime. Prior synthetic-profile and GAN-based studies suggest that generative models can capture multiscale temporal demand patterns when the training representation is sufficiently rich [13, 14]. Compared with interpolation, the physics-guided generator would be expected to produce more plausible ramps and peak shapes in concept, especially around daily transitions. However, this article does not report numerical results and treats these expectations as hypotheses for future synthetic experiments.

Accuracy metrics: MAE, RMSE, and MAPE

Future synthetic experiments could evaluate reconstruction accuracy using mean absolute error, root mean squared error, and mean absolute percentage error by comparing generated values with hidden synthetic ground truth. Imputation studies in energy contexts commonly rely on error metrics to assess how closely reconstructed values match the original sequence [8, 21]. In this framework, MAE would describe average deviation, RMSE would emphasise larger errors, and MAPE would express relative error where load magnitude allows meaningful interpretation. These metrics would be proposed for later validation, not used to present findings in the present conceptual article.

Peak load and event-period error

Peak-load error and event-period error would be important because demand response verification depends on whether reconstructed profiles preserve operationally meaningful differences, not only average pointwise accuracy. Baseline-estimation research shows that demand response impact can be sensitive to how expected consumption is estimated during event windows [6, 20]. A reconstruction that minimises overall error but suppresses peaks could still distort peak reduction measurement. Therefore, the framework would evaluate whether generated intervals maintain plausible event-period shape, rebound behaviour, and baseline-event contrast.

Demand response verification metrics

Demand response verification metrics could include baseline deviation, event-period deviation, estimated reduction consistency, and reconstruction-flag coverage. Customer baseline and virtual control-group studies demonstrate that verification depends on comparing observed or estimated event load against a credible counterfactual [7]. The proposed framework would treat reconstructed intervals as verification-supporting evidence only when uncertainty and physical

consistency are acceptable. These proposed metrics are summarized in **Table 4**, which defines each metric and describes the expected qualitative trend without reporting experimental results.

Table 4. Proposed Evaluation Metrics for Synthetic Reconstruction Experiments

Metric	Conceptual definition	Reconstruction aspect assessed	Expected qualitative trend	Demand response relevance
Mean absolute error	Average absolute difference between hidden synthetic values and reconstructed values	General pointwise accuracy	Would be expected to remain lower for short gaps than long gaps	Indicates basic reconstruction reliability
Root mean squared error	Square-root average of squared reconstruction deviations	Sensitivity to large deviations	Would be expected to increase when peaks are missing	Helps identify costly event-period errors
Mean absolute percentage error	Relative reconstruction error compared with hidden synthetic values	Scale-normalised accuracy	Would be less stable at very low loads	Useful for customer-level comparability
Ramp-rate violation count	Frequency of reconstructed changes exceeding plausible load transitions	Physical consistency	Would be expected to decline with physics-guided penalties	Protects against implausible event-window transitions
Non-negativity violation count	Frequency of reconstructed values below zero	Basic feasibility	Should be eliminated by hard or soft constraints	Prevents physically impossible verification inputs
Daily-shape consistency score	Agreement between reconstructed profile and expected daily pattern	Temporal realism	Would improve when daily periodicity is encoded	Supports credible baseline construction
Peak-load error	Difference between true synthetic and reconstructed peak demand	Peak preservation	Would rise when missing gaps cover peak periods	Important for peak reduction programmes
Event-period deviation	Difference between reconstructed and hidden synthetic event-window demand	Verification accuracy	Would depend strongly on event-window missingness	Directly affects estimated demand response performance
Uncertainty coverage	Degree to which uncertainty flags identify difficult reconstructions	Auditability	Would be expected to increase for longer or isolated gaps	Helps decide whether verification should be accepted or qualified

Error, Privacy, and Verification Analysis

Sources of reconstruction error

Reconstruction error could arise from insufficient neighbouring observations, unrealistic synthetic-profile assumptions, unmodelled weather sensitivity, atypical customer behaviour, or missing intervals that coincide with rapid load changes. Data-quality and imputation studies show that missingness is not a uniform problem, because error depends on the pattern and operational context of the missing data [2, 9]. A physics-guided generator could reduce infeasible outputs but

would not guarantee semantic correctness when the true load behaviour is unusual. The framework therefore treats reconstruction as uncertainty-aware inference rather than exact recovery.

Effect of missing-interval length

The length of the missing interval would be expected to influence reconstruction confidence because longer gaps remove more temporal context and create more possible trajectories. Autoencoder-based and masked reconstruction methods indicate that model performance depends on how much observed sequence information remains available to guide imputation [10, 24]. Short gaps could often be reconstructed using local continuity, while long gaps would rely more heavily on synthetic priors and physical penalties. Consequently, the framework would flag long-gap reconstructions as less certain even when they appear smooth and physically feasible.

Privacy–accuracy trade-offs

The framework recognises a privacy–accuracy trade-off because highly detailed reconstruction can improve analytical usefulness while also increasing the risk that sensitive behavioural patterns are inferred. Smart-meter privacy reviews and fine-grained consumption studies show that detailed electricity traces can expose occupancy, appliance behaviour, and other personal information [4, 18, 19]. Using synthetic profiles and privacy-aware reporting would reduce direct exposure, but it could also limit the specificity of reconstructed load behaviour. The proposed design therefore favours reconstructions that are adequate for verification while avoiding unnecessary individual behavioural detail.

Implications for demand response verification

For demand response verification, the main implication is that reconstructed intervals should be useful, auditable, and clearly distinguished from observed data. Baseline and aggregated demand response studies demonstrate that verification depends on credible counterfactual estimation and careful event-period comparison [6, 20]. A physics-guided generative reconstruction could support verification by preserving baseline shape, avoiding physically impossible event values, and quantifying uncertainty where missingness is severe. However, the framework would recommend that programme rules specify how reconstructed values may be used in settlement, reporting, and dispute resolution.

Figure 2 maps how the proposed framework aligns reconstruction fidelity, physical feasibility, grid-edge privacy protection, and demand response verification requirements while avoiding the overinterpretation of generated smart-meter values as directly observed consumption data.

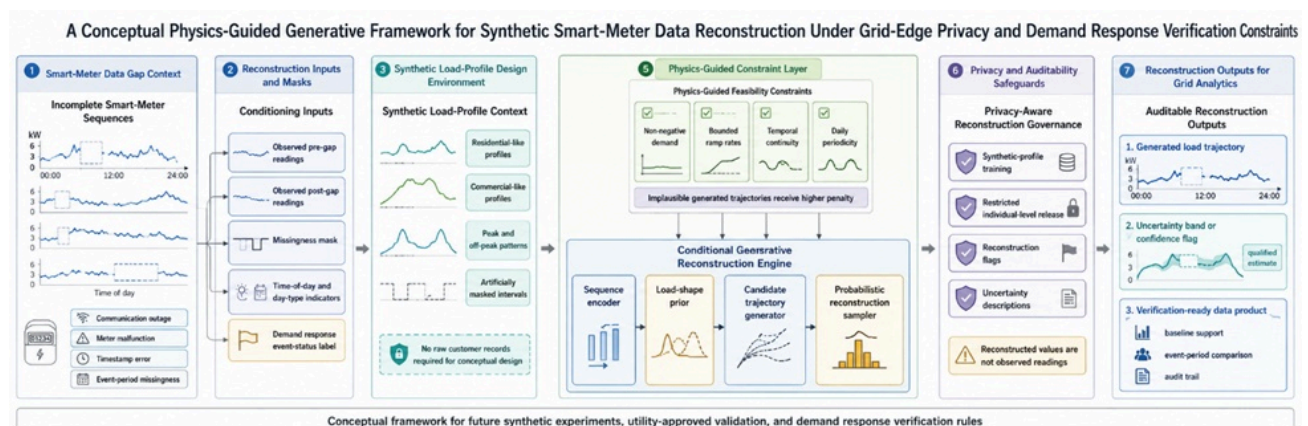


Figure 2. Privacy–Physics–Verification Alignment Map for Synthetic Smart-Meter Reconstruction

Conclusion

This article proposed a conceptual physics-guided generative framework for reconstructing missing smart-meter data using synthetic load profiles. The framework is designed to address the combined challenges of missing data, privacy protection, and demand response verification without relying on real customer records. Its core contribution is the integration of generative reconstruction with physical and operational constraints. This makes the approach suitable for conceptual development in privacy-sensitive grid-edge environments.

The framework emphasises that reconstructed smart-meter values should be plausible, auditable, and clearly distinguished from directly observed measurements. Non-negative demand, ramp-rate limits, temporal continuity, and daily periodicity provide practical forms of physics guidance. These constraints are intended to reduce unrealistic generated profiles while preserving the flexibility of modern generative models. The approach therefore seeks a balance between statistical realism and operational feasibility.

A central feature of the article is its reliance on synthetic load profiles as the primary design environment. Synthetic profiles allow missingness scenarios, benchmark comparisons, and verification logic to be specified without exposing confidential consumer data. This privacy-first orientation is especially important when reconstruction could otherwise complete behavioural traces and increase inference risks. Future work could extend the framework using public datasets, utility-approved synthetic datasets, or controlled validation environments.

The proposed framework remains conceptual and does not claim empirical superiority over existing methods. Its value lies in clarifying how physics-guided generation, privacy-aware design, and demand response verification could be integrated into a coherent reconstruction methodology. Subsequent research could formalise the architecture, test synthetic missingness scenarios, evaluate privacy safeguards, and compare reconstruction performance under approved data-governance conditions. Such work would help determine how privacy-preserving reconstruction can support trustworthy grid-edge analytics.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 18 Sep 2025

Revised: 25 Oct 2025

Accepted: 03 Dec 2025

Published online: 20 March 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Wang Y, Chen Q, Hong T, Kang C. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Trans smart Grid*. 2018;10(3):3125-48.
- Chen W, Zhou K, Yang S, Wu C. Data quality of electricity consumption data in a smart grid environment. *Renew Sustain Energy Rev*. 2017;75:98-105.
- Ryu S, Kim M, Kim H. Denoising autoencoder-based missing value imputation for smart meters. *IEEE Access*. 2020;8:40656-66.
- Razavi R, Gharipour A, Fleury M, Akpan IJ. Occupancy detection of residential buildings using smart meter data: A large-scale study. *Energy Build*. 2019;183:195-208.
- Eibl G, Engel D. Differential privacy for real smart metering data. *Comput Sci Res Dev*. 2017;32(1):173-82.
- Valentini O, Andreadou N, Bertoldi P, Lucas A, Saviuc I, Kotsakis E. Demand response impact evaluation: A review of methods for estimating the customer baseline load. *Energies*. 2022;15(14):5259.
- Lee E, Lee K, Lee H, Kim E, Rhee W. Defining virtual control group to improve customer baseline load calculation of residential demand response. *Appl Energy*. 2019;250:946-58.
- Hemanth GR, Raja SC. Proposing suitable data imputation methods by adopting a Stage wise approach for various classes of smart meters missing data—Practical approach. *Expert Syst Appl*. 2022;187:115911.
- Jeong D, Park C, Ko YM. Missing data imputation using mixture factor analysis for building electric load data. *Appl Energy*. 2021;304:117655.
- Duarte O, Duarte JE, Rosero-Garcia J. Data imputation in electricity consumption profiles through shape modeling with autoencoders. *Mathematics*. 2024;12(19):3004.
- Fekri MN, Ghosh AM, Grolinger K. Generating energy data for machine learning with recurrent generative adversarial networks. *Energies*. 2019;13(1):130.
- Yilmaz B, Korn R. Synthetic demand data generation for individual electricity consumers: Generative Adversarial Networks (GANs). *Energy and AI*. 2022;9:100161.
- Claeys R, Cleenwerck R, Knockaert J, Desmet J. Stochastic generation of residential load profiles with realistic variability based on wavelet-decomposed smart meter data. *Appl Energy*. 2023;350:121750.
- Claeys R, Cleenwerck R, Knockaert J, Desmet J. Capturing multiscale temporal dynamics in synthetic residential load profiles through Generative Adversarial Networks (GANs). *Appl Energy*. 2024;360:122831.
- Dobrovol'skij D, Stark HG. Synthetic demand data generation for individual electricity consumers: Inpainting. *Energy AI*. 2024;15:100312.
- Dalla Maria E, Secchi M, Macii D. A flexible top-down data-driven stochastic model for synthetic load profiles generation. *Energies*. 2021;15(1):269.

Kaneva T, Valova I, Gabrovska-Evstatieva K, Evstatiev B. A data-driven approach for generating synthetic load profiles with gans. *Appl Sci*. 2025;15(14):7835.

Kua J, Hossain MB, Natgunanathan I, Xiang Y. Privacy preservation in smart meters: current status, challenges and future directions. *Sensors*. 2023;23(7):3697.

Voyez A, Allard T, Avoine G, Cauchois P, Fromont E, Simonin M. The privacy cost of fine-grained electrical consumption data. *Sci Rep*. 2025;15(1):17391.

Tao P, Xu F, Dong Z, Zhang C, Peng X, Zhao J, et al. Graph convolutional network-based aggregated demand response baseline load estimation. *Energy*. 2022;251:123847.

Pazhoohesh M, Allahham A, Das R, Walker S. Investigating the impact of missing data imputation techniques on battery energy management system. *IET Smart Grid*. 2021;4(2):162-75.

Kezunovic M, Pinson P, Obradovic Z, Grijalva S, Hong T, Bessa R. Big data analytics for future electricity grids. *Electr Power Syst Res*. 2020;189:106788.

Chen S, Wen H, Wu J, Lei W, Hou W, Liu W, et al. Internet of things based smart grids supported by intelligent edge computing. *IEEE access*. 2019;7:74089-102.

Jeong J, Ku TY, Park WK. Denoising masked autoencoder-based missing imputation within constrained environments for electric load data. *Energies*. 2023;16(24):7933.

Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*. 2019;378:686-707.

Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys*. 2021;3(6):422-40.

Jia X, Willard J, Karpatne A, Read JS, Zwart JA, Steinbach M, et al. Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles. *ACM IMS Trans Data Sci*. 2021;2(3):1-26.