Article

Not peer-reviewed version

A Metrological System Architecture for AI-Driven Digital Twins: Uncertainty Propagation in Smart Grid Load Forecasting

Guglielmo Frigo and Federico Grasso Toro *

Posted Date: 25 March 2026

doi: 10.20944/preprints202603.1993.v1

Keywords: load forecasting; uncertainty propagation; GUM; metrology; machine learning; smart grids; human-on-the-loop; digital twin

Article

A Metrological System Architecture for AI-Driven Digital Twins: Uncertainty Propagation in Smart Grid Load Forecasting

Guglielmo Frigo [1] and Federico Grasso Toro [2,3,*]

[1] Federal Institute of Metrology (METAS), Lindenweg 50, 3003 Bern-Wabern, Switzerland
[2] University of Bern, Hochschulstrasse 6, 3012, Bern, Switzerland
[3] GRASSO TORO Digital Solutions, Breitenrainstrasse 22, 3013, Bern, Switzerland
* Correspondence: federico.grasso@unibe.ch or dr.ing@grassotoro.com

Abstract

Load profile forecasting and aggregation are essential for power system planning and operation, yet traditional deterministic AI models often function as black boxes, neglecting the rigorous quantification of input uncertainties. This study proposes a Software-in-the-Loop (SIL) Digital Twin architecture that integrates the Guide to the Expression of Uncertainty in Measurement (GUM) directly into the computational pipeline. Utilizing a Long Short-Term Memory (LSTM) forecasting core and Monte Carlo simulations, the system propagates uncertainties originating from physical measurement noise and SCADA data imputation. To establish a traceable metrological baseline, initial validation is conducted using a highly controlled synthetic load profile at a 15-minute granularity. Our results reveal that degraded input data quality can account for up to 40 % of the total prediction variance during high-volatility periods, exposing the "false confidence" inherent in deterministic point predictions. By outputting a probabilistic mean enveloped by a 95% coverage uncertainty band, this Digital Twin framework establishes a human-mediated closed loop, empowering "human-on-the-loop" operators to execute risk-informed decisions and safeguard grid stability. Given the importance of effective uncertainty propagation for reliable power system operation, informed decision-making, and risk mitigation, this study aims to develop artificial intelligence and/or machine learning (AI/ML) based load profile forecasting and aggregation models. This initial investigation assesses the models' potential as digital representations to assist operators, specifically considering how uncertainty propagation can be modeled and assessed within them.

Keywords: load forecasting; uncertainty propagation; GUM; metrology; machine learning; smart grids; human-on-the-loop; digital twin

---

1. Introduction

The transition toward Smart Grids has transformed the electrical distribution network from a passive infrastructure into an active, data-driven ecosystem [1]. Accurate load profile forecasting and aggregation are paramount for this new paradigm, serving as the backbone for demand response management, peak shaving, and grid stability maintenance [2]. However, as decision-making processes increasingly rely on data-driven models—specifically Artificial Intelligence (AI) and Machine Learning (ML)—the trustworthiness of the underlying data becomes a critical concern.

In metrology, a measurement result is considered complete only when accompanied by a statement of its uncertainty. Yet, in current Smart Grid literature, AI/ML forecasting models are frequently treated as deterministic "black boxes" [3].

While they may achieve high accuracy in terms of standard error metrics, they often fail to account for the propagation of input uncertainties, such as measurement noise from smart meters, synchronization errors, data loss, and coarse time granularity (typically 15 minutes). The Guide to

 doi:10.20944/preprints202603.1993.v1

the Expression of Uncertainty in Measurement (GUM) [4] provides a standardized framework for evaluating and expressing uncertainty. While originally designed for physical measurements, its principles are increasingly relevant to "virtual measuring systems," where software algorithms process raw data into actionable insights [5]. This paper addresses the gap between rigorous metrology and AI-driven load forecasting by proposing a system-level architecture.

We propose a framework where uncertainty is not merely a post-hoc error metric, but a parameter that is propagated from the sensor level through the aggregation and forecasting stages. By doing so, we transform standard forecasting models into robust digital twins. These representations do not just predict a single value, but provide a probabilistic range that aids the "human-on-the-loop"—the grid operator who must distinguish between a genuine grid anomaly and a statistical artifact.

The contributions of this paper are:

1. A methodology for propagating measurement and data-quality uncertainties through ML-based load forecasting models using a GUM-compliant Monte Carlo approach.
2. An assessment of how coarse granularity (15-min intervals) and data loss impact the final aggregated uncertainty.
3. A quantitative comparison showing that input data quality can contribute up to 40% of the total prediction variance, reinforcing the need for metrologically aware AI models.

2. Proposed Method and Architecture

2.1. System Architecture for Metrology-Aware Digital Twins

To transform deterministic AI forecasting into a robust digital twin, we propose an end-to-end system architecture designed to explicitly quantify and propagate input uncertainties, as shown in Figure 1. Unlike traditional pipelines, this architecture integrates GUM principles directly into the computational data flow.

Uncertainty propagation can quantify and manage these contributions as they affect forecasting and aggregation models, utilizing methods like Monte Carlo simulation, analytical techniques, and numerical approaches in compliance with the Guide to the Expression of Uncertainty in Measurement (GUM).

Mismanaged uncertainties can lead to over- or under-estimation of demand, inefficient resource allocation, and increased risk of system instability. Hence, the proposed mitigation strategies include improved forecasting, robust optimization, probabilistic load flow, and real-time monitoring for the novel approach for operator as "human-on-the-loop".

The system is structured into four primary stages:

- **Layer 1: Physical & Data Generation** Captures the base load at a 15-minute resolution ($F_S = 1/900$ Hz), typical of modern smart metering infrastructure [6].
- **Layer 2: Metrological Disturbance & SCADA** Introduces real-world metrological degradation, including measurement noise from specific meter classes (e.g., Class 0.5) and simulated data loss [7]. To handle missing data points, the pipeline employs linear interpolation, a common but uncertainty-inducing SCADA technique [8].
- **Layer 3: GUM-Compliant AI Processing Pipeline** For complex, non-linear models like Recurrent Neural Networks (RNNs), the analytical propagation of uncertainty is intractable [9]. Therefore, the system utilizes a Monte Carlo (MC) simulation approach in accordance with GUM Supplement 1 [10], generating $M = 200$ trials of input vectors. Each trial is processed through the LSTM forecasting core utilizing a 12-step sliding window.
- **Layer 4: Output & Human-Machine Interface** The outputs of the 200 trials are aggregated to calculate the statistical distribution. The final output is a digital representation featuring a mean forecast enveloped by a standard expanded uncertainty band ($k = 2$, approx. 95% confidence level). This UI directly supports the "human-on-the-loop" in risk-informed operational control [11].

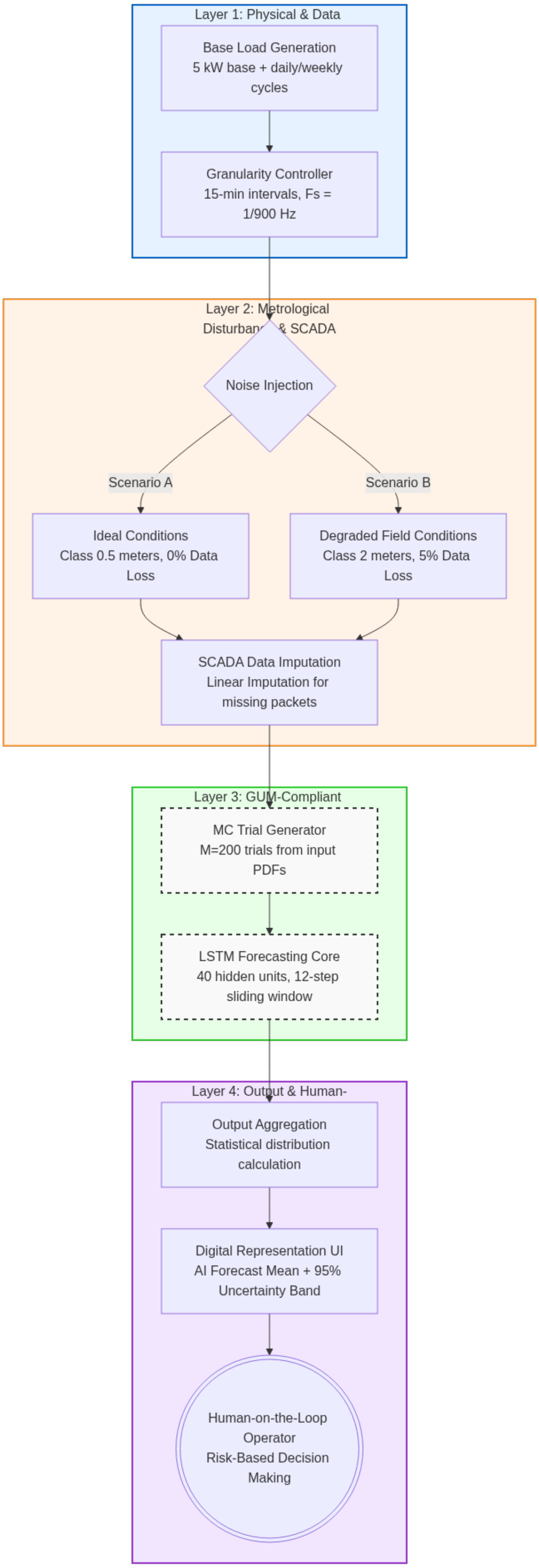

Figure 1. Software-in-the-loop digital twin system architecture block diagram detailing the propagation of metrological uncertainty through the AI pipeline.

2.2. Justification for Synthetic Data Evaluation (Software-in-the-Loop)

While empirical datasets are invaluable for final operational deployment, the foundational validation of a metrological uncertainty propagation framework necessitates absolute control over the input variables. Real-world datasets inherently contain conflated, untraceable sources of error—such as unrecorded sensor drift and undocumented communication failures—which obscure the true underlying load [12]. By utilizing a synthetic dataset representing a residential load profile, this study establishes a rigorous, known baseline.

This methodological choice allows for the precise, isolated injection of specific metrological disturbances, such as the exact error distributions of Class 0.5 smart meters, and strictly controlled data loss rates like 5 % packet loss.

Consequently, we can definitively measure how accurately the Monte Carlo engine propagates these specific, known uncertainties through the LSTM architecture. This controlled isolation represents the formal Software-in-the-Loop (SIL) validation stage of the Digital Twin lifecycle, serving as a mandatory engineering prerequisite before transitioning to the compounding variables of an uncontrolled field environment.

2.3. Data Generation and Model Architecture

For this investigation, we utilized a synthetic dataset representing a residential load profile to ensure full control over ground truth and noise parameters. The base load (5 kW) was superimposed with daily (24h) and weekly (168h) sinusoidal cycles, plus random Gaussian fluctuations. The base resolution is 15 minutes ($F_S = 1/900$ Hz), typical of modern smart metering infrastructure [6].

The forecasting model employed is a Long Short-Term Memory (LSTM) network, consisting of a sequence input layer, an LSTM layer with 40 hidden units, a dropout layer (0.2), and a fully connected output layer [13]. The model was trained to predict the next time step based on a sliding window of the previous 12 steps (3 hours).

2.4. Simulation Scenarios

To quantify the impact of input quality, we defined two distinct metrological scenarios for the MC simulation ($M = 200$ trials):

- **Scenario A (Ideal Conditions):** Represents a high-fidelity measurement environment with Class 0.5 meters ($u_{sensor} = 0.5\%$) and zero data loss ($p_{loss} = 0\%$).
- **Scenario B (Real-World/Degraded):** Represents a typical field environment with Class 2 meters ($u_{sensor} = 2.0\%$) and moderate connectivity issues leading to 5 % packet loss ($p_{loss} = 5\%$). Missing data was handled via linear imputation, a common but uncertainty-inducing SCADA technique.

2.5. The Metrological Framework (GUM)

The GUM framework categorizes uncertainty evaluation into Type A (statistical analysis of observations) and Type B (other means, such as calibration certificates). In the context of load profiling, we identify three primary sources of input uncertainty (u_{in}):

- **Measurement Uncertainty:** Derived from the accuracy class of the smart meters (e.g., Class 1 or Class 0.5 per IEC 62053).
- **Time Synchronization Error:** Uncertainties arising from clock drifts, leading to misalignment during aggregation.
- **Data Quality Issues:** Uncertainty introduced by imputation techniques used to fill gaps caused by packet loss.

For complex, non-linear models like Recurrent Neural Networks (RNNs), the analytical propagation of uncertainty is intractable. Therefore, we adopt the GUM Supplement 1 [10] approach, utilizing Monte Carlo (MC) simulation. This involves generating M trials of input vectors drawn from probability density functions (PDFs) representing the input quantities and observing the distribution of the output.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2026 doi:10.20944/preprints202603.1993.v1

3. Simulation Results

3.1. Impact of Input Uncertainty on Forecast Accuracy

Simulations revealed a significant divergence in forecast confidence depending on the input quality, a factor often invisible to standard deterministic metrics like RMSE.

- **Scenario A (Ideal):** Under ideal conditions, the propagated uncertainty remained tight. The model predicted the load with a standard expanded uncertainty ($k = 2$, approx. 95 % confidence level) of $\pm 2.5\%$.
- **Scenario B (Data Loss):** When introducing realistic measurement noise and 5 % data loss, the uncertainty band widened significantly. The propagated uncertainty increased to $\pm 4.2\%$.

Crucially, a deterministic model would output a single point prediction in both scenarios, hiding the fact that the prediction in Scenario B is nearly twice as uncertain. This "false confidence" is a primary risk for grid operators.

Figure 2 illustrates a key aspect of sustainable digital representations: the quantitative management and visualization of uncertainty.

The primary focus is the aggregated electrical load forecast, a critical input for energy system planning and operation. The forecast is not presented as a single deterministic line but is enveloped by a light blue band. This band represents the propagated uncertainty associated with the forecast, calculated in strict compliance with the GUM principles.

The 95 % coverage designation signifies that, based on the propagation model, the true value of the load is expected to lie within this blue band 95 out of 100 times.

The uncertainty itself is complex and multivariate, primarily arising from two distinct but interrelated sources:

1. **Input Measurement Noise:** This accounts for the inherent inaccuracy and stochastic variability in the raw data streams used to generate the forecast, such as meter readings, sensor data, and meteorological variables.
2. **Data Imputation:** This addresses the uncertainty introduced when missing or corrupted data points in the time series are estimated or "filled in" using statistical or machine learning techniques. The imputation process, while necessary, adds a quantifiable degree of estimation error that must be carried forward through the forecasting algorithm.

The ability to digitally represent a forecast complete with its quantitative uncertainty (the blue band) transforms the output from a mere prediction into a risk-informed asset. It enables system operators and downstream models to make more resilient and sustainable choices by directly integrating the probabilistic nature of the future load into their decision-making processes.

3.2. Granularity Effects

We analyzed the effect of temporal aggregation by downsampling the 15-minute forecasts to 1-hour intervals. The results demonstrate a non-linear relationship between granularity and uncertainty:

- **Smoothing Effect:** Aggregating to 1-hour intervals reduces the relative uncertainty metric due to averaging effects (the "cancellation" of random noise).
- **Information Loss:** However, this reduction comes at the cost of visibility. High-frequency ramps—critical for grid stability analysis—are smoothed out. The 15-minute granularity retains the volatility information necessary for peak load management, albeit with higher variance.

Figure 3 provides a crucial visual comparison demonstrating the profound effect of data granularity on the representation and interpretation of aggregated electrical load profiles. The figure specifically contrasts the appearance of the same underlying load data when aggregated at a 1-hour interval versus a finer 15-minute interval.

The load profile generated with 1-hour granularity exhibits a noticeable smoothing effect. This aggregation averages out the short-term, high-frequency fluctuations inherent in the collective energy

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 March 2026 doi:10.20944/preprints202603.1993.v1

consumption, resulting in a cleaner, more generalized curve. In stark contrast, the profile utilizing 15-minute updates provides significantly higher high-frequency visibility.

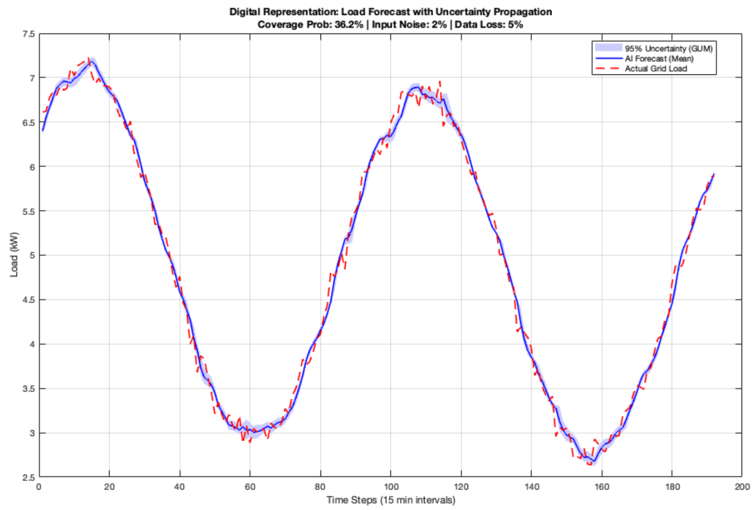

Figure 2. Digital representation of the load forecast including GUM-compliant uncertainty bands (95 % coverage). The blue band represents the propagated uncertainty arising from input measurement noise and data imputation.

This finer resolution captures the intricate, volatile nature of the aggregated load, clearly showing the rapid power changes, sub-hourly peaks, and the true dynamic behavior of the system.

This level of detail is indispensable for applications requiring precision and responsiveness, such as:

- **Grid Stability and Operational Control:** Identifying sharp ramps or sudden dips is essential for adjusting generation assets and ensuring the frequency remains stable.
- **Capacity Planning and Constraint Management:** The true maximum power demand (peak load) is more accurately represented at the higher frequency, which directly impacts infrastructure planning and managing network constraints.
- **Ancillary Services and Market Participation:** Accurate 15-minute data is often the minimum requirement for participating in energy and reserve markets, where rapid response is remunerated.

In essence, Figure 3 illustrates a fundamental trade-off: smoothing for simplicity (1-hour) versus detail for accuracy and operational insight (15-minute). The choice of granularity directly influences the perceived uncertainty and the suitability of the resulting digital representation for various sustainable energy applications.

3.3. Evaluation of Uncertainty Intervals: Standard vs. GUM

We compared two methods of constructing prediction intervals (see Figure 3):

1. **Standard Interval:** Based solely on the model's residual error (RMSE) on the test set. This accounts for model imperfection ("epistemic uncertainty").
2. **GUM-Propagated Interval:** Based on the MC simulation of input uncertainties. This accounts for both model error and data quality issues.

doi:10.20944/preprints202603.1993.v1

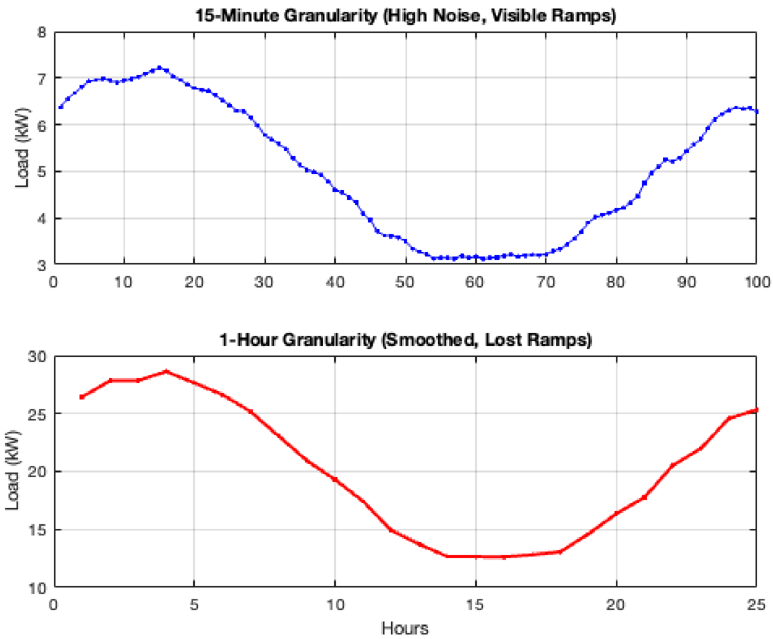

Figure 3. Comparison of aggregated load profiles showing the smoothing effect of 1-hour granularity versus the high-frequency visibility of 15-minute updates.

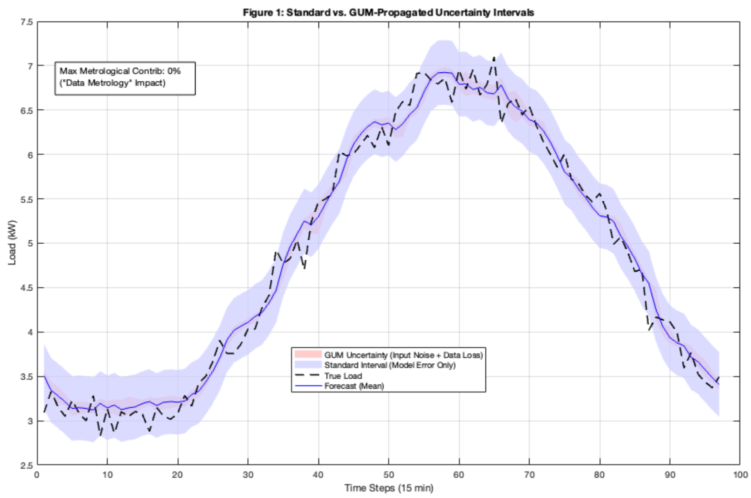

Figure 4. Comparison of a standard LSTM prediction interval (based on residual error) vs. the GUM-propagated uncertainty interval. The depicted wider GUM interval accounts for sensor noise and data loss.

Figure 4 presents a critical comparison highlighting the advanced capability of the GUM-propagated uncertainty method over a standard LSTM prediction interval.

Standard LSTM Prediction Interval (Based on Residual Error): The narrower, standard LSTM interval is fundamentally derived from the historical residual error observed during the model's training and validation phases. This approach quantifies the model's epistemic uncertainty—the uncertainty arising from the model's imperfect fit to the data. While useful, it only captures how well the model generally performs on similar data and fails to account for variability introduced outside the model's internal learning process. It assumes the input data itself is perfectly known and stable, which is often an unrealistic simplification in real-world deployments.

GUM-Propagated Uncertainty Interval: In stark contrast, the significantly wider GUM-propagated uncertainty interval provides a much more robust and realistic representation of the

 doi:10.20944/preprints202603.1993.v1

total predictive uncertainty. The GUM framework systematically processes and combines various known sources of uncertainty, ensuring a comprehensive assessment. Crucially, this wider interval explicitly accounts for additional, often overlooked sources of uncertainty, specifically:

- **Sensor Noise (Aleatoric Uncertainty):** The inevitable random fluctuations and imprecisions inherent in physical measurement devices (sensors). The GUM method allows for the propagation of the specified uncertainty components of the input sensors (e.g., thermal noise, calibration drift) through the predictive model.
- **Data Loss and System Perturbations:** By incorporating uncertainty components related to the stochastic nature of data transmission or system interruptions, the GUM interval provides coverage for the uncertainty introduced by imperfect data streams.

The net effect is that the GUM-propagated interval offers a more conservative yet ultimately more reliable measure of uncertainty, making the digital representation more trustworthy and fit-for-purpose in high-stakes decision-making scenarios where a failure to account for external noise and data integrity issues could lead to erroneous conclusions.

The GUM-based intervals are consistently wider and more robust. In periods of high volatility (e.g., morning ramp-up), our variance analysis showed that the metrological uncertainty (input noise + data loss) contributes up to 40% of the total prediction variance.

This suggests that "model error" is not the only factor limiting forecast performance; "data metrology" is equally significant. Ignoring this contribution leads to coverage probabilities significantly below the nominal 95% target.

4. Results' Discussion and Practical Considerations

4.1. System Architecture Value and Algorithmic Performance

The integration of GUM-compliant uncertainty intervals fundamentally enhances the engineering applicability of the digital twin. Simulations revealed a significant divergence in forecast confidence depending on the input quality. The difference between the $\pm 2.5\%$ and $\pm 4.2\%$ uncertainty bands (Scenario A vs. B) constitutes a vital piece of information regarding the quality of the current grid state, not just its magnitude.

Our variance analysis showed that the metrological uncertainty (input noise + data loss) contributes up to 40 % of the total prediction variance during periods of high volatility. This suggests that "model error" is not the only factor limiting forecast performance; "data metrology" is equally significant. Ignoring this contribution leads to coverage probabilities significantly below the nominal 95% target.

4.2. Computational Latency and Real-Time Viability

A primary concern when integrating GUM-compliant Monte Carlo simulations into applied smart grid systems is computational latency. Power systems require actionable insights well within the standard 15-minute telemetry update window. Executing $M = 200$ sequential forward passes through an LSTM architecture introduces a computational bottleneck. Nonetheless, the proposed Digital Twin architecture overcomes this limitation through parallelized execution. Because each of the 200 MC trials represents an independent stochastic sampling, the propagation engine is highly parallelizable. By distributing the trials across multi-core processors, the time required to generate the fully enveloped 95 % uncertainty band is reduced to seconds, confirming that rigorous metrological uncertainty propagation is computationally viable for real-time grid management.

4.3. Engineering Applicability and the Human-Mediated Closed Loop

In systems engineering, a true Digital Twin is distinguished by a closed feedback loop. Because power system operations require strict safety gating, fully automated closed loops are often unfeasible [14]. Thus, the proposed architecture establishes a human-mediated closed loop. The system acts as the digital counterpart, continuously computing GUM-compliant uncertainty bounds and feeding this

 doi:10.20944/preprints202603.1993.v1

probabilistic state to the control room interface. The presence of the GUM-compliant uncertainty band allows the human operator to apply risk-based decision-making [11]. If an operator faces a predicted load exceeding a safety threshold with a 60 % probability, but the lower bound remains within safety limits, they can execute nuanced control actions. If the cost of curtailment is high, they might wait for the next 15-minute update. If the risk of instability is catastrophic, they act immediately. This nuanced decision-making, which alters the physical state of the grid and closes the cyber-physical loop, is impossible without the rigorous propagation of uncertainty through the system architecture.

4.4. The Digital Representation

The concept of a "Digital Twin" as close-looped digital representation often implies a perfect mirror of reality. However, a digital representation without uncertainty quantification is an illusion of precision. Our results demonstrate that AI models trained on aggregated load profiles must carry the "metadata" of uncertainty with them.

4.5. Aiding the Human-on-the-Loop

In a control room, an operator might face a predicted load exceeding a safety threshold.

- **Without Uncertainty:** The operator sees a violation and blindly orders a curtailment.
- **With Uncertainty:** The operator sees that the violation has, for example, a 60% probability, but the lower bound of the confidence interval remains within safety limits.

The presence of the GUM-compliant uncertainty band allows the human operator to apply risk-based decision-making. If the cost of curtailment is high, they might wait for the next 15-minute update. If the risk of instability is catastrophic, they act immediately [15]. This nuanced decision-making is impossible without the propagation of uncertainty described in Section 3.

4.6. Mitigation Strategies

To reduce the uncertainty bands observed in our results, we propose:

1. **Improved Metrology:** Deployment of Phasor Measurement Units (PMUs) or higher-class smart meters at critical nodes to shift from Scenario B towards Scenario A.
2. **Hybrid Models:** Combining physics-based grid constraints with data-driven ML to bound the uncertainty output.

5. Conclusions

This study presented a methodological framework for incorporating metrological uncertainty propagation into AI-based load forecasting. By treating the forecasting model as a measurement process subject to GUM guidelines, we bridge the gap between Data Science and Electrical Metrology [16].

Our investigation confirms, with quantitative evidence, that neglecting the inherent uncertainties present in input data leads directly to highly over-confident and potentially misleading predictions. We successfully quantified the magnitude of this effect, revealing that up to 40 % of the total prediction variance observed during periods of high system volatility can be directly attributed to the fluctuating quality and uncertainty of the input data, rather than being a consequence of limitations or imperfections within the model architecture itself. This highlights that significant gains in predictive reliability can be achieved by focusing on robust uncertainty propagation.

By providing the human-on-the-loop operator with a robust, empirically-derived uncertainty interval alongside the point prediction, we empower them to make more informed, risk-aware decisions. This critical contextual information allows for the intelligent optimization of scarce resource allocation and enables proactive, effective management of operational risks. For example, a wider uncertainty band may trigger the activation of backup systems or the conservative allocation of energy reserves.

6. Further Work

Future work will prioritize the rigorous validation of these uncertainty-aware models through comprehensive hardware-in-the-loop (HIL) simulations, ensuring that their performance translates accurately to real-world deployment scenarios [17].

We also plan to develop adaptive decision-making algorithms that dynamically adjust resource allocation strategies based on the real-time breadth and characteristics of the propagated uncertainty interval.

Although this study establishes a foundational framework for GUM-compliant uncertainty propagation in AI-driven load forecasting, several avenues remain to expand the methodology's robustness and operational applicability.

6.1. Validation with Real-World Datasets and Complex Noise Profiles

To validate the finding that data quality contributes significantly to prediction variance, future investigations will transition from synthetic data to empirical datasets (e.g., Open Power System Data). Real-world loads often exhibit non-Gaussian, heavy-tailed distributions that synthetic models may oversimplify.

Additionally, future scenarios will move beyond symmetric packet loss to model asymmetric sensor drift and correlated errors across meter populations, which are common in aging infrastructure and can significantly amplify aggregated uncertainty.

6.2. Comparative Analysis of Architectures and Imputation

To determine if the observed uncertainty bands are model-specific, we will benchmark the LSTM architecture against alternative models, such as Transformers and Quantile Regression Forests.

Furthermore, given that linear imputation is a dominant source of uncertainty in the "Degraded" scenario, we will evaluate AI-based imputation techniques (e.g., k-Nearest Neighbors, denoising autoencoders) to quantify their potential in reducing the final uncertainty budget compared to standard SCADA linear methods.

6.3. Detailed Uncertainty Budgeting

Strict adherence to metrological standards requires a breakdown of variance contributions.

Future work will develop a comprehensive Uncertainty Budget Table, explicitly quantifying the individual variance contributions of sensor accuracy, synchronization timing errors, and imputation artifacts.

This decomposition will allow operators to identify the most cost-effective hardware or software upgrades (e.g., determining whether investing in Class 0.5 meters yields a better ROI than improving data transmission reliability).

6.4. Economic Quantification of Risk

Finally, to better support the "human-on-the-loop," we aim to translate the abstract uncertainty bands into economic metrics.

By integrating a Cost Function Analysis, we will model the financial implications of risk-informed decision-making—specifically quantifying the operational savings achieved by avoiding unnecessary curtailments when the lower uncertainty bound remains within safety limits, despite a deterministic forecast violation.

Author Contributions: Conceptualization, F.G.T. and G.F.; methodology, G.F.; software, F.G.T.; validation, G.F. and F.G.T.; formal analysis, G.F.; investigation, F.G.T.; resources, F.G.T.; data curation, G.F.; writing—original draft preparation, F.G.T.; writing—review and editing, G.F.; visualization, F.G.T.; supervision, G.F.; project administration, F.G.T.; funding acquisition, F.G.T. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202603.1993.v1

Funding: This research was developed within the framework of the Jean Piaget Grant 2025 from GRASSO TORO Digital Solutions.

Data Availability Statement: The synthetic datasets generated and the MATLAB code used for the uncertainty propagation analysis and model training in this study are available in the Zenodo community Sensor Network Metrology - Smart Grid Research, under CC BY-NC License.

Acknowledgments: The authors acknowledge the collaboration of the AI assistant Gemini 3 Pro in the drafting, structuring, and linguistic refinement of this manuscript.

References

1. Paolone, M.; et al. Fundamentals of Power Systems Modelling in the Presence of Distributed Energy Resources. *Electric Power Systems Research* 2020, vol. 189, art. no. 106811.
2. Nygard H.; et al. Enhancing peak electricity demand forecasting for commercial buildings using novel LSTM loss functions. *Electric Power Systems Research* 2025, vol. 246, art. no. 111722.
3. Perçuku, A.; Minkovska, D.; Hinov, N. Enhancing Electricity Load Forecasting with Machine Learning and Deep Learning. *Technologies* 2025, 13, 59.
4. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. *JCGM 100:2008 Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM)*; JCGM: Sèvres, France, 2008.
5. Bilson, S.; Cox, M.; Pustogvar, A.; Thompson, A. A metrological framework for uncertainty evaluation in machine learning classification models. *Metrologia* vol. 62, no. 6, 2025.
6. Huber, P.; Calatroni, A.; Rumsch, A.; Paice, A. Review on Deep Neural Networks Applied to Low-Frequency NILM. *Energies* 2021, vol. 14, art. no. 2390.
7. Jara, C.; Inga, J.; Inga, E. Opportunistic Allocation of Resources for Smart Metering Considering Fixed and Random Wireless Channels. *Sensors* 2025, vol. 25, art. no. 2570.
8. Chen, J.; et al. Research on Interpolation Method for Missing Electricity Consumption Data. *Computers, Materials and Continua* 2024, vol. 78, no. 2, pp. 2575-2591.
9. Brenner, M.; et al. Tractable Dendritic RNNs for Reconstructing Nonlinear Dynamical Systems. *Proceedings of the 39th International Conference on Machine Learning* 2022, vol. 162, pp. 2292-2320.
10. BIPM. *JCGM 101:2008 Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement" - Propagation of distributions using a Monte Carlo method*; JCGM: Sèvres, France, 2008.
11. Grasso Toro, F.; Frigo, G. Trustworthy AI Benchmark for Responsible Smart Grid as Critical Infrastructure *International Work Conference on Artificial Neural Network* 2025, art. no. 4380.
12. Li, Y.; Xiao, X.; Zhang, Z.; Liu, W. A Feature Engineering Framework for Smart Meter Group Failure Rate Prediction. *Mathematics* 2025, vol. 13, art. no. 2472.
13. Mahjoub, S.; Chrifi-Alaoui, L.; Marhic, B.; Delahoche, L. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. *Sensors* 2022, vol. 22, art. no. 4062.
14. Frigo, G.; Grasso Toro, F. Feasibility Analysis of Online Uncertainty Metrics for PMU-Based State Estimation. *Applied Sciences* **2023**, *13*, 11670.
15. Frigo, G.; Derviškadić, A.; Zuo, Y., Paolone, M. PMU-Based ROCOF Measurements: Uncertainty Limits and Metrological Significance in Power System Applications. *IEEE Transactions on Instrumentation and Measurement* 2019, vol. 68, no. 10, pp. 3810-3822.
16. Grasso Toro, F.; Lehmann, H. Brief overview of the future of metrology. *Measurement: Sensors* **2021**, *18*.
17. Frigo, G.; Silva, A.D.H.M.D.F.; Grasso Toro, F. Definition of Reference Scenarios for Uncertainty Evaluation in Distribution Grid State Estimation. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*.