



ICPEs 2025

40th INTERNATIONAL CONFERENCE ON PRODUCTION ENGINEERING - SERBIA 2025

DOI: [10.46793/ICPEs25.498T](https://doi.org/10.46793/ICPEs25.498T)



Nis, Serbia, 18 - 19th September 2025

SMART ENERGY PRODUCTION AND INDUSTRY 5.0: XAI AND IRREGULAR TIME SERIES FORECASTING

Milica TASIC¹, Ivan CIRIC¹, Predrag RAJKOVIC¹, Vladan JOVANOVIC¹, Milos SIMONOVIC¹, Marko IGNJATOVIC¹

Orcid: 0000-0002-5399-2870; Orcid: 0000-0002-2914-0985; Orcid: 0000-0003-1364-7764,

Orcid: 0000-0001-7205-7987

¹Faculty of Mechanical Engineering, University of Nis, Serbia

*Corresponding author: milica.tasic@masfak.ni.ac.rs

Abstract: Modern energy production systems, particularly in the district heating sector, face challenges caused by irregularly sampled time series resulting from asynchronous measurements, missing values, and variable operating conditions. These issues complicate the application of traditional forecasting methods. In this context, interpolation procedures play an important role, where the error depends on the curvature of the function and indicates the extent to which a polynomial of a given degree can approximate the observed signal, which is particularly important in irregular time series. In this way, a theoretical foundation is provided for understanding the limitations of data regularization and the irregularity of time series. The proposed work combines the theoretical analysis of interpolation with advanced time series processing and machine learning methods in order to support reliable forecasting, process optimization, and decision-making. The results highlight how theoretical insights into interpolation errors can guide the design of explainable and transparent forecasting models, thereby advancing smart energy production strategies aligned with Industry 5.0.

Keywords: irregular time series, polynomial interpolation error, artificial intelligence, explainable artificial intelligence, energy consumption forecasting, industry 5.0, smart energy production

1. INTRODUCTION

Energy systems in transition, particularly district heating systems, rely on heterogeneous SCADA data streams that are often irregularly sampled due to asynchronous measurements, missing readings, and variable operating conditions. Although modern machine

learning and deep learning models achieve high predictive accuracy, their lack of

transparency hinders trust and limits their practical application in critical infrastructure. These challenges further emphasize the need for transparent and explainable models, in line with the principles of Industry 5.0, which are based on human-centricity, resilience, and sustainability[1,2]. Recent reviews of XAI approaches in the energy sector highlight issues of traceability, accountability, and certifiability, offering guidelines and principles for the implementation of explainable solutions in areas such as demand and production forecasting, as well as grid

management [3-5]. In the context of irregular time series, interpolation plays a crucial role in filling gaps and regularizing data. However, the accuracy of such procedures depends on the local curvature of the function, which directly affects the interpolation error and the limits of polynomial approximation of a given order [6]. At the same time, the application of XAI techniques, particularly SHAP and LIME, has been gaining increasing attention, as they enable deeper insights into the functioning of predictive models and the importance of individual features. At the household level, research confirms that XAI contributes to the reliable identification of consumption patterns (both current and historical) and the validation of model architectures using standard performance metrics (R^2 , RMSE, MAE) [7]. Nevertheless, model explainability has no real value unless it is grounded in high-quality data and methodologically consistent approaches. Inadequate regularization or excessive interpolation may generate convincing but misleading model interpretations. For this reason, modern XAI frameworks emphasize the necessity of aligning technical solutions with organizational and societal requirements, including transparency, accountability, and security—the core principles of Industry 5.0.

Building on these challenges, this paper makes threefold contributions: (I) it links the theoretical framework of interpolation errors, derived from function curvature, with preprocessing procedures for irregular time series, (II) it integrates these insights into a modeling pipeline that seeks to preserve the authenticity of real system dynamics, and (III) it extends this pipeline with XAI analyses (SHAP/LIME) that enable the interpretation of predictions in the context of energy variables and operational decisions. In this way, the study contributes to the development of transparent and explainable predictive models that enhance trust, usability, and sustainability in line with the Industry 5.0 paradigm.

2. THEORETICAL FOUNDATIONS OF INTERPOLATION

The theoretical analysis of interpolation errors provides not only a mathematical framework but also a practical tool for processing consumption data and forecasting load demand in district heating systems. The curvature of a function, expressed through higher-order derivatives, together with the distribution of interpolation nodes and multidimensional data processing, directly determines the limits of prediction accuracy. In the context of energy systems, and particularly district heating networks, interpolation techniques are applied during the processing of SCADA data, where missing or irregularly sampled values frequently occur.

A classical result states that if

$$f \in C^{n+1}[a, b] \quad (1)$$

the interpolation polynomial $p_n(x)$ satisfies the relation:

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!} \prod_{i=1}^n (x - x_i) \quad (2)$$

which indicates that the error magnitude depends both on the higher-order derivative of the function and on the choice of interpolation nodes [8,9]. This formula is analogous to the Taylor expansion, but differs in that the polynomial is constructed from function values at several distinct points, making it particularly suitable for time series characterized by discreteness and irregular sampling. The error of interpolation is often smaller if we increase the number of nodes. But, there are some counterexamples.

2.1 Example 1. Consider the Runge function:

$$f(x) = \frac{1}{1+x^2}, \quad x \in [-5, 5]. \quad (3)$$

In contrast to our expectations, increasing the number of nodes especially in the segments $[-5, -3]$ and $[3, 5]$, has the consequence that the interpolation (blue curve) gives significant error. Even more,

$$\max_{-5 \leq x \leq 5} \|f(x) - Q_n(x)\| \rightarrow \infty, \quad n \rightarrow \infty \quad (4)$$

On the Figure 1, the divergence is obvious for $n=10$. This example illustrates that interpolation error may increase despite using more nodes. This naturally raises the question of how the choice of interpolation nodes affects the error. Until now, the discussion has been based on the assumption that the interpolation nodes are fixed.

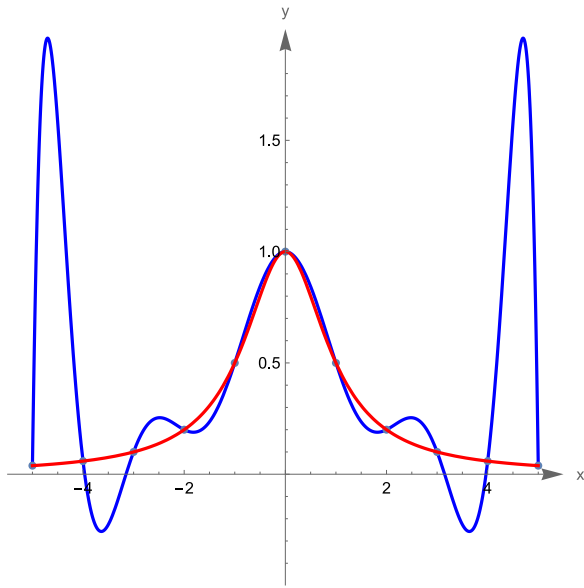


Figure 1. Illustration of Runge's phenomenon

In the entire discussion up to now, it is assumed that the interpolation nodes are given.

If this is the case, it would be reasonable to use this degree of freedom to minimize the interpolation error.

The distribution of error strongly depends on the selection of interpolation nodes. For uniformly distributed nodes, oscillations and significant growth of error occur near the interval boundaries, a phenomenon known as Runge's phenomenon. In energy applications, particularly in heat demand analysis, such an effect may lead to inaccurate forecasts during periods of sudden outdoor temperature changes. Therefore, special node distributions, such as Chebyshev nodes, are employed to provide a more balanced error distribution and to minimize the maximum deviation of the interpolation polynomial from the actual data [10-12].

Apart from node selection, the nature of the underlying function also plays an important role. Rapid growth of higher-order derivatives (e.g., sudden changes in heat load) can significantly increase the interpolation error, demonstrating that increasing the number of interpolation points does not necessarily improve the approximation. This demonstrates that increasing the number of interpolation points does not necessarily improve the approximation and that, in real district heating operations, a compromise must be made between complexity and stability [12].

In such circumstances, the use of spline functions becomes particularly relevant, as they enable stable interpolation through local low-degree polynomials. Unlike global polynomials, splines ensure smoothness of the function and continuity of derivatives, which is especially important in energy systems where changes in load and temperature must remain physically consistent. In this way, the risk of oscillations and error accumulation is reduced, and the reconstructed data provide a more reliable basis for modeling heat substations [13,14].

On some real interval $[a, b]$ notice its partition $\mathbb{P}$ with nodes

$$\mathbb{P}: a = x_0 < x_1 < \dots < x_{n-1} < x_n = b, \quad (n \in \mathbb{N})$$

Suppose that for a function $f(x)$ are known values $f_k = f(x_k)$ ($k = 0, 1, \dots, n$)

The function $S_m(x) = S_{m,k}(x; f; \mathbb{P})$ is polynomial spline with degree m and defect k ($1 \leq k \leq m$) for $f(x)$ if

- 1) For every $i \in \{1, 2, \dots, n\}$, and for every $x \in (x_{i-1}, x_i)$, $S_m(x)$ is a polynomial with degree at most m ;
- 2) $S_m(x)$ is a part-by-part interpolating polynomial, i.e. $S_m(x_i) = f_i$ ($i = 0, 1, \dots, n$);
- 3) $S_m(x)$ is a continuously differentiable function with its derivatives until order $m - k$ on $[a, b]$, i.e.

$$S_m^{(k)}(x_i^-) = S_m^{(k)}(x_i^+) \\ (k = 1, \dots, m - k; \quad i = 1, \dots, n - 1).$$

Figure 2 illustrates the concept of a spline as a piecewise polynomial that maintains smoothness across interval boundaries.

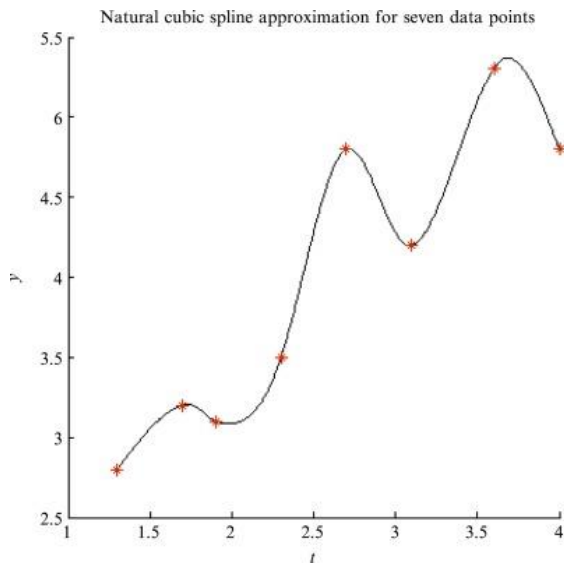


Figure 2. Natural cubic spline interpolation

These insights form the foundation for subsequent methods in this work that combine interpolation with machine learning and explainable artificial intelligence (XAI) to enhance transparency and reliability in district heating operations.

3. RELATED WORK

3.1 Data characteristics and preprocessing of irregular time series

Research focusing on data processing in district heating systems highlights that SCADA and smart meters typically provide key variables such as supply and return water temperatures, thermal energy, flow, and outdoor temperature. For example, Cvetković et al. document a publicly available SCADA dataset in which measurements were collected at the substation level and standardized to hourly time steps after cleaning, resulting in time series without errors or missing values [15]. Similarly, Schaffer et al. present a three-year dataset from 3,021 smart meters, where raw readings deviated by up to ± 30 minutes from the full hour and were therefore subjected to initial screening, linear interpolation, and

imputation to obtain a fully equidistant hourly series [16].

Typical challenges include irregularity and missing values. In practice, timestamps are not always aligned due to communication delays or technical faults, which leads to gaps in the series. Chatterjee et al. emphasize that raw smart meter data are often irregularly spaced and therefore require interpolation and resampling before being used in predictive models [17]. In addition, Steindl et al. point out that the transition from daylight saving time (DST) to standard time can produce incorrect measurement order, so it is recommended to explicitly remove readings recorded between 02:00–02:59 during this transition [18].

The proposed preprocessing procedures form a clear pipeline. In the first stage, data are checked for inconsistencies, especially cumulative variables such as energy or flow, where monotonicity is required (each new value must be greater than or equal to the previous one). A similar approach was applied in studies on fault detection based on smart meter data [19]. Interpolation and resampling are then performed—most often linear—in order to form an equidistant time grid (hourly or 15-minute). Short gaps up to one hour are typically filled by linear interpolation, while longer gaps are usually excluded from analysis [18]. Afterward, the remaining small fraction of missing values is imputed using moving-window methods. Schaffer et al. conducted a systematic comparison of different techniques and demonstrated that a weighted moving average with a 48-hour symmetric window, combined with scaling that respects the cumulative trend, yields the best results for smart meters [16]. Finally, when measurements are quantized or truncated at the hourly level (e.g., due to utility data storage practices), aggregation to daily values is recommended to reduce rounding errors and provide a more stable series for anomaly detection [20].

Beyond classical procedures, the importance of documenting the entire process through formal data-curation

guidelines is emphasized. Stecher et al. recommend an approach inspired by the CRISP-DM methodology, adapted to the DHS domain: defining objectives, understanding the data, integrating multiple sources (SCADA and meteorology), and performing systematic cleaning and annotation.

3.2 Forecasting heat demand in district heating systems

Conventional approaches to heat demand forecasting in district heating systems (DHS) are primarily rule-based, relying heavily on weather forecasts and, in particular, ambient temperature. This control strategy, known as weather compensation, can provide acceptable decisions in certain circumstances but rarely achieves optimality. In practice, this often results in overheating, observable through elevated supply line water temperatures and decreased heat load. While end-user comfort is usually unaffected, such inefficiencies lead to increased fuel consumption and negative environmental impacts.

Several empirical studies highlight the progress from conventional to advanced data-driven models.

Potočnik et al. [21] pioneered the use of Gaussian Process Regression (GPR) in Slovenia's largest DHS, demonstrating that nonlinear models substantially improve short-term forecasting accuracy. Building on this, Zdravković et al. [22] proposed a proactive control strategy based on deep learning, where different LSTM architectures enabled more adaptive and responsive system operation. In a complementary study, Kurek et al. [23] combined traditional approaches with fuzzy logic models, adapted to seasonal variations, to manage Warsaw's large district heating network, underlining the potential of ANNs in complex urban environments.

The scope of AI applications has been further expanded in comparative studies. Runge and Saloux [24] investigated predictive versus forecasting paradigms, showing that advanced machine learning methods,

particularly LSTM and XGBoost, yield significant accuracy improvements in Canadian DHS. Along similar lines, Shakeel et al. [25] introduced a hybrid LightGBM–FB Prophet model, demonstrating the ability to overcome sparse-data challenges and providing superior accuracy in operational heat demand forecasting. In China, Liu et al. [26] implemented a dynamic integrated control strategy, which incorporated multiple correction factors to better align supply with demand, achieving notable improvements in energy efficiency and emission reduction. Finally, Morteza et al. [27] explored Deep Recurrent Neural Networks (DRNNs) for medium- to long-term forecasting, emphasizing the importance of hyperparameter optimization. Their work confirmed that tailored deep learning architectures significantly outperform conventional models in energy demand prediction.

3.3 Explainable AI in district heating

Explainable AI in district heating has been advancing along two complementary tracks: data-centric XAI readiness—i.e., curated SCADA datasets and platforms that ensure reliable, interpretable inputs—and model-centric explainability—i.e., forecasting models whose inner workings (features, lags, temporal focus) can be understood and trusted by operators.

Recent work formalizes the data backbone needed for explainable control in DHS. An integrated data-acquisition platform aggregates SCADA, IoT, and weather streams into a unified time-series repository, enabling real-time monitoring and predictive control; the platform is positioned as the core of an explainable control approach (XAI4HEAT) that combines XAI algorithms with model-based predictive control for dynamic heat-supply adjustment. This line of work explicitly motivates explainability for operator trust, situational awareness, and decision support [28]. Building on the same program, the platform's conclusions emphasize that

curated, multi-source data ingestion is prerequisite to intelligent, explainable DHS operation [28].

Complementing the platform perspective, a curated SCADA dataset for enhanced explainability documents five years of substation-level measurements (supply/return temperatures on both primary and secondary circuits, energy transmission, local meteorology) and, crucially, standardizes the data to uniform hourly resolution without errors or missing values—a property that directly supports reliable model interpretation and feature attribution [29]. The dataset is explicitly positioned to facilitate explainable analyses in smart-energy research and DHS practice [29].

On the modeling side, Neubauer et al. develop a multi-step heating-load forecaster that marries an Encoder–Decoder architecture with SHAP feature attributions and temporal attention. They show that attention weights and SHAP values increase interpretability; feature selection guided by Deep SHAP reduces training time and error; and future weather features materially improve accuracy—findings that connect XAI directly to both performance and operator insight [30].

Going beyond post-hoc explanations, Souza et al. (HELIOS) embed explainability into the model structure by combining physics and expert knowledge. The approach decomposes heat load into space heating, domestic hot water, and transmission losses, links them through a causal graph, and leverages contextual mixture-of-experts to encode domain priors—yielding transparent predictions and superior accountability relative to black-box baselines [30]. The authors further argue that such inherent interpretability fosters stakeholder trust and supports operational decision-making [30].

Together, these strands outline a practical XAI framework for DHS: curate XAI-ready data (platform + dataset) to ensure valid, aligned inputs; adopt interpretable forecasting (SHAP + attention) to expose feature and temporal contributions; and, where possible, encode

physics/expert knowledge to make explanations intrinsic rather than purely post-hoc. This triangulation supports transparent forecasting, operator trust, and integration with predictive control in real networks [30].

4. METHODOLOGY

The data used in this study were obtained from the SCADA system of the Faculty of Mechanical Engineering in Nis, specifically from the heat substation in Lamela L8. For the purpose of this analysis, two input variables were selected: outdoor air temperature (t_{amb}) and supply water temperature in the primary circuit (t_{sup_prim}). These variables were chosen due to their direct physical influence on heat consumption and district heating system operation.

A specific characteristic of the dataset is that it is irregularly sampled, meaning that measurements were not always recorded at equal time intervals. The causes of such irregularities may be of a technical nature (interruptions in data acquisition, communication issues) but can also result from the dynamic operation of the system. In addition to irregular sampling, the dataset also contains missing values, which further complicates the analysis process. These challenges require the application of special preprocessing procedures to ensure a uniform time grid and consistent inputs for the prediction model. The focus of this study is precisely on addressing these issues, as the quality of preprocessing directly determines the reliability and accuracy of the subsequent forecasting model. By carefully regularizing the time series and handling missing values, the study emphasizes the importance of preparing real-world SCADA data for robust and interpretable predictive modeling.

4.1 Data Preprocessing

The dataset contains time-stamped measurements with irregular sampling intervals and occasional missing values. To enable numerical processing:

1. Time conversion: Original timestamps were transformed into numerical values expressed as the elapsed hours from the initial timestamp. This was calculated as the difference between serial date numbers, which allows the time variable to be treated as a continuous numerical input in the model.
2. Regularization of the time grid: The entire dataset was resampled to a 15-minute resolution to ensure comparability of observations.
3. Handling missing values: A local Lagrange interpolation method with barycentric weights was applied. The interpolation order was limited to a maximum of $n = 2$, with a fallback to lower orders if the data density was insufficient. Additionally, a guardrail mechanism was used to prevent interpolation across gaps longer than 90 minutes, where missing values were preserved as NaN.
4. Value clipping: Interpolated results were restricted within physically meaningful ranges (e.g., outdoor temperature between -40°C and 80°C) to prevent unrealistic estimates.

In this context, preprocessing is the central contribution of the research. The methodology demonstrates how irregular and incomplete SCADA data can be systematically transformed into a structured form suitable for machine learning, thereby ensuring both model accuracy and interpretability.

5. RESULTS AND DISCUSSION

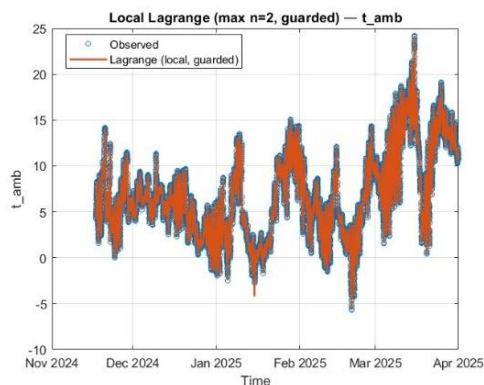


Figure 3. Observed and interpolated outdoor air temperature

The interpolation results reveal substantial differences in accuracy between the outdoor air temperature (t_{amb}) and the supply water temperature in the primary circuit ($t_{\text{sup_prim}}$).

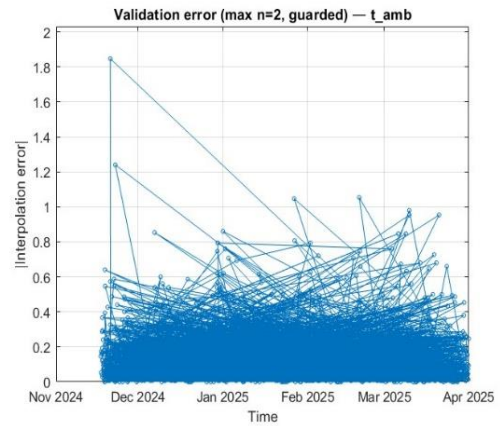


Figure 4. Validation error of local Lagrange interpolation for outdoor air temperature

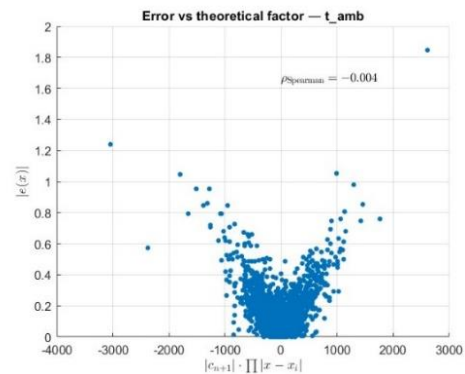


Figure 5. Interpolation error versus theoretical factor for outdoor air temperature

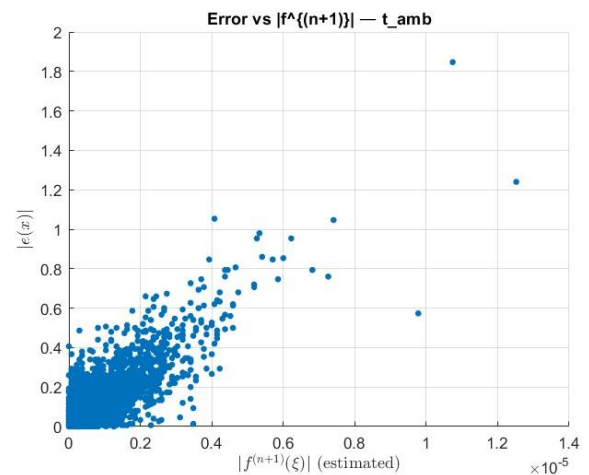


Figure 6. Interpolation error versus estimated higher-order derivative term for outdoor air temperature

For outdoor temperature, the local Lagrange interpolation ($n=2$) demonstrated very good agreement with the observed data. The interpolated curve (Fig.1) smoothly follows seasonal variations, while deviations remain minor. In most cases, the interpolation error remained below 0.5, with occasional peaks up to 1.8. The analysis with respect to the estimated curvature of the function $f^{(n+1)}$ confirms the expected positive correlation – errors increase as curvature grows, consistent with interpolation theory. This indicates that the method is sufficiently robust for processing meteorological data.

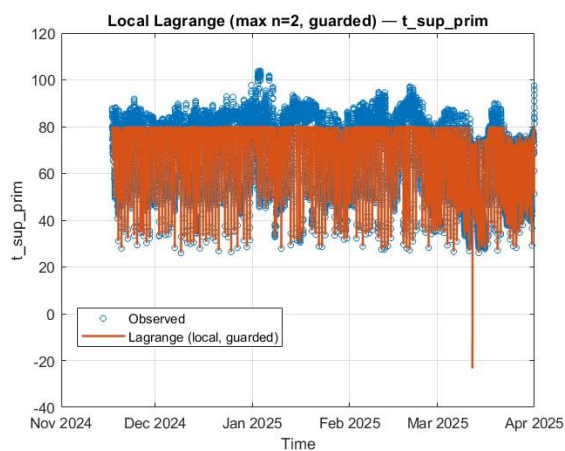


Figure 7. Observed and interpolated primary supply water temperature using local Lagrange method

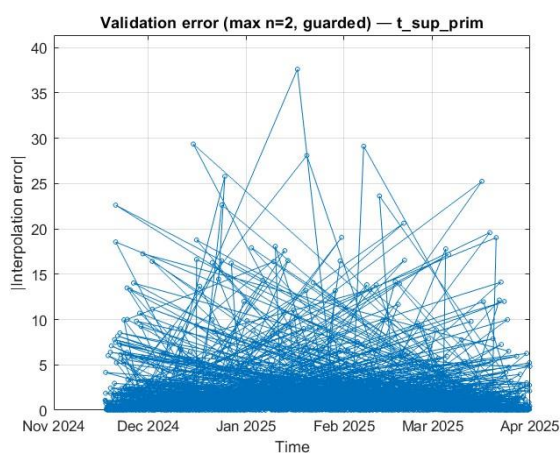


Figure 8. Validation error of local Lagrange interpolation for primary supply water temperature

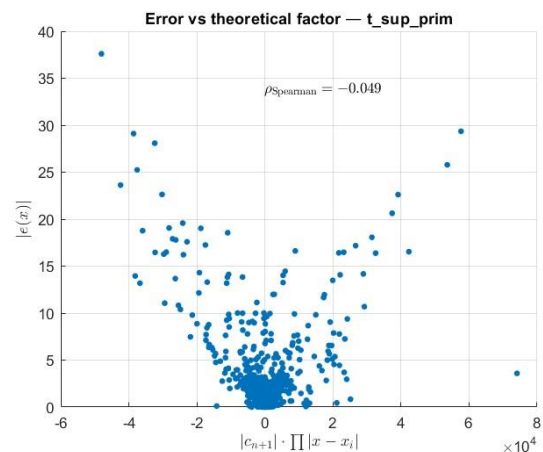


Figure 9. Interpolation error versus theoretical factor for primary supply water temperature

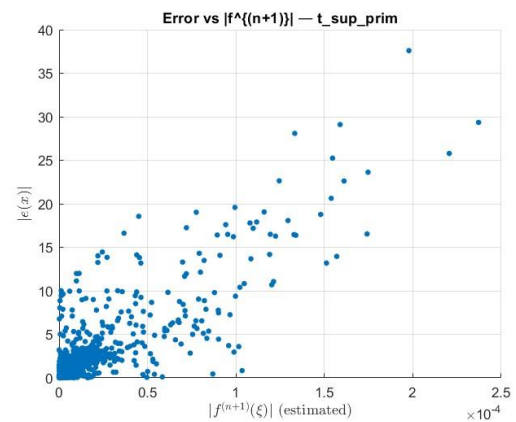


Figure 10. Interpolation error versus estimated higher-order derivative term for primary supply water temperature

In contrast, for supply water temperature in the primary circuit, interpolation accuracy is significantly lower. The results exhibit larger deviations between observed and interpolated values, as well as anomalies, including sudden drops that the method failed to capture. The interpolation error ranged widely, with values reaching up to 35–37, which is several times higher than for the outdoor temperature. The correlation with the theoretical factor proved negligible, while the dependency on the estimated curvature of the function was present but with considerable scattering.

Based on these findings, it can be concluded that the local Lagrange interpolation method is suitable and sufficiently accurate for outdoor temperature, while for the supply water temperature in the

primary circuit it requires refinement or replacement with more advanced approaches (e.g., spline interpolation, data filtering, or machine learning-based reconstruction). These results emphasize the importance of carefully selecting and combining preprocessing methods depending on the physical nature of the variables and the dynamics of the system.

6. CONCLUSION

This study demonstrated how the challenges of irregularly sampled SCADA data in district heating systems can be addressed through advanced preprocessing techniques and the application of explainable machine learning models. By focusing on outdoor air temperature and primary supply water temperature, the analysis highlighted substantial differences in interpolation accuracy, showing that some variables can be reliably reconstructed while others require more sophisticated approaches.

Beyond the specific case of district heating, the presented methodology illustrates a general framework applicable to production engineering systems, where data irregularities, missing values, and operational variability are common. The integration of mathematical interpolation theory with explainable AI provides not only accurate forecasts but also transparent insights into system behavior, which is essential for decision-making, process optimization, and sustainable operation. Future research will focus on the careful selection of prediction models and the systematic application of XAI methods fully aligned with Industry 5.0 principles, further enhancing predictive accuracy, interpretability, and operator trust in complex production systems.

REFERENCES

- [1] C. Trivedi, P. Bhattacharya, V. K. Prasad, *et al.*, "Explainable AI for Industry 5.0: Vision, Architecture, and Potential Directions," *IEEE Open Journal of Industry Applications*, vol. PP, no. 99, pp. 1–12, Sep. 2023, doi: 10.1109/OJIA.2024.3399057.
- [2] J.-L. Hu, Y. Li, and J.-C. Chew, "Industry 5.0 and Human-Centered Energy System: A Comprehensive Review with Socio-Economic Viewpoints," *Energies*, vol. 18, no. 9, p. 2345, 2025. doi: 10.3390/en18092345.
- [3] R. Alsaigh, R. Mehmood, and I. Katib, "AI Explainability and Governance in Smart Energy Systems: A Review," *arXiv preprint, arXiv:2211.00069*, Nov. 2022.
- [4] R. Machlev, L. Heistrene, M. Perl, *et al.*, "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," *Energy and AI*, vol. 9, p. 100169, May 2022, doi: 10.1016/j.egyai.2022.100169.
- [5] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, and N. Díaz-Rodríguez, "Explainable AI on Time Series Data: A Survey," *arXiv preprint, arXiv:2104.00950*, 2021.
- [6] J. Chen, J. Yuan, W. Chen, A. Zeb, M. Suzaiddola, and Y. A. Nanekharan, "Research on Interpolation Method for Missing Electricity Consumption Data," *Computers, Materials & Continua*, vol. 78, no. 2, pp. 2576–2591, Feb. 2024, doi:10.32604/cmc.2024.050414.
- [7] Bhandary, V. Dobariya, G. Yenduri, S. K. Pasupuleti, S. K. Maddikunta, and F. Benedetto, "Enhancing Household Energy Consumption Predictions Through Explainable AI Frameworks," *IEEE Access*, vol. 12, pp. 1–1, Jan. 2024, doi: 10.1109/ACCESS.2024.3373552.
- [8] Lecture Notes, *Errors in Polynomial Interpolation*, Lecture 15–16, 2010.
- [9] D. Levy, *Numerical Analysis – Interpolation*, University of Maryland, 2010.
- [10] Kværnø, *Polynomial Interpolation: Error Theory*, NTNU, 2021.
- [11] Ron, Y. Li, M. Cowlshaw, and N. Fillmore, "Polynomial Interpolation: Error Analysis and Introduction to Splines," *Lecture Notes in Numerical Analysis*, Univ. of Wisconsin, 2010.
- [12] Runge, "Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten," *Zeitschrift für Mathematik und Physik*, vol. 46, pp. 224–243, 1901.

- [13] Y. Mehari, "Easy Way to Find Multivariate Interpolation," *Int. J. Emerging Trends in Science and Technology*, vol. 4, no. 5, pp. 5189–5193, 2017.
- [14] M. Aldroubi and A. Papoulis, "Two-dimensional polynomial interpolation from nonuniform samples," *IEEE Trans. Signal Processing*, vol. 40, no. 2, pp. 333–339, 1992.
- [15] S. Cvetković, M. Zdravković, and M. Ignjatović, "Exploring district heating systems: A SCADA dataset for enhanced explainability," *Data in Brief*, vol. 59, p. 111320, 2025, doi: 10.1016/j.dib.2025.111320.
- [16] M. Schaffer, T. Tvedebrink, and A. Marszal-Pomianowska, "Three years of hourly data from 3021 smart heat meters installed in Danish residential buildings," *Scientific Data*, vol. 9, p. 420, 2022, doi: 10.1038/s41597-022-01502-3.
- [17] S. Chatterjee, P. K. Jana, and A. Bandyopadhyay, "Prediction of household-level heat consumption using PSO-kSVR with accumulated smart meter data," *arXiv preprint*, arXiv:2112.01908, Dec. 2021.
- [18] G. Steindl, M. Kathan, and G. N. Almbauer, "Comparison of black box models for load profile generation of district heating substations," *Proceedings of the 11th International Modelica Conference*, 2015, pp. 711–719, doi: 10.3384/ecp15118711.
- [19] F. Theusch, P. Diaconescu, J. Tuschy, and R. T. Stankovic, "Fault detection and condition monitoring in district heating using smart meter data," *Proceedings of the European Conference of the PHM Society*, vol. 6, no. 1, pp. 1–11, 2021, doi: 10.36001/phme.2021.v6i1.2786.
- [20] Stecher, A. A. Nielsen, M. K. Andersen, and H. Madsen, "Creating a labeled district heating data set: From anomaly detection towards fault detection," *Energy*, vol. 313, p. 134016, 2024, doi: 10.1016/j.energy.2024.134016.
- [21] Potočník, P., Škerl, P., and E. Govekar. 2021. Machine-learning-based multi-step heademand forecasting in a district heating system. *Energy & Buildings* 233:110673
- [22] Zdravković, M., Ćirić, I., and M. Ignjatović. 2022. Explainable heat demand forecasting for the novel control strategies of district heating systems. *Annual Reviews in Control* 53:405–413.
- [23] Kurek, T., Bielecki, A., Świrski, K., Wojdan, K., Guzek, M., Białek, J., Brzozowski, R., and R. Serafin. 2021. Heat demand forecasting algorithm for a Warsaw district heating network. *Energy* 217:119347.
- [24] Runge, J., and E. Saloux. 2023. A comparison of prediction and forecasting artificial intelligence models to estimate the future energy demand in a district heating system. *Energy* 269:126661.
- [25] Shakeel, A., Chong, D., and J. Wang. 2023. District heating load forecasting with a hybrid model based on LightGBM and FB-prophet. *Journal of Cleaner Production* 409:137130.
- [26] Liu, G., Zhou, X., Yan, J., and G. Yan. 2023. Dynamic integrated control for Chinese district heating system to balance the heat supply and heat demand. *Sustainable Cities and Society* 88:104286.
- [27] Morteza, A., Yahyaeian, A.A., Mirzaeibonehkhater, M., Sadeghi, S., Mohaimeni, A., and S. Taheri. 2023. Deep learning hyperparameter optimization: Application to electricity and heat demand prediction for buildings. *Energy and Buildings* 289:113036.
- [28] S. Cvetković, D. Stojiljković, I. Ćirić, R. Turudija, and M. Špeletić, "Integrated data acquisition platform for explainable control in district heating systems," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 23, no. 2, pp. 133–143, 2024.
- [29] S. Cvetković, M. Zdravković, and M. Ignjatović, "Exploring district heating systems: A SCADA dataset for enhanced explainability," *Data in Brief*, vol. 59, p. 111320, Jan. 2025.
- [30] Neubauer, S. Brandt, and M. Kriegel, "Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability," *Energy and AI*, vol. 20, p. 100480, Feb. 2025.