

UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

Personalised Comfort Metrology: Measuring the Reliability of Augmented Thermal Sensation Vote Data

This is the peer reviewd version of the followng article:

Original
Personalised Comfort Metrology: Measuring the Reliability of Augmented Thermal Sensation Vote Data / Alzahab, N.A., Kaculi, R., Casaccia, S., Serroni, S., Revel, G.M.. - (2026), pp. 277-282. (2026 IEEE International Workshop on Metrology for Living Environment, MetroLivEnv 2026 Cambridge 14 - 16 July 2026) [10.1109/MetroLivEnv70468.2026.11659858].

Availability:
This version is available at: 11566/363013 since: 2026-09-17T07:51:16Z

Publisher:
IEEE

Published
DOI:10.1109/MetroLivEnv70468.2026.11659858

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.
This item was downloaded from IRIS Università Politecnica delle Marche (https://iris.univpm.it). When citing, please refer to the published version.

Publisher copyright:

IEEE - Postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. To access the final edited and published work see 10.1109/MetroLivEnv70468.2026.11659858

(Article begins on next page)

02 November 2026

Personalised Comfort Metrology: Measuring the Reliability of Augmented Thermal Sensation Vote Data

1st Nibras Abo ALzahab
Department of Industrial Engineering and Mathematical Sciences
Marche Politicnic University
Ancona, Italy
n.abo@staff.univpm.it

2nd Richard Kaculi*
Department of Industrial Engineering and Mathematical Sciences
Marche Politicnic University
Ancona, Italy
r.kaculi@pm.univpm.it

3rd Sara Casaccia
Department of Industrial Engineering and Mathematical Sciences
Marche Politicnic University
Ancona, Italy
s.casaccia@staff.univpm.it

4th Serena Serroni
Department of Industrial Engineering and Mathematical Sciences
Marche Politicnic University
Ancona, Italy
s.serroni@pm.univpm.it

5th Gian Marco Revel
Department of Industrial Engineering and Mathematical Sciences
Marche Politicnic University
Ancona, Italy
gm.revel@staff.univpm.it

***Abstract*—Modern smart buildings use Personalized Comfort Models (PCMs) to balance energy efficiency and occupant comfort based on user feedback. However, their effectiveness is limited because most feedback data is missing (75–96%) due to occupants getting tired of frequent surveys.**

This study evaluates 16 univariate interpolation algorithms—including two ordinal-native comparators—to reliably impute missing categorical Thermal Sensation Votes (TSV) without distorting statistical distributions or environmental correlations. Using data from 10 apartments, all continuous imputations were subjected to discrete bounding and post-interpolation rounding to match valid ASHRAE integers. Performance was assessed through Pearson and Spearman correlation structure preservation and three distribution tests.

Results show that discrete rounding effectively neutralises non-physical artefacts. The linear family and the ordinal-native Ordinal Step method achieved identical lowest mean absolute correlation perturbation (0.113), while zero-order hold proved the most destructive ($1.81\times$ more accumulated error). Spearman and Pearson rankings were significantly consistent (Kendall's $\tau = 0.69$, $p<0.001$), validating the use of Pearson as a relative metric for ordinal TSV.

Ultimately, bounded and rounded linear or cubic interpolations offer a computationally lightweight and robust method to reconstruct highly sparse subjective feedback, ensuring continuous PCM training despite fluctuating occupant participation.

***Index Terms*—Thermal Sensation Vote, Personalised Comfort Models, Human-in-the-Loop, Data augmentation**

I. Introduction

Smart building automation systems are theoretically capable of operating efficiently, yet their practical deployment exposes a fundamental limitation: the scarcity of continuous, user-centric feedback data [1]. Such subjective data is essential for model robustness and personalized building behavior, but is rarely collected at scale or over long periods [2].

Thermal comfort is primarily quantified via the Thermal Sensation Vote (TSV), a standardized metric [3] on the ASHRAE [4] 7-point scale (-3 = very cold, 0 = neutral, $+3$ = very hot), capturing an occupant's real-time thermal perception [5]. With the residential sector accounting for approximately 27.4% of EU energy consumption [6], there is a growing shift from viewing thermal comfort and energy efficiency as competing objectives to jointly optimizing them [7]. This is the central goal of the European DEDALUS project, a data-driven Demand Response (DR) ecosystem leveraging digital twins and AI to enhance occupant well-being while optimizing energy use [8].

To realize this synergy, modern systems move beyond Fanger's population-average model toward Personalised Comfort Models (PCMs) [9] — ML-driven frameworks that learn individual thermal preferences from IoT sensor streams. These PCMs operate within a Human-in-the-Loop (HITL) paradigm [10], where real-time occupant feedback continuously refines the HVAC control loop [11].

In practice, TSV is collected via smartphone apps, wearables, or web dashboards [12], pairing subjective votes with concurrent environmental data (temperature, humidity) to train the PCM [13].

Despite these advantages, HITL systems face a critical data acquisition bottleneck: survey fatigue leads occupants to ignore requests or provide careless responses, while social desirability bias and perceived lack of HVAC responsiveness further reduce participation [14]. The result is highly fragmented TSV datasets with 75–96% missing values [14], heavily skewed

toward "neutral" (0) votes as occupants physically adapt rather than report discomfort. This sparsity fundamentally degrades PCM accuracy, creating an urgent need to reliably impute missing TSV observations.

A. State of the Art

To train supervised machine learning models on such datasets, the missing TSV values must be imputed. While sophisticated multivariate imputation techniques exist [15], [16], univariate interpolation along the temporal axis is widely used in practice owing to its simplicity, computational efficiency, and minimal assumptions about the data-generating process [17]. Despite its prevalence, the sensitivity of downstream model performance to the choice of interpolation method has received limited systematic attention [18]. Different interpolation strategies impose distinct smoothness constraints and may introduce distributional artefacts [19], particularly when applied to sparse, irregularly sampled subjective data [20]. Furthermore, the combined effect of the interpolation method, each room's thermal behavior, and the amount of missing data is still not well understood [21]. In thermal comfort research, missing TSV data is typically addressed using *listwise deletion* [22], *univariate temporal interpolation* [23], or *multivariate imputation* methods [24]. State-of-the-art multivariate frameworks and ensemble machine learning models predicting categorical TSV data typically report a Mean Absolute Error (MAE) between 0.45 and 0.65, exact classification accuracies around 55–65%, and robust ± 1 point tolerance accuracies exceeding 85%. Furthermore, advanced ordinal-aware imputation algorithms and generative models (e.g., missForest, CTGAN) successfully preserve underlying structural integrity, generally yielding a mean absolute correlation perturbation between 0.06 and 0.09 [25]. Benchmark studies for missing-data imputation in time series exist at the general level. Lepot et al. [17] compared interpolation methods for water-quality sensor data, and Toye et al. [26] recently benchmarked 12 imputation methods on health time series (glucose monitoring, heart rate), finding that linear interpolation achieved the lowest RMSE across all missingness mechanisms. However, no analogous domain-specific benchmark exists for categorical subjective comfort data. The distinct challenge of TSV imputation lies in its dual constraint: values must remain physically valid integers on the ASHRAE scale while preserving both the statistical distribution and the correlation structure with continuous environmental predictors. While heavy probabilistic algorithms set the baseline for structural preservation (achieving correlation errors below 0.10), the capacity of high-speed, univariate interpolation methods to meet these strict psychometric and multivariate constraints remains unquantified. Since TSV interpolation is treated purely as a preprocessing step, there is no empirical evidence focused solely on the interpolation stage in this domain. This gap motivates the present study. To address these methodological gaps, this study formulates the following primary research question: *How can missing categorical TSVs be reliably imputed in continuous indoor monitoring datasets without distorting their underlying statistical distribution and correlation with environmental predictors?* To systematically investigate this, the paper addresses two specific sub-questions:

1) Which interpolation method (e.g., linear, polynomial, splines, or shape-preserving) best preserves the original patterns and meaning of categorical data when filling in missing values?
2) How much does rounding interpolated values reduce unrealistic or meaningless results when filling in missing data?

II. Materials and Methods

This section outlines the experimental procedure used to assess various interpolation methods. First, the data collection and preprocessing steps are detailed in Sec. II-A. Next, the specific interpolation methods under consideration are listed in Sec. II-B. Finally, the evaluation criteria used to measure the performance of each method are described in Sec. II-C.

A. Dataset and Preprocessing

The data was collected during the DEDALUS project, which monitored elderly residents (ages 72-90 years old) living in 10 apartments in Treviso, Italy. The dataset pairs human feedback (TSV) with continuous environmental measurements, which are $\tilde{T}_r$ and $\tilde{T}_{\text{out}}$ indoor and outdoor air temperatures respectively, $\widetilde{RH}_r$ the indoor relative humidity, $\tilde{b}$ the forecasted energy consumption for that day, d, h, and s the day of week, hour of day, and season index respectively (each encoded as sine and cosine pairs to capture their cyclical nature), $\Delta\tilde{T}$ the indoor–outdoor temperature difference, and $\mathbb{1}_{\text{day}}$ a binary daytime indicator. To ensure the continuous sensor readings strictly aligned with the human feedback, the data for each apartment was cropped. Any sensor data recorded before a resident's first feedback or after their final feedback was discarded. The subjective feedback was collected using the Telegram messaging app. Residents received automated hourly notifications reminding them to submit their TSV. However, to prevent survey fatigue, participants were advised to submit at least three surveys per day (morning, after lunch and evening) rather than responding to every single hourly notification. A summary of the TSV feedback data per apartment is provided in Table I. Because residents did not sent their feedback every hour, there is a high proportion of missing hourly TSV values. This missing data ranges from 85.2% (Apartment 144, with 9 out of 61 samples) to 96.3% (Apartment 157, with 156 out of 4,262 samples).

B. Interpolation Methods

Sixteen interpolation methods were evaluated, spanning five functional families. The Linear methods for interpolation include (01) Linear, (02) First-order spline, (03) Time-weighted linear interpolation. The next group is the Polynomial Methods which includes (04) Quadratic interpolation, (05) Cubic interpolation, (06) Global polynomial (order 3), (07) Barycentric rational interpolation, (08) Krogh interpolation. The Piecewise-Constant interpolation Methods consist of

TABLE I
MISSING DATA STATISTICS PER APARTMENT. SHOWING THE TOTAL NUMBER OF HOURLY QUESTIONNAIRES SENT N_{total}, THE ACTUAL NUMBER OF QUESTIONNARIES SUBMITTED BY EACH TENANT N_{avail}, AND THE MISSING FEEDBACK COUNT AND PERCENT N_{miss}.

Apartment	N_{total}	N_{avail}	N_{miss}	Missing (%)
144	61	9	52	85.2
145	64	9	55	85.9
147	62	9	53	85.5
149	829	108	721	87.0
151	831	103	728	87.6
153	830	88	742	89.4
154	4 264	162	4 102	96.2
155	63	9	54	85.7
157	4 262	156	4 106	96.3
158	996	126	870	87.3

(09) Nearest-neighbour, and (10) Zero-order hold. Next, Spline Methods includes (11) Smoothing B-spline (order 3), and (12) Natural cubic spline. Finally, Shape-Preserving Methods include (13) Piecewise Cubic Hermite Interpolating Polynomial (PCHIP), and (14) Akima sub-spline.

To provide ordinal-native comparators that inherently respect the discrete ASHRAE scale without requiring post-hoc rounding, two additional methods were included: (15) *Ordinal Step* advances through ASHRAE categories exactly one step at a time, distributing category transitions proportionally across the gap duration; (16) *Ordinal Midpoint* assigns the rounded median of the two bounding observed votes to every position within a gap, creating a piecewise-constant approximation grounded in ordinal arithmetic.

Although methods (01)–(14) apply continuous mathematical interpolation to an ordinal variable, which assumes equal spacing between ASHRAE scale, this is standard practice in TSV literature [27] and is justified by the mandatory post-processing step: all interpolated values are clipped to $[-3, +3]$ and rounded to the nearest valid integer before any analysis, ensuring the imputed series never contains non-physical fractional votes. The ordinal-native methods (15)–(16) serve as empirical reference comparators to test whether natively respecting the discrete scale provides any performance advantage over bounded continuous methods.

C. Evaluation Framework

Two complementary evaluation criteria were employed:

1) Correlation Structure Preservation: To evaluate the impact of interpolation on the data relationships, Pearson correlation [28] matrices were computed for each apartment both before and after applying the imputation methods. The overall structural perturbation was quantified by calculating the mean absolute difference and the Frobenius norm between the original and interpolated matrices [29]. Furthermore, changes in Pearson correlation specifically involving the TSV were isolated to assess how each method affected the target variable's relationship with the environmental predictors.

To assess whether the choice of correlation measure affects method rankings given the ordinal nature of TSV, Spearman rank correlation (ρ_s) [30] was computed in parallel with Pearson for each apartment and method. The mean absolute perturbation in Spearman ρ_s was computed identically to the Pearson metric. Kendall's τ [31] between the Pearson-based and Spearman-based method rankings was then calculated to quantify their consistency.

2) Statistical Distribution Tests: To verify that the interpolation methods preserved the original data's structural characteristics, three independent two-sample tests were applied to each apartment-method pair. Unpaired tests were utilized due to the unequal lengths of the original and interpolated datasets.

The evaluation employed the Kolmogorov-Smirnov (KS) test [32] to detect overall deviations between the empirical cumulative distributions, the Mann-Whitney U test [33] to assess stochastic equality, and the Anderson-Darling k-sample test [34] to specifically capture discrepancies in the distribution tails. All tests were evaluated at a 0.05 significance level. To ensure statistical validity, the performance of each method is reported as a rejection rate, representing the proportion of apartments where the null hypothesis of equal distributions was rejected. [35].

III. RESULTS

A total of $16 \times 10 = 160$ method–apartment experiments were conducted. All sixteen methods converged for every apartment, yielding a complete factorial design.

A. Correlation Structure Preservation

Figure 1 visualises how each interpolation method perturbs the TSV–feature correlation, with methods ordered from best (linear family) to worst (zero-hold) preservation. Table II quantifies these perturbations, showing that the linear family—Linear, Slinear, and Time—achieves the smallest mean absolute correlation change (0.113), with the ordinal-native Ordinal Step method (0.113) matching identically, closely followed by PCHIP and Nearest-neighbour (0.114), while zero-order hold induces the largest perturbation (0.207). The accumulated absolute change follows the same ranking, with zero-order hold accumulating $1.81\times$ more perturbation than the best-performing methods (1.528 vs 0.845). Maximum individual perturbations range from 0.309 (Linear family) to 0.483 (zero-hold), with indoor temperature ($\tilde{T}_r$) being the most frequently perturbed feature across interpolation strategies.

Critically, the Pearson-based and Spearman rank-based method rankings are significantly consistent across all 16 methods (Kendall's $\tau = 0.69$, $p < 0.001$), confirming that the use of Pearson correlation as a relative perturbation metric is equivalent to an ordinal-compatible measure for this dataset.

B. Statistical Distribution Analysis

To assess the distributional fidelity of the imputed data, Figure 2 illustrates the KS rejection rates for each interpolation method across all evaluated apartments. A lower rejection

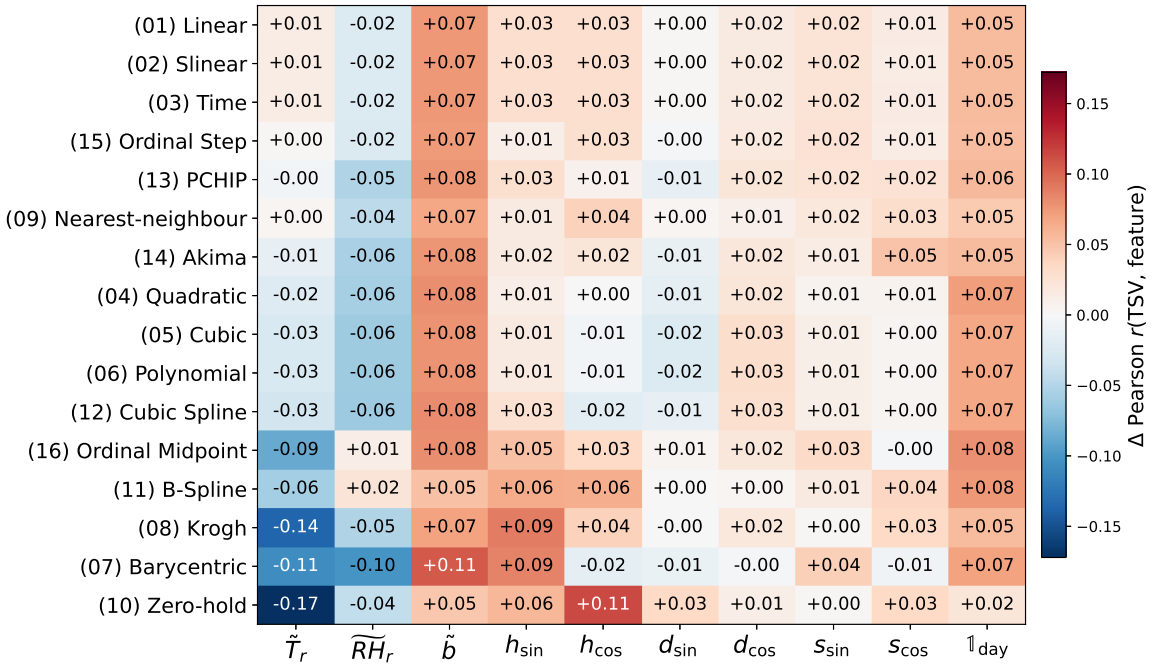


Fig. 1. Mean change in Pearson correlation between TSV and each predictor feature after interpolation, averaged across ten apartments. Rows are ordered by mean absolute perturbation (best preservation at top); the linear family and Ordinal Step are jointly ranked first.

rate (20) indicates a higher success in preserving the original statistical distribution of TSV.

The results demonstrate a robust baseline performance across the majority of the algorithms. 13 out of the 16 evaluated methods, encompassing the linear family, the Ordinal Step method, cubic splines, shape-preserving methods (Akima, PCHIP), nearest-neighbour, and Krogh approaches, achieved a minimal KS rejection rate of 20%. This signifies that in 80% of the test cases, the null hypothesis (that the original and interpolated distributions are stochastically equivalent) could not be rejected. This widespread stability underscores the effectiveness of bounding and rounding the imputed values to match the discrete ASHRAE scale [4], successfully preventing the introduction of non-physical continuous artifacts.

Conversely, three methods exhibited elevated rejection rates. B-Spline reached the highest Anderson-Darling rejection rate of 70% (KS: 40%, Mann-Whitney U: 60%), and Barycentric reached 50% across KS, Mann-Whitney U, and Anderson-Darling tests. Notably, the Ordinal Midpoint method—despite being a natively ordinal approach—also exhibited elevated rates (KS: 30%, Mann-Whitney U: 50%, Anderson-Darling: 50%), revealing that its piecewise-constant structure artificially concentrates mass at intermediate categories and distorts the empirical distribution. This result demonstrates that structural ordinality alone does not guarantee distributional fidelity.

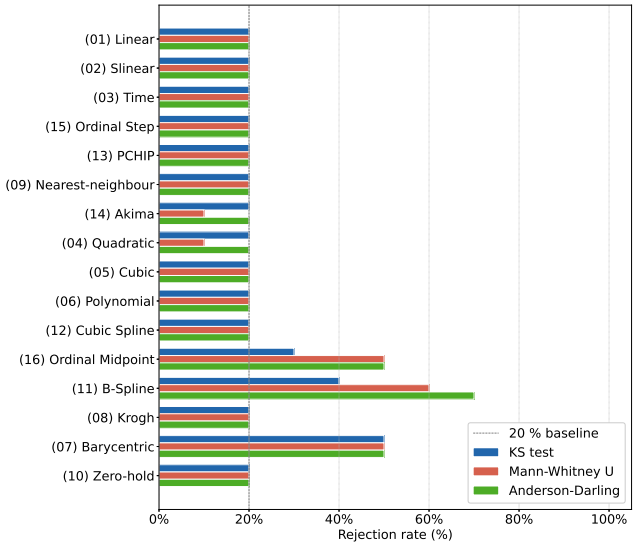


Fig. 2. Rejection rates (proportion of apartments where H_0 was rejected at $\alpha = 0.05$) for three two-sample distribution tests. Methods ordered best-to-worst by mean absolute Pearson correlation perturbation (top to bottom). Dashed line marks the 20% baseline achieved by 13 of 16 methods.

IV. Discussion and Conclusion

Missing data is a pervasive challenge for HITL building control systems and PCMs, where survey fatigue severely fragments essential TSV datasets. Because TSV provides the fundamental ground truth for thermal perception, data continuity is vital for downstream predictive models. To avoid the fractional artifacts of continuous imputation, studies using the ordinal 7-point ASHRAE scale have historically relied on listwise deletion, causing severe information loss [22]. Conversely, advanced imputation techniques like GANs [24] and Spline Regression [23] have strictly targeted continuous

TABLE II
CORRELATION PERTURBATION (CHANGE AFTER THE INTERPOLATION) (MEAN ACROSS APARTMENTS).
MEAN: MEAN ABSOLUTE CORRELATION CHANGE (BEFORE AND AFTER THE INTERPOLATION) ACROSS THE FEATURES;
ACCUM: ACCUMULATED ABSOLUTE CHANGE IN CORRELATION AFTER THE INTERPOLATION ACROSS THE FEATURES;
MAX: MAXIMUM CORRELATION PERTURBATION;
FEATURE: MOST CORRELATION CHANGED FEATURE.

Method	Mean	Accum	Max	Feature
(01) Linear	0.113	0.845	0.309	$\tilde{T}_r$
(02) Slinear	0.113	0.845	0.309	$\tilde{T}_r$
(03) Time	0.113	0.845	0.309	$\tilde{T}_r$
(15) Ordinal Step	0.113	0.845	0.322	$\tilde{T}_r$
(13) PCHIP	0.114	0.856	0.328	$\tilde{T}_r$
(09) Nearest-neighbour	0.114	0.859	0.325	$\tilde{T}_r$
(14) Akima	0.123	0.928	0.346	$\tilde{T}_r$
(04) Quadratic	0.133	1.006	0.350	$\tilde{T}_r$
(05) Cubic	0.136	1.026	0.350	$\tilde{T}_r$
(06) Polynomial	0.136	1.026	0.350	$\tilde{T}_r$
(12) Cubic Spline	0.138	1.039	0.345	$\tilde{T}_r$
(16) Ordinal Midpoint	0.150	1.119	0.329	$h_{\sin}$
(11) B-Spline	0.154	1.153	0.394	$h_{\sin}$
(08) Krogh	0.166	1.269	0.407	$\tilde{T}_r$
(07) Barycentric	0.179	1.393	0.460	$\tilde{T}_r$
(10) Zero-hold	0.207	1.528	0.483	$\tilde{T}_r$

Note: Numbers (XX) refer to the interpolation methods' numbering stated in Section II-B

physical variables, leaving a critical gap in handling ordinal, human-centric data.

To address this gap, this study systematically evaluated 16 univariate interpolation algorithms across 10 distinct indoor environments to reconstruct highly sparse TSV datasets. We empirically demonstrate that applying discrete bounding and post-interpolation rounding to standard linear and cubic interpolations successfully imputes categorical subjective data, offering a robust alternative to both naive data deletion and unconstrained continuous imputation.

The results establish a clear hierarchy in interpolation efficacy and highlight the necessity of data processing constraints to maintain high measurement quality:

- **High Fidelity vs. State-of-the-Art Generative Models:** The linear family (Linear, Slinear, Time) and the natively ordinal Ordinal Step method (15) achieved the lowest mean absolute correlation perturbation (0.113), closely followed by PCHIP and Nearest-neighbour (0.114). In recent literature, computationally massive, ordinal-aware generative models (e.g., missForest, CTGAN) establish the baseline for structural preservation, typically yielding correlation perturbations between 0.06 and 0.09 [25], [36]. While the proposed linear interpolation exhibits a marginally higher perturbation (0.113), it confirms that linear approximations—when adequately bounded and rounded—successfully preserve the underlying multivariate dependencies without suffering structural distortion.
- **The Critical Role of Rounding:** Enforcing the discrete ASHRAE boundaries effectively neutralized the non-physical overshooting typically associated with continuous splines. Consequently, 13 of the 16 methods maintained robust distributional fidelity, achieving a low 20% rejection rate in KS tests.
- **Underperforming Methods:** Zero-order hold proved highly destructive to the correlation structure, accumulating 1.81× more total error than the best linear methods (1.528 vs 0.845, mean perturbation 0.207). B-Spline and Barycentric failed distributional tests (B-Spline: 70% AD, 60% MW; Barycentric: 50% across all three tests). Importantly, the Ordinal Midpoint method—despite being natively ordinal—also failed (KS: 30%, MW: 50%), confirming that the type of interpolation framework matters more than its philosophical alignment with the scale type.

By benchmarking these methods, this study advances the foundational knowledge required to build robust PCMs. The slight trade-off in correlation fidelity (0.113 versus the ∼0.08 of advanced probabilistic models) is outweighed by the elimination of algorithmic latency. When time-series data for apartments are cropped to start at the first human feedback observation and end at the last, creating bounded temporal windows, deploying iterative matrix operations becomes a severe bottleneck. Within protocols requiring high-frequency subjective questionnaires—such as those administered every 4 minutes during the acclimatization period—the latency of complex ML imputations prevents real-time responsiveness. Thus, researchers should employ constrained cubic or linear interpolation as a computationally lightweight and highly effective data augmentation pipeline that can be directly integrated into smart building dashboards and DR ecosystems.

The results of this assessment demonstrate that the choice of interpolation method significantly impacts the reliability of TSV. In the context of the DEDALUS project, where data from elderly inhabitants often contain gaps due to a lack of TSVs, robust interpolation ensures continuous longitudinal thermal analysis. By recovering missing TSVs, researchers can better understand the nuances of thermal vulnerability in aging populations. By rescuing fragmented data, this work actively pushes the boundaries of the comfort field, shifting the paradigm from discarding imperfect subjective surveys to rigorously recovering them. Furthermore, within the broader measurement field, this study provides a benchmark for selecting algorithms that balance computational efficiency with physical accuracy, ensuring that transient comfort states are not lost to data sparsity.

Despite these promising findings, this study is subject to certain limitations. The evaluation relied exclusively on univariate temporal interpolation, and the dataset was restricted to a specific demographic (elderly occupants) within two buildings.

Most importantly, while this study establishes the statistical and structural fidelity of the imputed data, it did not evaluate the performance of downstream machine learning models trained on these datasets. Therefore, a critical next step for future work is to explicitly quantify the direct impact of the interpolation choice on ML model accuracy. Specifically,

future research will train tree-based and ordinal predictive architectures on the augmented datasets to compare model performance with and without the interpolation step.

Furthermore, future work should compare these univariate baselines against multivariate imputation techniques, such as MICE or deep learning-based autoencoders, to determine the optimal preprocessing pipeline for PCMs.

ACKNOWLEDGMENT

This work is under the framework of DEDALUS project financed by the European Union's Horizon Europe research and innovation program under the Grant Agreement No. 101103998. The authors want to thank the DEDALUS partners for the fruitful and precious collaboration.

REFERENCES

[1] P. Zech, S. Hammes, D. Geisler-Moroder, M. Hauer, R. Pfluger, and R. Breu, "Quantifying the simulation-reality gap in building simulations," in *Building Simulation 2025*, vol. 19, pp. 0–0, IBPSA, 2025.

[2] O. Badreddine, H. Radoine, and R. Hajji, "Twincity: An urban digital twin framework for data-scarce environments—a case study of benguerir, morocco," *Smart Cities*, vol. 9, no. 2, p. 23, 2026.

[3] N. Morresi, S. Casaccia, M. Sorcinelli, M. Arnesano, A. Uriarte, J. I. Torrens-Galdiz, and G. M. Revel, "Sensing physiological and environmental quantities to measure human thermal comfort through machine learning techniques," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 12322–12337, 2021.

[4] ASHRAE, "ANSI/ASHRAE Standard 55-2020 – Thermal Environmental Conditions for Human Occupancy," 2020.

[5] G. Cosoli, N. Abo-Alzahab, D. Ago, S. Murazzo, R. Christoforou, M. Moayyedi, M. Schweiker, M. Arnesano, I. Ciuffreda, and G. M. Revel, "A measurement approach for the assessment of indoor multidomain comfort: The multiclimact experience," in *2025 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pp. 1319–1324, IEEE, 2025.

[6] S. Stancu, A. M. Hristea, C. Kailani, D. E. Bălă, and A. Pernici, "Adoption of voluntary measures for reducing electricity consumption in the repowereu plan context: A romanian consumer perspective," *Amfiteatru Economic*, vol. 25, no. 64, pp. 760–779, 2023.

[7] M. M. Sequeira, J. P. Gouveia, and J. Joanaz de Melo, "Can local organizations act as middle actors in energy support? exploring their functions, motivations, challenges, and needs," *Energy Efficiency*, vol. 17, no. 7, p. 81, 2024.

[8] L. Toderean, T. Cioara, I. Anghel, E. Sarmas, V. Michalakopoulos, and V. Marinakis, "Demand response optimization for smart grid integrated buildings: Review of technology enablers landscape and innovation challenges," *Energy and Buildings*, vol. 326, p. 115067, 2025.

[9] B. Amangeldy, T. Imankulov, N. Tasmurzayev, S. Aibagarov, N. Azatbekuly, G. Dikhanbayeva, and A. Mukhanbet, "A pilot study on thermal comfort in young adults: Context-aware classification using machine learning and multimodal sensors," *Buildings*, vol. 15, no. 15, p. 2694, 2025.

[10] X. Liang, F. de Nijs, B. Say, and H. Wang, "Human-in-the-loop ai for hvac management enhancing comfort and energy efficiency," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, pp. 359–370, 2025.

[11] T. Sawada, M. Mizuno, T. Hasegawa, K. Yokoyama, and M. Kono, "Office-in-the-loop: an investigation into agentic ai for advanced building hvac control systems," *Data-Centric Engineering*, vol. 6, p. e31, 2025.

[12] Q. Zheng, X. Zhou, X. Li, Y. Liang, B. Cao, Y. He, P. Li, and M. Luo, "Usage behavior and comfort effects of iot-connected personalized environmental control systems," *Journal of Building Engineering*, vol. 98, p. 111487, 2024.

[13] Z. D. Tekler, Y. Lei, Y. Peng, C. Miller, and A. Chong, "A hybrid active learning framework for personal thermal comfort models," *Building and Environment*, vol. 234, p. 110148, 2023.

[14] S. Zhang, Y. Cheng, Z. Fang, C. Huan, and Z. Lin, "Optimization of room air temperature in stratum-ventilated rooms for both thermal comfort and energy saving," *Applied Energy*, vol. 204, pp. 420–431, 2022.

[15] S. v. Buuren, K. Groothuis-Oudshoorn, A. Robitzsch, G. Vink, L. Doove, and S. Jolani, "Multivariate imputation by chained equations," *Comprehensive R Archive*, vol. 587, 2019.

[16] W. M. Campion, "Book review: multiple imputation for nonresponse in surveys," 1989.

[17] M. Lepot, J.-B. Aubin, and F. H. Clemens, "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment," *Water*, vol. 9, no. 10, p. 796, 2017.

[18] Q. Aumann and I. V. Gosea, "Practical challenges in data-driven interpolation: dealing with noise, enforcing stability, and computing realizations," *International Journal of Adaptive Control and Signal Processing*, vol. 39, no. 10, pp. 2062–2080, 2025.

[19] M. Hu, J. Yin, Z. Ma, J. Ma, F. Zhu, B. Wu, Y. Wen, M. Wu, C. Hu, B. Hu, *et al.*, "beta-fft: Nonlinear interpolation and differentiated training strategies for semi-supervised medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30839–30849, 2025.

[20] A. S. AlSalehy and M. Bailey, "Improving time series data quality: Identifying outliers and handling missing values in a multilocation gas and weather dataset," *Smart Cities*, vol. 8, no. 3, p. 82, 2025.

[21] I. A. Campodonico-Avendano, S. Erba, P. Papadopoulos, S. Carlucci, A. Luparelli, A. Ingrosso, G. Tresoldi, M. S. Shahid, F. Wurtz, B. Delinchant, *et al.*, "Collectief dataset: A high-resolution indoor environmental dataset from european buildings across diverse climates supporting thermal, air-quality, and visual-comfort assessments," *Data in Brief*, p. 112486, 2026.

[22] D. Geethamani, A. Bhuvaneshwari, A. Sagayapriya, *et al.*, "Thermal comfort prediction using machine learning: A comparative study of algorithms and features," *PREPRINT (Version 1) available at Research Square*, 2025.

[23] Y. Kim, S. Kelly, D. Krishnan, J. Falletta, and K. Wilmot, "Strategies for imputation of high-resolution environmental data in clinical randomized controlled trials," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1307, 2022.

[24] M. Muslimsyah, S. Safwan, and A. Novandri, "Comprehensive assessment of indoor thermal in vernacular building using machine learning model with gan-based data imputation: A case of aceh region, indonesia," *Buildings*, vol. 15, no. 14, p. 2448, 2025.

[25] E. Albu, S. Gao, L. Wynants, and B. Van Calster, "missforestpredict—missing data imputation for prediction settings," *Plos one*, vol. 20, no. 11, p. e0334125, 2025.

[26] A. A. Toye, A. Celik, and S. Kleinberg, "Benchmarking missing data imputation methods for time series using real-world test cases," *JAMIA Open*, vol. 8, no. 2, p. ooaf040, 2025.

[27] M. Schweiker, G. M. Huebner, B. R. M. Kingma, R. Kramer, and H. Pallubinsky, "Analysis of subjective thermal comfort data: A statistical point of view," *Building and Environment*, vol. 225, p. 109563, 2022.

[28] P. Sedgwick, "Pearson's correlation coefficient," *Bmj*, vol. 345, 2012.

[29] G. Cheng, Z. Liu, and Q. Xiong, "Testing equality of two high-dimensional correlation matrices," *Communications in Mathematics and Statistics*, pp. 1–23, 2025.

[30] C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[31] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1–2, pp. 81–93, 1938.

[32] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[33] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.

[34] F. W. Scholz and M. A. Stephens, "K-sample anderson–darling tests," *Journal of the American Statistical Association*, vol. 82, no. 399, pp. 918–924, 1987.

[35] N. A. Heard and P. Rubin-Delanchy, "Choosing between methods of combining-values," *Biometrika*, vol. 105, no. 1, pp. 239–246, 2018.

[36] A. Danholt Lautrup, T. Hyrup, A. Zimek, and P. Schneider-Kamp, "Syntheval: A framework for detailed utility and privacy evaluation of tabular synthetic data," *arXiv e-prints*, pp. arXiv–2404, 2024.