A BIM-INTEGRATED ADAPTIVE SURROGATE FRAMEWORK FOR EFFICIENT UNCERTAINTY QUANTIFICATION IN BUILDING ENERGY SIMULATION

Arash B. Naamrudi[1], Reihaneh Aghamolaei[1]
[1] School of Mechanical & Manufacturing Engineering, Faculty of Engineering and Computing, Dublin City University, Whitehall, Dublin 9, Ireland

Abstract

Uncertainty analysis is critical for reliable building energy predictions, yet classical Monte Carlo simulation is computationally prohibitive for detailed physics-based models. This study proposes a BIM-integrated adaptive surrogate uncertainty analysis framework combining adaptive sampling with sparse Polynomial Chaos Expansion to efficiently propagate uncertainty and perform global sensitivity analysis. The approach is demonstrated on four Irish residential archetypes with uncertain envelope thermal parameters. The adaptive surrogates achieve stable accuracy (CV-RMSE 0.74–2.05 kWh/m² for annual EUI) while reducing computational cost by around 99% compared to Monte Carlo simulation, enabling scalable and interpretable decision support for building design and retrofit.

Introduction

Buildings are a significant contributor to global energy use and carbon emissions, accounting for about 34–37% of total energy consumption worldwide (United Nations Environment Programme, 2024). In the European Union, residential buildings consume nearly 40% of total energy and are responsible for roughly 33% of energy-related greenhouse gas emissions, largely due to an ageing and inefficient building stock with low renovation rates (Energy Performance of Buildings Directive, 2025). Improving the energy performance of existing buildings is therefore critical to reducing energy demand and achieving a fully decarbonised building stock by 2050 (Energy Performance of Buildings Directive, 2025). In response, energy standards such as ASHRAE Standard 90.1 (ASHRAE Standard, 2022) and the International Energy Conservation Code (IECC) (International Code Council (ICC), 2021) have progressively tightened efficiency requirements, increasing the need for accurate yet computationally efficient Building Energy Models (BEMs) to support compliance, retrofit evaluation, and design optimisation (ElSayed, Shultz and Kurtz, 2025). BEM is a computational simulation approach used to predict the energy performance of buildings under defined operating conditions. The typical inputs to a BEM can be broadly categorised into five groups, including geometric and spatial parameters, envelope thermal properties, occupancy and internal load schedules, HVAC system parameters and boundary conditions. The primary outputs of interest in building energy analysis include the annual and monthly Energy Use Intensity (EUI, in kWh/m²), total heating and cooling demand (kWh), peak load (kW), and thermal comfort indicators.

Building Information Modelling (BIM) has increasingly been integrated with BEM to improve data consistency and automation in energy analysis (Heydari and Heravi, 2023). BIM provides a structured, data-rich representation of buildings, incorporating geometry alongside material properties and performance attributes (Shadram *et al.*, 2016; Wang *et al.*, 2022), which can be translated into energy simulations (Arbulu, Oregi and Etxepare, 2024; Forth and Borrmann, 2024). Within BIM-based energy modelling workflows, building performance is most commonly evaluated using deterministic simulation approaches, in which all model inputs are assigned fixed, single-point values (Alwan *et al.*, 2021; Heydari and Heravi, 2023). These values are typically derived from design specifications, standards, or nominal assumptions, implicitly assuming that building components behave exactly as prescribed and that variability arising from construction quality, ageing, occupant behaviour, and environmental conditions is negligible (Liu *et al.*, 2022; Prataviera *et al.*, 2022). While deterministic BIM-based simulations provide a streamlined pathway from design data to energy performance prediction, they are inherently limited by their reliance on single-point input assumptions, which obscure the effects of uncertainty and variability in building parameters.

To address this limitation and consider variability and uncertainty in building parameters, uncertainty-aware modelling frameworks have been developed that explicitly represent uncertain inputs using probabilistic or stochastic descriptions and propagate their effects through building energy models. Within this context, Forward Uncertainty Analysis (FUA) defines the analytical task of quantifying how uncertainty in model inputs translates into uncertainty in energy performance outputs.

FUA can be implemented using a range of numerical and surrogate-based techniques, including Monte Carlo simulation, Latin Hypercube sampling, and surrogate models such as Polynomial Chaos Expansion or Gaussian process regression. Yet, existing approaches face significant computational constraints since they require a high number of simulations over fixed and predefined samples. For example, Monte Carlo Simulation (MCS), the conventional benchmark for FUA, propagates uncertainties through repeated random sampling (Liu *et al.*, 2020; Chadly *et al.*, 2023; Prilliman *et al.*, 2023). Although theoretically robust, MCS converges slowly, with accuracy improving only with the inverse square root of sample size, often requiring thousands of simulations (Zhao, Lee and Augenbroe, 2016). Given the high computational cost of individual BEM evaluations, this renders MCS impractical for many real-world applications.

To improve the efficiency MCS, static sampling strategies such as Latin Hypercube Sampling (LHS) have been proposed. LHS can reduce the required number of simulations by approximately 50% compared to MCS for similar accuracy (Tian *et al.*, 2020; Koniorczyk *et al.*, 2022). However, LHS allocates samples uniformly across the input domain, irrespective of model sensitivity or non-linearity, leading to inefficient use of computational resources in complex regions of the input space. Adaptive sampling strategies address this limitation by sequentially selecting new samples based on output variance, achieving reported simulation reductions of 25–68% relative to static designs (Yu *et al.*, 2025). Despite these gains, adaptive sampling becomes challenging when applied to high-fidelity BEMs with multiple outputs and temporal resolutions, where the dimensionality of objectives significantly increases computational burden.

Surrogate modelling offers a pathway to overcoming these computational barriers by approximating expensive BEM simulations with inexpensive metamodels trained on limited high-fidelity data (Zhao, Lee and Augenbroe, 2016; Sood *et al.*, 2025). Among surrogate approaches, Polynomial Chaos Expansion (PCE) is particularly attractive due to its explicit polynomial representation of input–output relationships, enabling analytical computation of statistical moments and Sobol' sensitivity indices directly from model coefficients (Tian, 2024). This analytical structure provides both uncertainty and sensitivity information within a single framework and enhances interpretability through direct examination of polynomial terms. However, classical PCE suffers from the curse of dimensionality, as the number of basis terms grows combinatorially with input dimension and polynomial order, making it impractical for building models with moderate to high numbers of uncertain parameters (Thapa, Mulani and Walters, 2020).

To address this challenge, adaptive PCE methods have been developed that iteratively refine the polynomial basis by retaining only influential terms and pruning negligible contributions, reducing required simulations by 40–60% compared to full-order PCE (Cheng and Lu, 2018; Thapa *et al.*, 2024). When combined with sequential sampling strategies, adaptive PCE further improves efficiency by allowing both the surrogate structure and training data to evolve based on observed model behaviour (Thapa, Mulani and Walters, 2020). Despite these advances, adaptive PCE has not been systematically integrated into BIM-based BEM workflows, particularly for simultaneous multi-output and multi-timescale analysis.

The literature, therefore, reveals three critical gaps. First, no existing framework unifies adaptive sampling and adaptive PCE within automated BIM-to-BEM workflows, with current approaches relying on either computationally expensive Monte Carlo methods or static surrogate training designs. Second, most uncertainty and sensitivity studies focus on single outputs and single temporal scales, failing to capture how parameter importance varies across different performance metrics and timescales, a limitation that can misinform retrofit decisions. Third, the scalability of adaptive PCE for multi-output, multi-timescale building energy analysis remains largely unvalidated. Addressing these gaps, this study proposes a unified BIM–BEM framework that integrates adaptive sampling with adaptive PCE to efficiently quantify uncertainty and sensitivity across multiple outputs and temporal resolutions and benchmarks its performance against Monte Carlo simulation.

Methodology

BIM-based BEM Transformation

The proposed framework (Figure 1) integrates BIM-based building modelling with adaptive surrogate-assisted uncertainty and sensitivity analysis. A BEM is first derived from a BIM representation, focusing on envelope-related parameters, including wall, roof, and window U-values and window solar heat gain coefficient (SHGC). These parameters are treated as uncertain inputs and assigned probability distributions. An adaptive PCE surrogate, coupled with sequential sampling, is then constructed to propagate uncertainties and evaluate their influence on multiple performance metrics. The framework employs a two-level adaptivity strategy, combining variance-based basis refinement with sequential experimental design. At each iteration, surrogate accuracy is improved through targeted basis enrichment and adaptive sampling, while convergence is assessed using distribution-based criteria.

The BIM-based building model was developed by identifying geometric and non-geometric parameters that define the physical and thermal characteristics of the building, including story configuration, ceiling height, openings, construction assemblies, and thermal properties. Parameter values were obtained from literature and established databases to ensure representativeness. Autodesk Revit was selected as the BIM authoring platform due to its parametric data structure, interoperability, and industry adoption. Thermal attributes such as U-values, layer thicknesses, and SHGC were extracted directly using a single-step BIM-to-BEM approach to minimise data loss and improve reliability.

To bridge BIM and energy simulation domains, a transitional energy model was generated in HBJSON format using the Pollination plugin. This intermediate representation preserves geometry, zoning, and aperture definitions while enabling quality control of spatial and thermal data. Aggregating spaces with similar thermal behaviours defined thermal zones. Fenestration elements were mapped directly from Revit window families, and

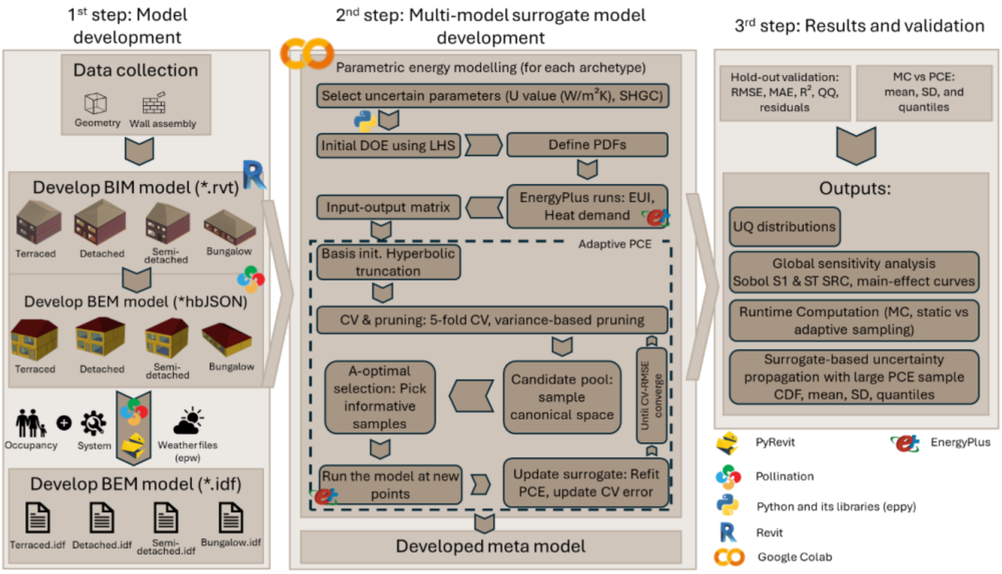

Figure 1. Overarching methodology diagram of the proposed framework.

boundary locations were assigned to wall centrelines to ensure surface alignment and shared interfaces. The transitional model was subsequently validated to correct envelope discontinuities, geometric gaps, and orthogonality issues. Once validated, the model was exported in EnergyPlus IDF format and enriched with material properties, weather data (EPW), occupancy schedules, HVAC parameters, and internal loads, forming the final simulation-ready BEM (Figure 2).

Adaptive Uncertainty Quantification Framework

Uncertainty in envelope parameters was propagated through the BEM using an adaptive PCE framework. Uncertain inputs were modelled as independent normally distributed random variables and transformed into standard normal space to satisfy orthogonality requirements. The building energy response was approximated using a truncated generalised PCE, where multivariate orthogonal basis functions are constructed as tensor products of univariate polynomials. Statistical moments of the output, including mean and variance, were obtained analytically from the PCE coefficients.

When using PCE, the uncertainty is represented through a set of independent random variables $\mathbf{X} = (X_1, \ldots, X_d)$, each characterised by a normal probability distribution. In order to allow the construction of a generalised Polynomial Chaos Expansion (gPCE), each variable is transformed to a standardised form $\xi_i = (X_i - \mu_i)/\sigma_i$, such that $\xi_i \sim \mathcal{N}(0,1)$ and μ_i, and σ_i are the mean and standard deviation of the i-th parameter, respectively. The mapping between physical parameters $\mathbf{x}$ and standard normal variables ξ ensures the orthogonality properties required for PCE convergence (Thapa, Mulani and Walters, 2020; Thapa *et al.*, 2024).

In PCE, $u(x(\xi))$, which is a second-order stochastic response with finite variance and dependent on a set of independent random inputs $x = \{x_i\}_{i=1}^{N}$, is represented as a spectral expansion using multivariate orthogonal polynomials $\psi_k(\xi)$ and deterministic PCE coefficients a_k as in Eq 1.

$$u(x(\xi)) \approx \sum_{k=1}^{\infty} a_k \Psi_k(\xi) \qquad 1$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, is a multi-index with non-negative integer components, $a_{\boldsymbol{\alpha}}$ are the PCE coefficients to be determined, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ are multivariate orthonormal polynomial basis functions. More detailed descriptions and formulas of adaptive PCE are presented in the related literature (Sudret, 2008, 2008; Blatman and Sudret, 2011).

To address the curse of dimensionality, an adaptive PCE formulation (VARPCE) was employed, in which polynomial basis functions are introduced incrementally in small subsets. Variance-based pruning was applied to remove basis terms with negligible contributions while ensuring that the relative loss of total variance remained below a predefined tolerance. Polynomial order selection and overfitting prevention were managed using k-fold cross-validation, with model complexity increased only when cross-validation error showed meaningful improvement.

To validate the adaptive PCE surrogate, a hold-out set of N = 50 direct EnergyPlus simulations, drawn from an independent random seed withheld from the training process, was used as the ground truth. Moreover, the computational efficiency of the adaptive PCE surrogate was benchmarked against direct Monte Carlo simulation using EnergyPlus as the ground truth solver.

Adaptive sampling was integrated with adaptive PCE to efficiently enrich the experimental design. Initial samples were generated using Latin Hypercube Sampling to ensure global coverage of the input space. As the active basis evolved, additional samples were selected using an A-optimal design criterion, which minimises the variance of PCE coefficient estimates and targets regions of high surrogate uncertainty. This strategy balances global exploration with local refinement, reducing redundant simulations and improving surrogate stability (Figure 3). The adaptive sampling procedure is:

1- Generate an initial Latin Hypercube Sample (LHS) of size N_0
2- Evaluate computational model at initial design points
3- For polynomial order $p = 1,2,\ldots,p_{\max}$:
   a. Fit PCE using current data
   b. Apply variance-based pruning
   c. Compute cross-validation error
   d. If the CV convergence criterion is met, stop
   e. Otherwise, for $\ell = 1,\ldots,L_{\text{adapt}}$ (here $L_{\text{adapt}} = 5$):
      - Generate a candidate set of size N_{cand} via LHS
      - Select n_{new} points (here $n_{\text{new}} = 3$) using the A-optimal criterion
      - Evaluate model at new points
      - Update PCE and check CV convergence

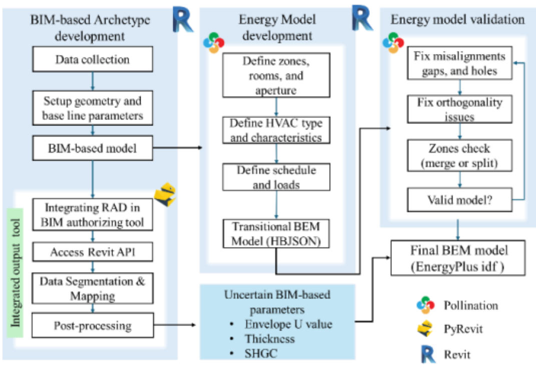

Figure 2. Developing a BIM-based energy-simulation-ready file format.

Sensitivity Analysis

Global sensitivity analysis was performed analytically using the final PCE coefficients. First-order and total-effect Sobol' indices were computed to quantify both main effects and interaction contributions of each uncertain input parameter. Convergence of the adaptive process was primarily assessed using the Kullback–Leibler divergence, which measures stability of output probability distributions across iterations and aligns directly with uncertainty quantification objectives. Secondary validation metrics, including leave-one-out error and the coefficient of determination (Q^2), were used to confirm predictive accuracy. Once the convergence criteria were satisfied, uncertainty and sensitivity results were finalised and benchmarked against Monte Carlo simulation to evaluate computational efficiency and accuracy.

Case study and data preparation

The proposed methodology is applied to four representative Irish residential archetypes: terraced, bungalow, detached, and semi-detached houses. These archetypes are chosen as they have been tested and validated in previous studies (Ali *et al.*, 2024; Sood *et al.*, 2025) and are among the most common archetypes in Ireland. BIM models for each archetype are developed in Revit (Figure 4) to capture both geometric and non-geometric attributes, including floor areas, ceiling heights, fenestration, and thermal envelope properties.

A PyRevit-based workflow is used to directly extract data via the Revit API, enabling a unified, automated, and user-friendly BIM data integration. Envelope-related thermal parameters are treated as uncertain inputs with normal distributions, while deterministic parameters are fixed at literature-based mean values.

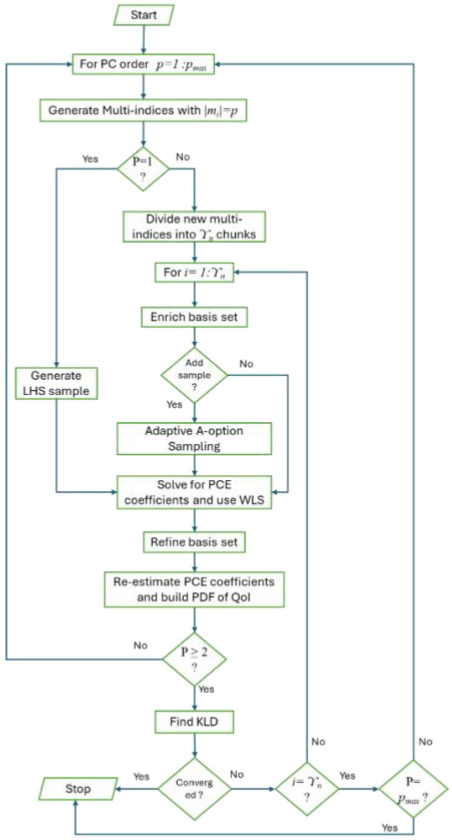

Figure 3. Flowchart of integrating adaptive sampling and adaptive PCE.

The uncertain inputs comprised envelope-related parameters, including U-values of the external walls, roof, and floor, the window U-value, and the solar heat gain coefficient (SHGC), each assigned normal probability distributions as described in the previous Section. Deterministic parameters, such as setpoint temperatures and lighting density, were fixed at mean values derived from the literature (Table 1).

Initial uncertainty propagation was performed using 30 Latin Hypercube samples to generate the training dataset. The adaptive PCE framework was then applied, with polynomial order incrementally increased up to the fourth order. A variance-based pruning strategy was employed to discard basis terms with negligible contribution, while new input samples were iteratively added using an A-optimal sequential design until convergence was achieved. Convergence was determined when the Kullback–Leibler divergence (KLD) between successive output distributions fell below 0.01 and the coefficient of determination (Q^2) exceeded 0.9. The outputs of interest were annual and monthly energy use intensity and annual and monthly heating demand (kWh).

After convergence, global sensitivity analysis was performed using Sobol' indices derived directly from the final PCE coefficients. Both first-order and total-order Sobol' indices were computed for each uncertain parameter, providing interpretable measures of their individual contributions and interaction effects on the outputs. In the final stage, the computational cost of this method will be compared with that of the classic Monte Carlo.

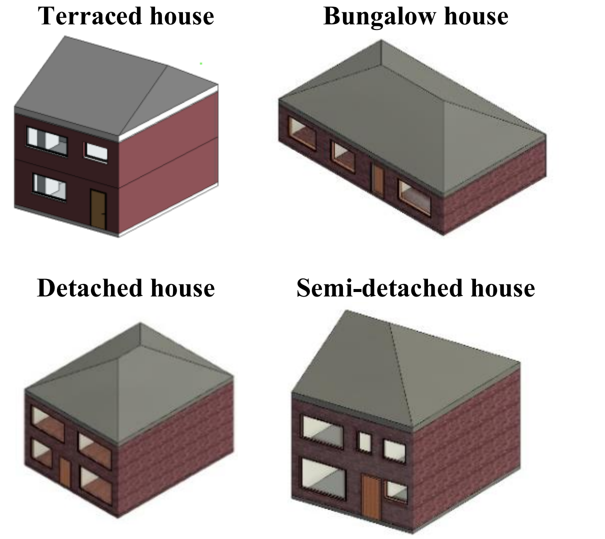

Figure 4. Common archetypes for Irish residential buildings in Revit.

Results and discussion

As shown in Table 2, the convergence history demonstrates that the adaptive sparse PCE model reaches stability very quickly while requiring only a small number of EnergyPlus simulations. At first order with 30 samples, for Bungalow as an example, the meta model contains six terms and already achieves a low Cross-Validation Root Mean Square Error (CV-RMSE) of 2.05 kWh/m², showing that annual EUI is dominated by linear effects, which is typical for U-value perturbations in building envelope analysis. When the adaptive routine adds new samples, the temporary rise in CV-RMSE reflects normal cross-validation variability rather than deteriorating model quality, and the consistent retention of only six terms confirms that higher-order interactions remain negligible.

Table 1. Deterministic parameter values and range of the variables with uncertainty for enriching final energy models (Egan et al.*, 2018; Ali* et al.*, 2024; Sood* et al.*, 2025).*

HVAC and system	Utility energy consumption (kWh/year.m2)			3
	Heating setback (°C)			12
	HVAC efficiency			2.2
	Heating setpoint (°C)			20.5
	DHW (l/m2/day)			2
	Lighting density (W/m2)			5
Internal loads	Occupants (person)			3.5
	Appliance density (W/m2)			11
Envelope	External wall U value (W/m2K)	0.1	1.1	
	Floor U value (W/m2 K)	0.1	1.1	
	Roof U value (W/m2 K)	0.1	0.9	
	Windows U value (W/m2 K)	0.6	4.3	
	SHGC of glass	0.1	0.9	
	Infiltration (ACH at 50 Pa)			1.85

After increasing the polynomial order to two, the meta model for semi-detached retains just 11 terms following pruning, indicating that most quadratic and interaction components do not meaningfully contribute to output variance. This modest growth in active terms aligns with the weak nonlinearities expected from EnergyPlus under small parameter perturbations. Throughout all iterations, the stability of total variance, which remains within the range of approximately 255–284, shows that the model captures the essential variance structure early and preserves it across refinements.

Table 2. Analytical variation of PCE convergence history.

Archetype	Order	No. terms	No. samples	CV-RMSE	Action	Itr.
Semi-detached	1	6	30	1.82	pp	
	1	6	33	2.44		1
	1	6	36	2.92		2
	2	11	36		pp	
	2	11	39	0.88		1
	2	11	42	0.82		2
	2	11	45	0.80		3
	2	11	48	0.78		4
	2	11	51	0.72		5
Detached	1	6	30	1.84	pp	
	1	6	33	2.26		1
	1	6	36	2.25		2
	2	11	36		pp	
	2	11	39	0.81		1
	2	11	42	0.77		2
	2	11	45	0.87		3
	2	11	48	0.70		4
Terraced	1	6	30	1.865	pp	
	1	6	33	2.48		1
	1	6	36	2.23		2
	2	11	36		pp	
	2	11	39	0.83		1
	2	11	42	0.85		2
	2	11	45	0.82		3
Bungalow	1	6	30	2.05	pp	
	1	6	33	2.27		1
	1	6	36	2.20		2
	2	11	36		Pp	
	2	11	39	1.06		1
	2	11	42	0.85		2
	2	11	45	0.84		3
	2	11	48	0.82		4
	2	11	51	0.74		5

*pp: post-prune; Itr: Iteration

The final CV-RMSE of the surrogate model for semi-detached, detached, terraced and bungalow are 0.72, 0.70, 0.82, 0.74 kWh/m² at second order, respectively confirming that the surrogate has reached its optimal balance between accuracy and complexity. Additional samples or higher-order terms do not improve performance, signalling classical convergence in adaptive PCE. Overall, the results show that the relationship between envelope parameters and annual EUI is smooth and predominantly linear, allowing an accurate surrogate to be built using approximately 50 EnergyPlus simulations for the semi-detached archetype with excellent computational efficiency.

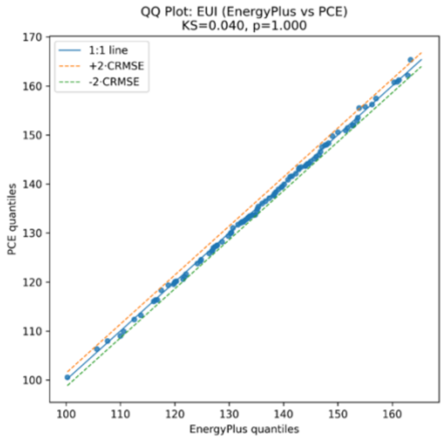

Figure 5. QQ comparison between the semi-detached surrogate predictions and the Energy Plus results.

Figure 5 shows excellent distributional agreement between PCE surrogate predictions and EnergyPlus outputs, with QQ points closely aligned to the 1:1 line across the full EUI range. The KS test (KS = 0.040, p = 1.000) confirms that both distributions are statistically indistinguishable. Nearly all points lie within ±2×CRMSE bounds, indicating small, well-controlled errors that are predominantly random rather than systematic, demonstrating high-fidelity probabilistic emulation by the surrogate.

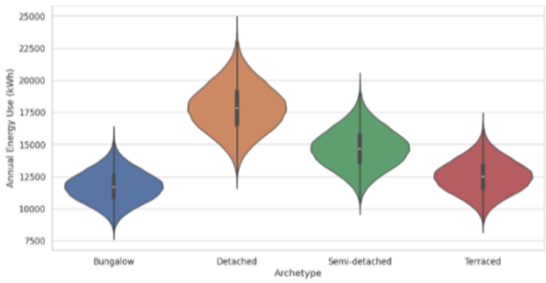

Figure 6. Violin plot of annual energy use for different archetypes.

The violin plot in Figure 6 illustrates the distribution of annual energy use (kWh) across four distinct building archetypes. This illustrates how input uncertainties in envelope U-values and SHGC propagate into annual energy variability. Notably, Detached properties exhibit the highest median annual energy consumption, as well as the broadest distribution, indicating a wider range of energy use within this archetype. Semi-detached properties show a slightly lower median energy use compared to detached, with a similarly wide but slightly less dispersed distribution. Terraced houses consistently demonstrate the lowest median annual energy use, with a comparatively tighter distribution, suggesting more uniform energy consumption patterns. Bungalows, while generally having lower energy use than Detached and Semi-detached, present a distribution that is somewhat wider than Terraced homes but narrower than detached properties. These findings suggest a strong correlation between building archetype (and implicitly, average floor area) and annual energy demand, with larger, standalone properties typically requiring more energy.

The similarity between distributions is because all four archetypes were generated under the same uncertainty space and parameter ranges from identical probabilistic assumptions. As a result, differences in built form alone introduce only second-order effects on annual energy use, leading to overlapping distributions with highly comparable central tendency and spread.

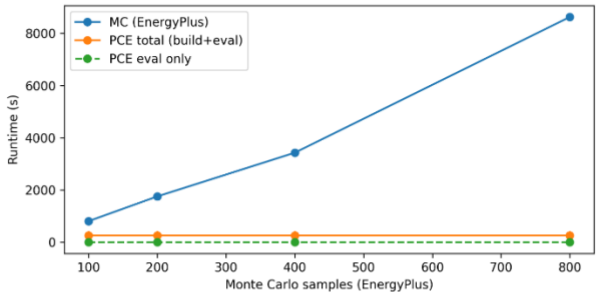

Figure 7. Runtime comparison between direct MCS using EnergyPlus and adaptive PCE surrogate for annual EUI.

Figure 7 compares the computational runtime of direct MCS with that of a PCE surrogate for annual EUI. The MC runtime increases approximately linearly with the number of samples, reflecting the fixed cost of repeated high-fidelity EnergyPlus evaluations and rapidly becoming computationally prohibitive as the number of runs grows. In contrast, the total PCE runtime, comprising surrogate construction and evaluation, remains nearly constant across different numbers of simulations, indicating that the computational cost is dominated by the one-time surrogate construction. Once constructed, the PCE surrogate can be evaluated at negligible cost, enabling large-scale uncertainty propagation and statistical analysis with minimal additional runtime. These results highlight the fundamentally different scaling behaviour of surrogate-based uncertainty quantification and demonstrate the suitability of PCE for applications requiring extensive sampling, repeated analyses, or computationally efficient sensitivity studies. However, the final decision on which method to choose depends on the trade-off between computational budget and required accuracy.

The monthly Sobol' sensitivity analysis shows a strong seasonal pattern in how envelope uncertainties influence energy demand (Table 3). During the heating-dominated months (January–March and September–December), Uwall is the main driver of variability, explaining 47-66% of the monthly energy variance. In contrast, during April–July, Uwin becomes the dominant factor, reflecting the greater role of glazing heat transfer and solar gains in warmer periods. Across all months, first-order and total-order indices are nearly identical (difference < 0.003), indicating that model behaviour is almost entirely additive with minimal interaction effects. This also confirms the strong convergence and stability of the adaptive sparse

PCE surrogate. Overall, the results demonstrate that the dominant envelope parameter changes seasonally, and that the PCE model reliably captures these monthly dynamics for interpreting energy behaviour and guiding retrofit priorities.

Table 3. Top first-order and total Sobol monthly indices.

Month	Top S_1 variable	Top S_1 value	Top S_T variable	Top S_T value
Jan	U_{wall}	0.66	U_{wall}	0.66
Feb	U_{wall}	0.53	U_{wall}	0.53
Mar	U_{wall}	0.47	U_{wall}	0.47
Apr	U_{win}	0.43	U_{win}	0.43
May	U_{win}	0.40	U_{win}	0.40
Jun	U_{win}	0.38	U_{win}	0.38
Jul	U_{win}	0.41	U_{win}	0.41
Aug	U_{wall}	0.40	U_{wall}	0.40
Sep	U_{wall}	0.53	U_{wall}	0.53
Oct	U_{wall}	0.53	U_{wall}	0.53
Nov	U_{wall}	0.57	U_{wall}	0.57
Dec	U_{wall}	0.58	U_{wall}	0.58

Conclusions

This study introduced a unified BIM–BEM uncertainty framework that integrates adaptive PCE with sequential A-optimal sampling. Unlike conventional BIM-based energy analyses that rely on deterministic simulations or Monte Carlo simulations, the proposed framework embeds uncertainty quantification and global sensitivity analysis directly into the modelling workflow. The results demonstrate that adaptive basis refinement combined with informed sampling is sufficient to capture the dominant variance structure of residential building energy models with high fidelity.

The adaptive sparse PCE surrogate converged using approximately 50 EnergyPlus simulations per archetype, achieving RMSE values below 0.8 kWh/m² and an R^2 exceeding 0.99 for annual EUI. In contrast, a Monte Carlo simulation would require several orders of magnitude more model evaluations to reach comparable distributional accuracy. This confirms that the energy response surface induced by envelope uncertainty is smooth and low-dimensional enough to be efficiently approximated by a variance-pruned polynomial basis, validating the core hypothesis of this study. Additionally, analytical Sobol' indices derived from the final PCE coefficients revealed a clear hierarchy of envelope parameter influence. External wall U-value and window U-value jointly explain the majority of annual EUI variance, while monthly analysis exposes a strong seasonal shift in dominance from opaque envelope components in winter to glazing properties in shoulder and summer months.

Future research should extend the uncertainty space beyond envelope properties to include occupant behaviour, infiltration variability, HVAC details, and system efficiencies, where stronger nonlinearities and interaction effects are expected. It is also valuable to check the effectiveness of LHS vs adaptive sampling and normal PCE vs adaptive PCE. Finally, comparing the results with the measured data could increase the credibility and robustness of the model's validation.

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Funding Program Grant No. 22/RDD/822. The opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI.

References

Ali, U. et al. (2024) 'Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach', Energy and Buildings, 303, p. 113768. Available at: https://doi.org/10.1016/J.ENBUILD.2023.113768.

Alwan, Z. et al. (2021) 'Framework for parametric assessment of operational and embodied energy impacts utilising BIM', Journal of Building Engineering, 42, p. 102768. Available at: https://doi.org/10.1016/J.JOBE.2021.102768.

Arbulu, M., Oregi, X. and Etxepare, L. (2024) 'Parametric simulation tool for the enviro-economic evaluation of energy renovation strategies in residential buildings with life cycle thinking: PARARENOVATE-LCT', Energy and Buildings, 312. Available at: https://doi.org/10.1016/j.enbuild.2024.114182.

ASHRAE Standard (2022) 'Standard 90.1–2022: Energy standard for sites and buildings except low-rise residential buildings', American Society of Heating, Refrigerating and Air-conditioning Engineers, Atlanta, GA [Preprint].

Blatman, G. and Sudret, B. (2011) 'Adaptive sparse polynomial chaos expansion based on least angle regression', Journal of Computational Physics, 230(6), pp. 2345–2367. Available at: https://doi.org/10.1016/j.jcp.2010.12.021.

Chadly, A. et al. (2023) 'Techno-economic assessment of energy storage systems in green buildings while considering demand uncertainty', Energy and Buildings, 291, p. 113130. Available at: https://doi.org/10.1016/j.enbuild.2023.113130.

Cheng, K. and Lu, Z. (2018) 'Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression', Computers & Structures, 194, pp. 86–96. Available at: https://doi.org/10.1016/j.compstruc.2017.09.002.

Egan, J. et al. (2018) 'Definition of a useful minimal-set of accurately-specified input data for Building Energy Performance Simulation', Energy and Buildings, 165, pp. 172–183. Available at: https://doi.org/10.1016/j.enbuild.2018.01.012.

ElSayed, M., Shultz, J. and Kurtz, J. (2025) 'User-friendly AI-driven automation for rapid building energy model generation', Energy and Buildings, 345, p. 116092. Available at: https://doi.org/10.1016/j.enbuild.2025.116092.

Energy Performance of Buildings Directive (2025) An official website of the European Union. Available at: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (Accessed: 7 May 2025).

Forth, K. and Borrmann, A. (2024) 'Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM', Journal of Building Engineering, 95, p. 110312. Available at: https://doi.org/10.1016/j.jobe.2024.110312.

Heydari, M.H. and Heravi, G. (2023) 'A BIM-based framework for optimization and assessment of buildings' cost and carbon emissions', Journal of Building Engineering, 79, p. 107762. Available at: https://doi.org/10.1016/J.JOBE.2023.107762.

International Code Council (ICC) (2021) 2021 International Energy Conservation Code (IECC), International Code Council, Inc. Available at: https://codes.iccsafe.org/content/IECC2021P2 (Accessed: 7 October 2024).

Koniorczyk, M. et al. (2022) 'Stochastic energy-demand analyses with random input parameters for the single-family house', Building Simulation, 15(3), pp. 357–371. Available at: https://doi.org/10.1007/s12273-021-0812-9.

Liu, Z. et al. (2020) 'Surrogate modeling based on resampled polynomial chaos expansions', Reliability Engineering & System Safety, 202, p. 107008. Available at: https://doi.org/10.1016/j.ress.2020.107008.

Liu, Z. et al. (2022) 'Impacts of uncertainty in building envelope thermal transmittance on heating/cooling demand in the urban context', Energy and Buildings, 273, p. 112363. Available at: https://doi.org/10.1016/J.ENBUILD.2022.112363.

Prataviera, E. et al. (2022) 'Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis', Applied Energy, 311, p. 118691. Available at: https://doi.org/10.1016/j.apenergy.2022.118691.

Prilliman, M. et al. (2023) Quantifying Uncertainty in PV Energy Estimates Final Report. NREL/TP-7A40-84993, 1961370, MainId:85766, p. NREL/TP-7A40-84993, 1961370, MainId:85766. Available at: https://doi.org/10.2172/1961370.

Shadram, F. et al. (2016) 'An integrated BIM-based framework for minimizing embodied energy during building design', Energy and Buildings, 128, pp. 592–604. Available at: https://doi.org/10.1016/j.enbuild.2016.07.007.

Sood, D. et al. (2025) 'Machine learning based multi-objective optimisation of energy consumption, thermal comfort and CO2 concentration in energy-efficient naturally ventilated residential dwellings', Building and Environment, 267, p. 112255. Available at: https://doi.org/10.1016/j.buildenv.2024.112255.

Sudret, B. (2008) 'Global sensitivity analysis using polynomial chaos expansions', Reliability Engineering & System Safety, 93(7), pp. 964–979. Available at: https://doi.org/10.1016/j.ress.2007.04.002.

Thapa, M. et al. (2024) 'Classifier-based adaptive polynomial chaos expansion for high-dimensional uncertainty quantification', Computer Methods in Applied Mechanics and Engineering, 422, p. 116829. Available at: https://doi.org/10.1016/j.cma.2024.116829.

Thapa, M., Mulani, S.B. and Walters, R.W. (2020) 'Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling', Computer Methods in Applied Mechanics and Engineering, 360, p. 112759. Available at: https://doi.org/10.1016/j.cma.2019.112759.

Tian, W. et al. (2020) 'Sensitivity analysis of building energy performance based on polynomial chaos expansion', Journal of Green Building, 15(4), pp. 173–183. Available at: https://doi.org/10.3992/jgb.15.4.173.

Tian, W. (2024) 'Towards advanced uncertainty and sensitivity analysis of building energy performance using machine learning techniques', Journal of Building Performance Simulation, 17(6), pp. 655–662. Available at: https://doi.org/10.1080/19401493.2024.2387071.

United Nations Environment Programme, G.A. for B. and C. (2024) 'Global Status Report for Buildings and Construction - Beyond foundations: Mainstreaming sustainable solutions to cut emissions from the buildings sector'. Available at: https://wedocs.unep.org/20.500.11822/45095.

Wang, H. et al. (2022) 'BIM-based automated design for HVAC system of office buildings—An experimental study', Building Simulation, 15(7), pp. 1177–1192. Available at: https://doi.org/10.1007/S12273-021-0883-7/METRICS.

Yu, G. et al. (2025) 'A novel global prediction framework for multi-response models in reliability engineering using adaptive sampling and active subspace methods', Computer Methods in Applied Mechanics and Engineering, 433, p. 117506. Available at: https://doi.org/10.1016/j.cma.2024.117506.

Zhao, F., Lee, S.H. and Augenbroe, G. (2016) 'Reconstructing building stock to replicate energy consumption data', Energy and Buildings, 117, pp. 301–312. Available at: https://doi.org/10.1016/j.enbuild.2015.10.001.