Data-Driven Optimization of Building Envelope Geometry for Thermal Comfort and Energy Conservation

Ignazio Ferrari
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
ignazio.ferrari114@unh.edu

Raffaella Alighieri
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
raffaella980@uc.edu

Abstract

Building envelope geometry governs solar exposure, conductive heat loss, daylight availability, and the spatial distribution of thermal comfort, yet early-stage optimization frequently reduces envelope decisions to a small set of catalog options. This manuscript develops a data-driven methodological framework for optimizing envelope geometry under coupled thermal-comfort and energy objectives. The framework parameterizes continuous and discrete geometric variables, maps them to envelope area, glazing distribution, shading, and enclosed volume, and couples a building-performance simulator with a surrogate model that supports multi-objective search. Thermal comfort is represented through predicted mean vote exceedance and long-term discomfort, while energy demand aggregates heating, cooling, lighting, and fan electricity. A constrained Pareto optimization formulation balances energy, comfort, and envelope cost; Gaussian-process surrogates with uncertainty estimates allow adaptive infill and reduce reliance on exhaustive simulation. Robustness is addressed through stochastic weather and occupancy scenarios, and global sensitivity analysis identifies geometry variables that dominate performance variance. The paper specifies validation criteria, failure modes, and decision rules for early design. Its contribution is an integrated, reproducible optimization logic rather than a single optimized form: it clarifies how geometry, comfort constraints, and surrogate uncertainty interact, and it shows why unconstrained energy minimization can degrade thermal comfort or produce fragile designs. The framework is relevant to architects and engineers who need transparent, simulation-grounded evidence for envelope decisions before detailed design is fixed.

Keywords

building envelope geometry; thermal comfort; surrogate modeling; multi-objective optimization; global sensitivity analysis; robust design

1. Introduction

Building envelope geometry is one of the few design decisions that simultaneously conditions solar gain, heat loss, daylight access, ventilation potential, and the radiant environment experienced by occupants. Once massing, orientation, glazing distribution, and shading are fixed, later mechanical and control measures can compensate for poor decisions only at additional cost and energy use. Computational optimization has therefore been applied to sustainable building design for well over a decade, with reviews documenting the promise

and the practical limits of simulation-based search [1], [2], [3]. Recent reviews focused on building geometry and envelope design show that energy performance can be improved through systematic variation of form, orientation, and fenestration, but also that optimization studies often remain bespoke, computationally heavy, and weakly connected to thermal-comfort constraints [4]. The central gap is not the absence of optimization algorithms. It is the absence of a coherent formulation that treats envelope geometry as a mixed continuous–discrete design problem, represents comfort as a time-varying constraint rather than a single nominal condition, and accounts for surrogate error and environmental uncertainty before a design is accepted.

Building performance simulation provides the physical basis for such a formulation [5]. Simulation engines can represent conduction, convection, radiation, infiltration, HVAC operation, and daylight, but each annual or multi-year evaluation is expensive. Direct coupling of a simulation engine to an evolutionary algorithm can require thousands of evaluations, which is rarely compatible with early-stage design. Surrogate modeling addresses this bottleneck by fitting a statistical approximation to simulation outputs and using the approximation to guide search [11], [12]. Yet surrogate-based optimization introduces a second problem: the optimizer may exploit regions where the surrogate is inaccurate. A defensible framework must therefore distinguish between predicted performance and predictive confidence, and it must use that confidence to decide where additional simulations are worthwhile.

This paper develops a data-driven optimization framework for building envelope geometry with explicit thermal-comfort and energy-conservation objectives. The contribution is methodological and integrative. First, it defines a geometry parameterization that connects architectural variables to measurable envelope quantities. Second, it formulates energy demand and thermal-comfort violation as annual performance functionals. Third, it combines a Gaussian-process surrogate with constrained Pareto optimization and adaptive sampling. Fourth, it incorporates robustness to weather and occupancy variation and uses global sensitivity analysis to identify geometry variables that dominate performance variance. The framework does not report a single optimized building; it specifies how optimization evidence should be generated, validated, and interpreted when comfort and energy are optimized together.

2. Problem Formulation

2.1 Geometry and performance variables

The envelope design is represented by a design vector whose components may be continuous, integer, or categorical. Continuous variables include plan aspect ratio, roof pitch, shading-device depth, courtyard width, and facade articulation. Integer or categorical variables include orientation sectors, glazing types, and the number of shading elements. The vector is written as:

where x is a candidate design vector, d is the number of design variables, and X is the feasible design domain defined by site boundaries, floor-area requirements, structural span limits, and architectural program constraints. Categorical variables are encoded through integer indices or periodic transforms; orientation, for example, is represented by sine and cosine components so that 0° and 360° are adjacent rather than artificially distant.

For each x, a geometry generator produces a closed envelope surface and an enclosed volume. These quantities are not merely descriptive: they determine exposed heat-transfer

area and the volume of air that must be conditioned. The envelope area and volume are defined as:

$$A_env(x) = \sum_\{s \in S\} A_s(x),\ V(x) = \int_\{\Omega(x)\} dV$$

where S is the set of envelope surfaces, A_s(x) is the area of surface s, Ω(x) is the enclosed building volume, and V(x) is that volume. The ratio A_env/V provides a first-order indicator of form compactness. It is not a performance metric by itself, because orientation and glazing distribution can reverse the expected effect of compactness, but it helps explain why two designs with similar floor area can have different heating and cooling loads.

2.2 Energy and thermal comfort objectives

Annual energy demand aggregates heating, cooling, lighting, and fan electricity. For a simulation time step t, the annual demand is:

$$\mathrm{E}(x) = \sum_{t=1}^{T} \left[Q_h \frac{(t,x)}{C} OP_h + Q_c \frac{(t,x)}{C} OP_c + P_l(t,x) + P_f(t,x) \right] \Delta t$$

where T is the number of time steps in the annual evaluation, Q_h(t,x) and Q_c(t,x) are heating and cooling loads, COP_h and COP_c are representative system coefficients of performance, P_l(t,x) is lighting power, P_f(t,x) is fan power, and Δt is the time-step duration. The coefficients may be constants or may depend on outdoor temperature and part-load operation; the formulation remains valid in either case. Energy is minimized, but it is not the only objective, because a low-energy design that produces unacceptable thermal conditions does not satisfy the building's purpose.

Thermal comfort is evaluated with the predicted mean vote (PMV) index and its associated predicted percentage dissatisfied. PMV combines air temperature, mean radiant temperature, air velocity, humidity, clothing, and metabolic rate [6]. For long-term evaluation, a single annual mean is inadequate because occupants experience sequences of warm and cool conditions. A comfort-violation functional is therefore defined as:

$$D_T(x) = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \max(0, |PMV(t,x)| - 0.5)^2$$

where PMV(t,x) is the predicted mean vote at time step t and the interval $-0.5 \le \mathrm{PMV} \le 0.5$ represents the central comfort band associated with approximately 10% predicted dissatisfied occupants under steady-state assumptions. The squared exceedance penalizes large deviations more strongly than small ones, which prevents a design from trading many slightly uncomfortable hours for a few severely uncomfortable hours. For naturally ventilated or mixed-mode buildings, adaptive comfort criteria may replace PMV for the relevant seasons [7], [8]; the optimization formulation can accommodate either index by redefining D_T.

2.3 Constrained multi-objective formulation

The design problem has three competing objectives: annual energy demand, long-term thermal discomfort, and envelope cost. Envelope cost is included because otherwise the optimizer has an incentive to add insulation, glazing layers, and shading without limit. The resulting vector optimization is:

$$min_\{x \in X\} F(x) = [E(x), D_T(x), C(x)]\ s.t.\ g_j(x) \le 0$$

where C(x) is a normalized envelope-cost indicator derived from material quantities and glazing areas, and g_j(x) are constraints such as minimum daylight autonomy, maximum window-to-wall ratio on each facade, structural span limits, and maximum construction cost. The constraint set should be defined before optimization and not adjusted after seeing preferred solutions; otherwise the optimization becomes a post hoc justification rather than a design method. Because the objectives have different units, Pareto optimization is used rather than a single weighted sum. A design is Pareto optimal if no other feasible design improves one objective without worsening another. This property allows architects and engineers to inspect trade-offs explicitly instead of accepting an arbitrary weighting chosen by the analyst.

3. Data-Driven Surrogate Modeling

3.1 Design of experiments and simulation data

The term data-driven refers here to the use of simulation-generated data to fit and validate surrogate models. The design space is sampled with a space-filling plan, such as a Latin hypercube or a low-discrepancy sequence, stratified across categorical variables. Each sample is passed to a geometry generator, checked for geometric validity, and evaluated with a building-performance simulator [5]. The output record contains annual energy, comfort-violation components, peak loads, daylight indicators, and material quantities. Before any surrogate is fitted, invalid geometries, failed simulations, and physically implausible outputs are flagged. A surrogate fitted to corrupted simulation data will reproduce the corruption efficiently, so data quality control is part of the modeling method rather than an auxiliary step.

The sampling plan should cover the feasible domain rather than a narrow neighborhood of a preferred design. Early-stage design benefits from global coverage because the geometry–performance relationship is often non-monotonic: increasing window area may improve daylight and winter solar gain while worsening summer overheating and heat loss. Similarly, a deeper overhang may reduce cooling load in one orientation and block beneficial winter radiation in another. Space-filling initial designs provide the information needed to detect such interactions before the optimizer concentrates on a local Pareto front.

3.2 Gaussian-process surrogate

A Gaussian-process surrogate is used because it provides both a prediction and a predictive variance, which is essential for adaptive sampling. For a scalar response y, the posterior mean at an untested design x is:

$$\hat{y}(x) = \mu + \psi(x)^T \Psi^{-1}(y - \mu 1)$$

where μ is the estimated mean, ψ(x) is the vector of covariances between x and the training designs, Ψ is the covariance matrix among training designs, y is the vector of observed responses, and 1 is a vector of ones. The covariance function may be a squared-exponential or Matérn kernel with automatic relevance determination, which estimates a separate length scale for each geometry variable. Automatic relevance determination is useful in this setting because it can reveal that orientation or window-to-wall ratio has a shorter length scale than roof pitch, meaning that the response changes more rapidly along the former variables. The formulation is standard in engineering surrogate modeling [11], [12]; its value here is that the predictive variance can be compared with simulation noise and used to prioritize new simulations.

3.3 Validation and adaptive infill

Surrogate validation should use hold-out designs or cross-validation, with error reported separately for energy, comfort, and cost. A single aggregate error is insufficient because a surrogate may predict annual energy accurately while misclassifying comfort violations. The infill criterion should therefore combine predicted improvement with predictive uncertainty. A candidate design is simulated if it is predicted to lie near or beyond the current Pareto front and if its predictive variance is large relative to the performance difference that matters for the decision. This criterion prevents wasted simulations in well-explored regions and reduces the risk that the optimizer selects a design whose apparent performance is an artifact of surrogate interpolation.

4. Multi-Objective Optimization of Envelope Geometry

4.1 Pareto search with geometric constraints

The surrogate is coupled to a multi-objective evolutionary algorithm, such as NSGA-II, which maintains a population of candidate designs and evolves toward the Pareto front [14]. The algorithm is appropriate because the design variables include discrete choices, the feasible domain is non-convex, and the objective functions are not required to be differentiable. Constraint handling can use dominance with constraint violation or a feasibility-first ranking scheme. The output is a set of non-dominated designs, not a single answer. This set is useful only if it is filtered for geometric constructability and architectural coherence; a mathematically feasible combination of courtyard width, roof pitch, and glazing distribution may be physically valid but spatially incoherent. A geometry-checking module should therefore reject self-intersecting surfaces, unusable floor plates, and facade regions that cannot be constructed with the assumed structural system.

Recent work on generative optimization of building blocks shows that geometric decision variables can be coupled across density, solar access, and structural performance [13]. That line of work reinforces a broader point: envelope geometry is not an isolated aesthetic variable but a carrier of multiple performance regimes. In the present framework, the same logic applies at the building scale, where window-to-wall ratio and shading depth modify both energy and comfort, and where plan aspect ratio modifies both envelope area and daylight distribution. The optimizer must therefore evaluate geometry as a coupled system rather than as a list of independent measures.

4.2 Decision logic and post-optimality analysis

The Pareto front is not self-interpreting. A design that minimizes annual energy may have a high discomfort score because it relies on small windows and limited solar gain. A design that minimizes discomfort may have high energy demand because it uses large glazed areas and active conditioning to manage overheating. The decision rule should be explicit: first, eliminate designs that violate mandatory comfort or daylight constraints; second, inspect the remaining front for knee points where further energy reduction requires disproportionate comfort loss; third, test the sensitivity of shortlisted designs to weather and occupancy assumptions. This sequence preserves the multi-objective character of the problem while acknowledging that not all trade-offs are equally acceptable in practice.

5. Thermal Comfort and Energy Conservation Trade-offs

5.1 Comfort constraints and long-term evaluation

Thermal comfort is often treated as a post-processing check, but in this framework it is an objective and a constraint. The difference matters. As an objective, discomfort is minimized

jointly with energy. As a constraint, a maximum allowable discomfort violation or a maximum percentage of occupied hours outside the comfort band is imposed. The constraint prevents the optimizer from purchasing energy savings with unacceptable comfort degradation. Long-term evaluation should distinguish occupied from unoccupied hours and should use realistic clothing and activity schedules. A design that performs well under a constant metabolic rate may perform poorly when occupants adapt clothing seasonally or when internal gains vary by zone.

The PMV index is most defensible for mechanically conditioned spaces with steady conditions, while adaptive criteria are more appropriate for naturally ventilated spaces [6], [7], [8]. A mixed-mode building may require a hybrid evaluation in which PMV governs air-conditioned periods and adaptive comfort governs free-running periods. The optimization code should record which index is active at each time step so that comfort violations can be attributed to the correct operating regime. Without this attribution, the optimizer may trade one type of discomfort for another without improving the occupant experience.

5.2 Geometry mechanisms and sensitivity

The geometry–energy–comfort relationship is mediated by several mechanisms. Increasing window-to-wall ratio raises daylight availability and winter solar gain but also increases conductive heat loss and summer solar gain; the net effect depends on orientation, glazing properties, and shading. External shading reduces peak solar gain but can block beneficial radiation and daylight, especially in heating-dominated climates. Plan aspect ratio changes the balance between perimeter heat loss and core daylight; a compact form reduces exposed area but may reduce daylight penetration in deep floor plates. Roof pitch and courtyard geometry modify sky view, stack ventilation, and solar exposure. These mechanisms interact, which means that a variable that is beneficial in one geometry may be detrimental in another.

Variance-based sensitivity analysis is appropriate for quantifying these interactions. First-order and total-order indices can identify variables whose variation explains most of the output variance and variables whose influence is mainly through interactions [15]. Building-energy applications require careful treatment of correlated inputs, non-linear responses, and computational cost [16]. In the present framework, sensitivity analysis is used after surrogate fitting to rank geometry variables and to determine whether the sampling plan has covered the influential dimensions. A variable with a small first-order index but a large total-order index should not be discarded, because its effect may be realized only in combination with orientation or glazing ratio.

6. Robustness, Uncertainty, and Implementation

6.1 Robust objective under weather and occupancy uncertainty

Deterministic optimization to a single typical meteorological year can produce designs that are tuned to one climate realization. Weather variability, occupant behavior, internal gains, and control schedules introduce uncertainty. A robust objective is therefore introduced:

$$J(x) = E_\{\theta \sim P\}[E(x,\theta)] + \beta \sqrt{Var_\{\theta \sim P\}[E(x,\theta)]}$$

where θ is a random scenario vector containing weather, occupancy, and operational variables, P is the assumed scenario distribution, E_{θ~P} denotes expectation over that distribution, Var denotes variance, and β is a risk-aversion parameter. A β of zero recovers expected-energy minimization; larger β penalizes designs whose energy performance is

highly variable across scenarios. The same construction can be applied to the comfort-violation functional D_T(x). The scenario distribution should be documented because the robust optimum depends on it. If weather files are limited, a resampling or synthetic-weather procedure can be used, but its assumptions must be reported.

6.2 Global sensitivity and decision support

Global sensitivity analysis closes the loop between optimization and interpretation. The first-order Sobol index for variable x_i is:

$$S_i \; = \; Var_\{x_i\}(E_\{x \sim i\}[y|x_i]) \; / \; Var(y)$$

where E_{x~i}[y|x_i] is the expected value of the response y conditional on x_i, x~i denotes all variables except x_i, and Var(y) is the total output variance. The index S_i measures the fraction of output variance explained by x_i alone. Total-order indices, which include interactions, should be reported alongside first-order indices. Because surrogate predictions are not exact, sensitivity indices should be computed from the validated surrogate and checked against a small set of direct simulations in the influential dimensions. This check prevents over-interpretation of surrogate-derived rankings.

Implementation of the framework requires a geometry generator, a simulation engine, a surrogate library, an optimizer, and a reporting layer. The reporting layer is not cosmetic: it must display the Pareto front, constraint status, uncertainty bands, and sensitivity rankings so that decision-makers can see why a design is recommended. Uncertainty analysis in building performance simulation is most useful when it is connected to design decisions rather than presented as a separate statistical exercise [17]. Early-design decision support also requires that optimization outputs be expressed in terms familiar to architects, such as glazing ratio, shading depth, and plan compactness, rather than only as abstract objective values [18]. Global sensitivity methods provide a defensible way to prioritize which variables deserve further simulation or physical testing [19]. Finally, gaps between optimization research and professional practice persist because tools are often difficult to integrate into design workflows and because results are not transparently tied to design intent [20]. The framework addresses these gaps by making constraints, uncertainty, and trade-offs explicit at the point of decision.

The main failure modes are also clear. A surrogate can be accurate on average but inaccurate near constraint boundaries, which is exactly where feasibility is decided. A Pareto front can be dominated by geometrically invalid or architecturally incoherent designs if constructability checks are weak. A comfort model can be inappropriate for the operating mode, especially in mixed-mode buildings. A robust objective can be misinterpreted as a guarantee rather than a risk-adjusted preference. These are not reasons to abandon optimization; they are reasons to treat optimization as an evidence-generating method that requires validation, transparent assumptions, and disciplined interpretation.

7. Conclusion

This manuscript has developed a data-driven optimization framework for building envelope geometry that treats thermal comfort and energy conservation as coupled objectives rather than sequential checks. The framework parameterizes envelope geometry as a mixed continuous–discrete design vector, relates that vector to envelope area and volume, formulates annual energy demand and long-term comfort violation, and embeds both in a constrained Pareto optimization problem. Gaussian-process surrogates provide predictions and predictive

uncertainty, enabling adaptive infill rather than exhaustive simulation. Robustness is incorporated through scenario-based expected energy and variance penalties, and global sensitivity analysis identifies the geometry variables that most strongly govern performance variance.

The central argument is that envelope optimization is not a search for a single low-energy shape. It is a search for a set of defensible trade-offs among energy, comfort, cost, and robustness. A design that minimizes annual energy but produces unacceptable PMV exceedance is not an optimum in any practical sense. A design that performs well under one weather year but fails under plausible occupancy variation is similarly fragile. The proposed framework makes these distinctions explicit through constraints, Pareto analysis, surrogate validation, and sensitivity reporting. Its practical value lies in the structure it imposes on early-stage decisions: geometry variables are defined, performance mechanisms are represented, uncertainty is acknowledged, and the resulting evidence is presented in a form that can be reviewed by architects, engineers, and clients.

The framework also clarifies where further analytical work is needed. Comfort indices must be matched to operating mode, surrogate error must be checked near constraint boundaries, and constructability filters must be integrated into the optimization loop rather than applied afterward. These requirements follow from the physics and decision context of envelope design, not from a preference for computational complexity. When implemented with validated simulation models and transparent assumptions, data-driven geometry optimization can support envelope decisions that conserve energy without sacrificing the thermal conditions that make buildings usable.

References

1. Evins, R. (2013). A review of computational optimisation methods applied to sustainable building design. Renewable and Sustainable Energy Reviews, 22, 230–245.

2. Nguyen, A. T., Reiter, S., & Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. Applied Energy, 113, 1043–1058.

3. Machairas, V., Tsangrassoulis, A., & Axarli, K. (2014). Algorithms for optimization of building design: A review. Renewable and Sustainable Energy Reviews, 31, 101–112.

4. Kheiri, F. (2018). A review on optimization methods applied in energy-efficient building geometry and envelope design. Renewable and Sustainable Energy Reviews, 92, 897–920.

5. Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. Building and Environment, 43(4), 661–673.

6. Fanger, P. O. (1972). Thermal comfort: Analysis and applications in environmental engineering. McGraw-Hill.

7. Carlucci, S., & Pagliano, L. (2012). A review of indices for the long-term evaluation of the general thermal comfort conditions in buildings. Energy and Buildings, 53, 194–205.

8. Djongyang, N., Tchinda, R., & Njomo, D. (2010). Thermal comfort: A review paper. Renewable and Sustainable Energy Reviews, 14(9), 2626–2640.

9. Caldas, L. G., & Norford, L. K. (2002). A design optimization tool based on a genetic algorithm. Automation in Construction, 11(2), 173–184.

10. Wetter, M., & Wright, J. (2004). A comparison of deterministic and probabilistic optimization algorithms for nonsmooth simulation-based optimization. Building and Environment, 39(8), 989–999.

11. Eisenhower, B., O'Neill, Z., Narayanan, S., & Fonoberov, V. A. (2012). A methodology for meta-model based optimization in building energy models. Energy and Buildings, 47, 292–301.

12. Forrester, A. I. J., Sóbester, A., & Keane, A. J. (2008). Engineering design via surrogate modelling: A practical guide. Wiley.

13. Tabasi, S. F., Luo, D., Rafizadeh, H. R., Le Nguyen, K., & Banihashemi, S. (2025). Generative optimization of building blocks for density, solar and structural performance. Journal of Building Engineering, 111, 113307.

14. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.

15. Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1–3), 271–280.

16. Tian, W. (2013). A review of sensitivity analysis methods in building energy analysis. Renewable and Sustainable Energy Reviews, 20, 411–419.

17. Hopfe, C. J., & Hensen, J. L. M. (2011). Uncertainty analysis in building performance simulation for design support. Energy and Buildings, 43(10), 2798–2805.

18. Østergård, T., Jensen, R. L., & Maagaard, S. E. (2016). Building simulations supporting decision making in early design – A review. Renewable and Sustainable Energy Reviews, 61, 187–201.

19. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). Global sensitivity analysis: The primer. Wiley.

20. Attia, S., Hamdy, M., O'Brien, W., & Carlucci, S. (2013). Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design. Energy and Buildings, 60, 110–124.