AI-Driven Multi-Objective Optimization of Urban Building Morphology for Climate-Responsive Architectural Design

Gaja Harasimiuk
Department of Computer Science, George Mason University, Fairfax, VA, USA.
hellogaja@gmu.edu

Maida Usada
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
contactmaida@oregonstate.edu

Abstract

Urban building morphology is a first-order control on solar access, daylight, ventilation, and operational energy, yet climate-responsive design must reconcile conflicting objectives under site, regulatory, and structural constraints. This manuscript develops an AI-driven multi-objective optimization framework for urban building morphology that treats design as a constrained, simulation-based, many-criteria decision problem rather than a single-score search. The framework encodes footprints, heights, orientation, setbacks, and porosity as mixed design variables; maps morphology and climate-scenario inputs to energy demand, thermal discomfort, daylight availability, and solar potential through probabilistic surrogate models; and couples surrogate inference with Pareto-based evolutionary search and hypervolume-based batch selection. Uncertainty from surrogate error, weather variability, and model assumptions is propagated into robust compromise selection, so that recommended morphologies are evaluated by both predicted performance and predictive variance. The contribution is methodological: it clarifies how AI surrogates, physical performance metrics, and multi-objective decision rules can be joined without reducing urban climate performance to a single objective or treating machine-learning predictions as ground truth. The framework also specifies validation logic, constraint handling, and failure modes, including extrapolation under novel morphologies and smoothing of shadow boundaries. The approach supports early-stage architectural and urban design by producing transparent Pareto sets and uncertainty-aware recommendations, while retaining simulation-based evaluation for final verification. It is relevant to climate-responsive master planning, district-scale retrofit, and performance-based design review.

Keywords

urban building morphology; climate-responsive design; multi-objective optimization; surrogate modeling; hypervolume; uncertainty quantification

1. Introduction

Urban morphology—the arrangement of blocks, streets, heights, setbacks, and open spaces—modifies solar radiation, wind, and heat exchange at the district scale [1]. These modifications are not secondary effects of architectural style. They determine whether streets and courtyards overheat, whether lower floors receive adequate daylight, whether prevailing

breezes can remove heat and pollutants, and how much heating and cooling energy buildings require. Heat-mitigation research has shown that reflective surfaces, vegetation, and urban geometry interact in ways that cannot be captured by isolated building-scale metrics [2]. Urban building energy modeling has likewise established that district-scale energy demand depends on building stock geometry, envelope properties, occupancy, and microclimate, not merely on individual building systems [3].

Computational optimization has therefore become central to performance-driven design [4]. Generative optimization of building blocks has demonstrated how density, solar, and structural performance can be coupled in a single search process [5]. Algorithmic reviews have catalogued evolutionary search, gradient-based methods, and heuristic strategies for building design [6], while simulation-based optimization studies have clarified how simulation cost and objective formulation constrain practical workflows [7]. Multi-objective evolutionary algorithms such as NSGA-II and strength Pareto methods provide a natural language for design trade-offs [8], [9], and many-objective theory extends these ideas to problems with four or more criteria [10]. Yet a persistent gap remains between algorithmic optimization and climate-responsive architectural decision-making. Many workflows optimize a small set of late-stage envelope parameters, rely on deterministic surrogate predictions, or collapse daylight, energy, comfort, and solar potential into a weighted score before the designer can inspect the trade-offs. The resulting optimum may be numerically efficient but architecturally opaque, physically fragile, or unresponsive to regulatory and social constraints.

This paper addresses that gap by developing an integrated methodological framework for AI-driven multi-objective optimization of urban building morphology. The central contribution is a constrained, uncertainty-aware design pipeline that links a morphologically explicit design space to probabilistic performance surrogates and Pareto-based decision analysis. The framework does not treat machine-learning predictions as ground truth. Instead, surrogate predictions are used to accelerate exploration, while uncertainty estimates, physical constraints, and final simulation checks remain part of the decision logic. The argument proceeds from problem formulation to surrogate design, constrained Pareto optimization, climate-responsive metrics, robust decision rules, and technical discussion of trade-offs and failure modes.

2. Problem Formulation and Morphology Representation

The design problem is defined over a district containing N buildings. Each building has a footprint geometry, height, and position relative to streets and open spaces. At the district level, orientation, setback patterns, and porosity govern solar admittance, wind channels, and sky-view exposure. A compact representation is necessary because optimization over raw geometry is computationally unstable. Let p_i denote the parameterized footprint of building i, h_i its height, θ a district orientation variable, and σ a vector of setback or porosity parameters. The design vector belongs to a bounded, mixed-dimensional space.

Here X is the feasible design space, and d is the total number of encoded variables. The footprint parameters p_i may describe polygon vertices, aspect ratios, courtyard openings, or other low-dimensional shape controls. The vector σ may encode street-edge setbacks, passage widths, or ground-floor porosity. This representation is intentionally not a complete geometric model. It is a design grammar that preserves the morphological features most relevant to climate performance while excluding arbitrary detail that would make the search unnecessarily large.

Performance is evaluated against a climate-scenario vector c, which may contain solar irradiance, wind direction and speed, outdoor temperature, humidity, and scenario weights. The objective vector is written in minimization form, with daylight and solar potential negated when they are benefits.

$$F(x;c) = [\mathrm{E}(x;c), T(x;c), -D(x;c), -S(x;c)]$$

E denotes annual operational energy demand, T denotes thermal discomfort such as discomfort degree-hours, D denotes daylight availability, and S denotes usable solar potential. The semicolon indicates that climate conditions are conditioning inputs rather than design variables. This formulation keeps the physical meaning of each objective visible. A designer can inspect whether an apparent improvement arises from energy reduction, daylight loss, overheating, or solar sacrifice.

Feasibility is imposed through site and regulatory constraints. Let g_j(x) be the jth constraint function, including maximum site coverage, floor-area ratio limits, height limits, solar-access rights, fire separation, and minimum daylight or ventilation requirements. Bound constraints ensure that the encoded geometry remains within the design grammar.

The vector x_low contains lower bounds, x_up contains upper bounds, and m is the number of inequality constraints. Some constraints are hard legal limits; others are performance thresholds that may be negotiated. The optimization framework must distinguish these categories. A hard constraint violation cannot be repaired by a compensating performance gain, whereas a soft constraint may enter a feasibility margin or penalty term. This distinction matters because climate-responsive design often involves trading daylight against density, or solar access against street enclosure, and the decision logic should reveal those trades rather than hide them inside an aggregate penalty.

3. AI-Assisted Surrogate Modeling and Climate Performance Inference

High-fidelity evaluation of E, T, D, and S requires coupled simulation of building energy, daylight, solar radiation, and outdoor microclimate. The cost of such simulation makes direct evolutionary search impractical for large morphological populations. The framework therefore uses probabilistic surrogate models to approximate each objective component. Gaussian process regression is appropriate for this role because it provides both a predictive mean and a predictive variance, which can be propagated into robust selection [11]. It is also compatible with active learning, in which new simulation samples are chosen where expected improvement or disagreement is high [12]. Reviews of surrogate modeling in sustainable building design show that the choice of surrogate must reflect the smoothness, dimensionality, and simulation noise of the performance landscape [13].

For each objective k, the true performance function is modeled as a latent function drawn from a Gaussian process conditioned on morphology and climate features.

Here f_k is the kth component of F, m_k is a prior mean function, and κ_k is a covariance kernel. The kernel may combine a linear trend for monotonic effects, a Matérn component for local smoothness, and categorical terms for scenario or building-type effects. The prior mean should encode physically defensible expectations, such as increasing solar obstruction with height or decreasing sky-view factor with street-canyon enclosure. The covariance function controls how rapidly performance is assumed to vary between similar morphologies.

The surrogate is not trained on morphology alone. It must also receive climate and context features, including solar azimuth and altitude ranges, prevailing wind directions, site

latitude, adjacent obstructions, and surface albedo. Feature fusion is necessary because the same height field can produce opposite effects under different solar or wind conditions. A compact kernel can be written as a sum of morphology, climate, and interaction terms, but the important methodological point is that the model must represent interactions rather than treat climate as a constant backdrop.

Daylight and solar availability in urban fabric are especially sensitive to shadow boundaries and sky-view obstruction [14]. A surrogate that is accurate on average can still misrepresent the number of daylight hours in a courtyard or the winter solar gain on a facade. The framework therefore uses validation by spatial holdout, scenario holdout, and error decomposition by objective. Residual analysis should distinguish random simulation noise from systematic errors near geometric thresholds, such as the height at which one building begins to shade another. The surrogate should also be checked for physical plausibility: predicted energy should not decrease arbitrarily with solar exposure in a heating-dominated season, and predicted daylight should not increase when a blocking mass is added without a compensating opening.

Uncertainty is not a byproduct of the surrogate; it is part of the design object. Building performance simulation itself contains uncertainty from weather, occupancy, material properties, and model assumptions [15]. The surrogate adds approximation error. The optimization framework therefore maintains a predictive distribution for each objective and does not reduce the design problem to deterministic point predictions. This choice shapes the search, the Pareto analysis, and the final recommendation.

4. Multi-Objective Optimization and Pareto Decision Analysis

The optimization problem seeks feasible morphologies that approach the Pareto front of F. A design x dominates y when it is no worse in every objective and strictly better in at least one. Pareto dominance provides the formal basis for rejecting aggregate scores that conceal trade-offs.

$$x \prec y \Leftrightarrow f_i(x) \leq f_i(y)\ \forall i \wedge \exists j: f_j(x) < f_j(y)$$

The symbol $\prec$ denotes dominance, i and j index objective components, and the strict inequality in the second clause prevents mutual dominance between identical designs. The set of non-dominated feasible designs forms the Pareto set P. Evolutionary algorithms such as NSGA-II are well suited to approximating P because they maintain diversity along the front and do not require objective weights during search [8]. Strength Pareto approaches provide complementary selection pressure and archive mechanisms [9]. For four objectives, many-objective methods and reference-point strategies are needed to avoid the degeneration of dominance pressure [10].

Constraint handling must be explicit. Feasible designs are preferred to infeasible ones, and among infeasible designs those with smaller normalized constraint violation are retained to preserve search progress. Repair operators can project a candidate back into the feasible set by adjusting height, setback, or footprint variables, but repair must be physically meaningful. A repaired design that satisfies floor-area limits by creating unbuildable slivers is not a valid architectural solution. The optimizer should therefore include a buildability check in addition to numerical constraint satisfaction.

Pareto front quality is assessed with the hypervolume indicator, which measures the dominated objective space relative to a reference point. Hypervolume rewards both convergence toward good performance and diversity along the front.

Λ is the Lebesgue measure, r_i is the reference value for objective i, and m is the number of objectives. A larger hypervolume indicates that the set P dominates a larger region of objective space. The reference point must be chosen consistently across runs and must be worse than the worst acceptable design in every objective. Hypervolume is useful for comparing Pareto sets, but it is not a substitute for design judgment. A high hypervolume can still contain front segments that are architecturally implausible or irrelevant to a specific site.

Because simulation and surrogate evaluation are expensive, candidate selection should be batch-oriented. Expected hypervolume improvement selects designs whose predicted outcomes are likely to enlarge the dominated region after accounting for predictive uncertainty.

$$EHVI(x) = \mathbb{E}[max(0, HV(P \cup \{y\}) - HV(P)) \mid y \sim p(y|x)]$$

denotes expectation, y is a predicted objective vector drawn from the surrogate predictive distribution p(y|x), and P is the current Pareto archive. This criterion balances exploitation of promising regions with exploration of uncertain regions that may contain better trade-offs. Hypervolume-based many-objective search provides a practical algorithmic foundation for this selection logic [16]. The resulting workflow alternates between surrogate-based candidate generation, selected high-fidelity simulation, model updating, and Pareto archive maintenance.

5. Climate-Responsive Performance Metrics and Physical Constraints

The performance vector must remain tied to observable physical mechanisms. Energy demand E should include heating, cooling, lighting, and, where relevant, ventilation and hot-water loads. Thermal discomfort T should be defined through an outdoor or indoor comfort index appropriate to the design question; for outdoor spaces, a rational index such as PET or UTCI can be used, while indoor comfort may be expressed as overheating degree-hours. Daylight D may be represented by daylight autonomy, useful daylight illuminance, or spatial daylight autonomy, depending on the design stage. Solar potential S may refer to roof and facade irradiation or to usable passive-solar gain during specified seasons.

Morphology affects these metrics through several coupled pathways. Height and spacing control shadow length and sky-view factor. Street-canyon aspect ratio influences both solar penetration and longwave exchange. Courtyard geometry affects ventilation and daylight distribution. Surface albedo and vegetation modify radiant load and air temperature. Building-form archetypes can reveal these relationships qualitatively, but district-scale optimization must also account for neighboring obstructions and street orientation [17]. Urban physics research has shown that microclimate, comfort, health, and energy demand are linked through wind, radiation, and heat exchange at multiple scales [18]. The choice of simulation engine and coupling strategy therefore affects the objective values themselves, not merely the computational cost [19].

The constraints should encode climate-responsive minimum performance. Examples include a minimum number of daylight hours for habitable rooms, a maximum allowable overheating risk, preservation of solar access to adjacent buildings, and minimum ventilation opening ratios. These constraints must be expressed in the same spatial and temporal

resolution as the corresponding metrics. A site that satisfies an annual average daylight target may still fail in winter or at lower floors. A district that reduces mean cooling load may still create local outdoor discomfort in a courtyard. The framework should therefore use temporal aggregation rules that preserve seasonal and diurnal distinctions where they matter.

6. Robustness, Uncertainty, and Decision Workflow

A Pareto front is a representation of possible trade-offs, not a final decision. The final recommendation must account for uncertainty in surrogate predictions, climate scenarios, and stakeholder priorities. Robust selection can be formalized by replacing each deterministic objective with a risk-adjusted value and then applying a compromise decision rule over the Pareto set. Let μ_i(x) and σ_i(x) be the surrogate predictive mean and standard deviation for objective i. Let w_i be a nonnegative stakeholder weight, β a risk-aversion parameter, and z_i^I and z_i^N the ideal and nadir values of objective i within the considered Pareto set.

$$x* = argmin_\{x \in P\}\, max_i\, w_i\, (\mu_i(x) + \beta\sigma_i(x) - z_i\text{^}I)/(z_i\text{^}N - z_i\text{^}I)$$

The selected design x minimizes the maximum weighted normalized regret across objectives. This is a robust Tchebycheff compromise rule. The term βσ_i(x) penalizes designs whose predicted performance is favorable but uncertain. The normalization keeps objectives with different units comparable, and the max operator prevents a design from being selected solely because it performs well on one dominant criterion. The weights encode design priorities, but they should not be fixed prematurely. An interactive workflow can present several Pareto subsets under different weightings and let decision-makers inspect the consequences of changing priorities.

The decision workflow has four stages. First, the design space and constraints are formalized with architects, urban planners, and engineers. Second, a training set of morphologies is evaluated with high-fidelity simulation and used to fit the surrogate models. Third, multi-objective search and expected hypervolume improvement generate candidate designs for additional simulation. Fourth, the Pareto archive is analyzed under multiple climate scenarios and risk settings, and selected designs are re-evaluated with the original simulation models. This final re-evaluation is essential because surrogate predictions can be systematically biased near unseen geometric configurations.

Sensitivity analysis should accompany the recommendation. If a design remains optimal across a plausible range of weights, its selection is robust to priority uncertainty. If small changes in β or in the climate scenario reverse the ranking, the decision is fragile and should be presented as such. The framework should also report which constraints are active. An active height limit or solar-access constraint explains why a design cannot improve further, while an inactive constraint suggests remaining design freedom. This information is more useful to designers than a single numerical optimum.

7. Technical Discussion: Trade-offs, Failure Modes, and Implementation Implications

The central trade-off in urban morphology is between density and environmental access. Compact forms can reduce envelope heat loss and land consumption, but they can also block winter solar gain, reduce daylight in lower floors, and trap heat in street canyons. Increasing height can improve solar exposure at upper levels while shading courtyards and neighboring buildings. Increasing porosity can improve ventilation and daylight but may reduce developable floor area or weaken street enclosure. These trade-offs are not resolved by a universal weighting. They depend on climate, building use, occupant vulnerability, and

regulatory context. The optimization framework should make these dependencies explicit rather than assuming that one morphological pattern is universally sustainable.

AI surrogates introduce specific failure modes. Extrapolation is the most serious: a surrogate trained on moderate block heights may predict implausibly low energy demand for a novel high-rise courtyard configuration. Shadow boundaries create discontinuities that smooth kernels can misrepresent. Correlated features can produce spurious associations if the training set lacks morphological diversity. Multi-objective search can exploit these errors by concentrating candidates near surrogate artifacts. Mitigations include physics-informed feature engineering, constrained candidate generation, ensemble disagreement as an additional uncertainty signal, and mandatory high-fidelity re-evaluation of Pareto candidates. Surrogate validation should be reported by objective and by morphological regime, not only as a global average error.

Implementation requires coupling to existing design workflows. Morphological variables should map to building information models, geographic information systems, and planning rules so that optimized designs are buildable and reviewable. Simulation engines should be configured consistently across candidates, with fixed weather files, occupancy assumptions, and material libraries unless those are explicitly part of the uncertainty analysis. The optimizer should also expose the Pareto archive in a form that designers can interrogate: which variables distinguish one front segment from another, which constraints bind, and which objectives improve at the expense of others. Reviews of optimization from an architectural perspective emphasize that computational performance must remain integrated with design intent, constructability, and professional judgment [20]. The same principle applies here. AI-driven optimization can expand the search space and accelerate comparison, but it cannot replace the architectural reasoning that defines what a good district is.

8. Conclusion

This manuscript has developed an AI-driven multi-objective optimization framework for urban building morphology in climate-responsive architectural design. The framework represents morphology as a constrained design vector, evaluates climate performance through energy, thermal discomfort, daylight, and solar-potential objectives, and uses probabilistic surrogates to accelerate evolutionary search. Pareto dominance, hypervolume, and expected hypervolume improvement provide a transparent basis for exploring trade-offs, while a robust compromise rule incorporates predictive uncertainty and stakeholder priorities. The main methodological insight is that AI surrogates should function as uncertainty-aware accelerators within a physically grounded decision process, not as substitutes for simulation or architectural judgment. The framework also identifies technical boundary conditions: shadow discontinuities, extrapolation under novel morphologies, constraint ambiguity, and the sensitivity of decisions to weights and climate scenarios. These conditions are not peripheral. They determine whether an optimized morphology is robust, buildable, and defensible. By combining explicit morphological variables, climate-responsive metrics, constrained Pareto optimization, and uncertainty-aware selection, the approach offers a coherent basis for district-scale design exploration and performance-based review.

References

1. Oke, T. R., Mills, G., Christen, A., & Voogt, J. A. (2017). Urban climates. Cambridge University Press.

2. Santamouris, M. (2014). Cooling the cities – A review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. Solar Energy, 103, 682–703.

3. Reinhart, C. F., & Cerezo, C. (2016). Urban building energy modeling – A review of a nascent field. Building and Environment, 97, 196–202.

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

5. 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.

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

7. 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.

8. 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.

9. Zitzler, E., & Thiele, L. (1999). Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation, 3(4), 257–271.

10. Emmerich, M. T. M., & Deutz, A. H. (2018). A tutorial on multiobjective optimization: Fundamentals and evolutionary methods. Natural Computing, 17(3), 585–609.

11. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

12. Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global Optimization, 13(4), 455–492.

13. Westermann, P., & Evins, R. (2019). Surrogate modelling for sustainable building design – A review. Energy and Buildings, 198, 170–186.

14. Compagnon, R. (2004). Solar and daylight availability in the urban fabric. Energy and Buildings, 36(4), 321–328.

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

16. Bader, J., & Zitzler, E. (2011). HypE: An algorithm for fast hypervolume-based many-objective optimization. Evolutionary Computation, 19(1), 45–76.

17. Ratti, C., Raydan, D., & Steemers, K. (2003). Building form and environmental performance: Archetypes, analysis and an arid climate. Energy and Buildings, 35(1), 49–59.

18. Moonen, P., Defraeye, T., Dorer, V., Blocken, B., & Carmeliet, J. (2012). Urban physics: Effect of the micro-climate on comfort, health and energy demand. Frontiers of Architectural Research, 1(3), 197–228.

19. 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.

20. Shi, X., Tian, Z., Chen, W., Si, B., & Jin, X. (2016). A review on building energy efficient design optimization from the perspective of architects. Renewable and Sustainable Energy Reviews, 65, 872–884.