Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

INFORMING BUILDING RETROFITS USING SURROGATES OF PHYSICS-BASED SIMULATION MODELS: A COMPARISON OF MULTI-OBJECTIVE OPTIMIZATION ALGORITHMS

Seif Qiblawi[1], Elin Markarian[1], Shivram Krishnan[2], Anagha Divakaran[2], Albert Thomas[2], and Elie Azar[1]

[1]Dept. of Civil and Environmental Eng., Carleton University, Ottawa, ON, CANADA
[2]Dept. of Civil Eng., IIT Bombay, Mumbai, MH, INDIA

ABSTRACT

Surrogate models are increasingly used to reduce the computational costs of building performance simulation (BPS) models. However, they are rarely coupled with optimization algorithms to inform retrofitting decisions. The goal of this paper is to compare the performance of several commonly used meta-heuristic algorithms on a machine learning (ML) surrogate of a BPS model to better shape best practices in future work on surrogate model optimization. A surrogate model with an hourly prediction scheme is developed, and retrofits are optimized to minimize annual electric and natural gas loads, as well as PMVSET as a measure of discomfort. This study incorporates exploratory landscape analysis (ELA) to better understand the problem, then separately applies four optimization algorithms (random search, OMOPSO, NSGA-II, AMOSA). NSGA-II is the best-performing algorithm, finding 74% of the final set of Pareto efficient points and converging in 9% of the time required by the second fastest algorithm.

1 INTRODUCTION

Energy consumption in buildings is a large contributing factor to increasing greenhouse gas (GHG) emissions worldwide, with buildings accounting for more than 36% of total energy consumption and carbon emissions (Olu-Ajayi et al. 2023). The pressing need to mitigate the negative effects of climate change has pushed several governments to adopt policy changes with the aim of reducing energy consumption in buildings, such as Canada's aim to achieve net-zero and climate-resilient buildings by 2050 (Natural Resources Canada 2022). A major component of this push is retrofitting existing buildings in order to improve on the existing building stock (National Research Council Canada 2020).

To best guide the retrofitting process, physics-based Building Performance Simulation (BPS) tools are routinely used to model existing buildings to simulate particular retrofits and predict the energy consumption of the retrofitted building. For example, Barbosa and Azar (2018) modeled retrofits using a BPS of a commercial building in the United Arab Emirates (UAE), achieving a 24.4% reduction in energy consumption through simple operation strategies. However, there are some limitations to building design and retrofitting workflows using the mentioned BPS-based approach. Bell (2023) found that while BPS toolkits can be used to bridge industry challenges in predicting energy use, they are most effective when the scope of simulated retrofits is narrow, making it time-intensive to utilize BPS tools for comprehensive investigations of retrofits. Similarly, Lee et al. (2015) surveyed 18 different energy retrofit toolkits, which included 10 physics-based toolkits. They found that while these tools represented high-fidelity models, they required large input data, which is not always available, as well as longer simulation time.

To mitigate the effects of slow simulation times, data-driven surrogate models have been developed to mimic the performance of complex BPS models and leverage their predictive capabilities at reduced computational cost. For example, Ali et al. (2023) trained different machine learning (ML) models to predict annual energy consumption and peak loads using a dataset generated by EnergyPlus, indicating that even basic models (e.g., linear regression (LR)) can attain acceptable levels of accuracy with R^2 values exceeding 0.9 at very low computational times. In another study, Papadopoulos et al. (2018) developed tree-based ML models (i.e., random forests (RF), extra trees (ET) and gradient-boosted regression trees (GBRT) to predict heating and cooling loads based on geometrical features. Their results showed that GBRT outperformed the other models with a shorter training time.

BPS models have also been extensively used with optimization algorithms to help with the decision-making process of retrofits. Hashempour et al. (2020) reviewed a large body of literature on the topic, and found more than 50 papers that dealt with the optimization of building retrofits. Examples of this process include the work by Rosso et al. (2020), who coupled a BPS model of a residential building in Rome with an optimization algorithm and found retrofits that could reduce the annual energy demand by 49.2%.

To date, there has been little work that has coupled the surrogate modelling of buildings with optimization algorithms. Current literature on the subject is limited, with only a few papers detailing this workflow. Studies such as Bamdad et al. (2017) and Bamdad et al. (2020), who used Artificial Neural Network (ANN) surrogates alongside several optimization algorithms, found that ant colony optimization was the most effective algorithm. Older studies by Tresidder et al. (2011) and Tresidder et al. (2012) found that their surrogate could find optimum solutions faster than typical evolutionary algorithms almost 95% of the time. All four studies mentioned used annual metrics for their objective functions, which limits the ability to look at outputs on a more granular scale. This was partially mitigated in the recent study by Zhan et al. (2023), who utilized an ANN-based surrogate of an educational building to output predictions on shorter timescales, finding a 4% improvement in energy consumption and comfort when optimizing for two objectives.

Overall, papers that coupled surrogate modelling with optimization algorithms typically share the limitation that they were designed to handle, at most, two objective functions. For papers that compared different optimization algorithms for a single model, their surrogate models output annual data. Finally, to the best of the author's knowledge, no attempt has been made to test and compare different multi-objective optimization algorithms on surrogate models with sub-annual prediction schemes in currently published scientific literature.

The goal of this paper is to compare the performance of several commonly used meta-heuristic algorithms on an ML surrogate of a BPS model. The study thus aims to inform building retrofits along three performance metrics: electricity consumption, natural gas consumption, and thermal comfort, to better shape best practices in future work on surrogate model optimization. As data-driven methods for simulation become more common in streamlining and accelerating scientific efforts, a study such as the one to be undertaken here is helpful to gauge the performance of different algorithms on similar black-box optimization problems.

This paper is organized as follows: Section 2 outlines the methodology followed. Section 3 details the results of the proposed methods, and Section 4 discusses the limitations of this study and concludes the work.

2 METHODOLOGY

The methodology of this paper consists of three main phases, shown in Figure 1 and detailed in the following subsections. Phase 1 consists of creating the ML surrogate model based on a validated BPS model of an archetypal office building, Phase 2 details the selection of optimization algorithms to test as well as the basis of comparison to be used, and Phase 3 consists of the implementation and comparison of the selected optimization algorithms' capabilities to inform building retrofit measures.

2.1 ML Surrogate Model Development

A previously developed BPS-based dataset is used to train the surrogate model (Markarian et al. 2024). The BPS model is representative of an archetype three-story office building that was compliant with the National

Energy Code of Canada for Buildings (NECB). Detailed information on the building's characteristics can be found in Kharvari et al. (2022).

The dataset was generated by varying key input parameters (e.g., heating and cooling setpoints and window U-value) and estimating hourly electricity and natural gas consumption as well as Predictive Mean Vote at Standard Effective Temperature (PMVSET). PMVSET is a comfort metric from the Pierce two-node model that simplifies the human body into two isothermal compartments: the core, where all metabolic heat is generated, and the skin (Chen et al. 2017). The model additionally standardizes the actual environment to a hypothetical one at SET, defined as the dry-bulb temperature in an imaginary setting with 50% relative humidity and appropriate clothing for occupant activity. PMVSET is obtained from $PMVSET = (0.303e^{-0.036M} + 0.028)(H - L_{SET^*})$, where M, H and L_{SET*} represent metabolic rate, occupant's internal heat production rate and energy loss from body at SET, respectively. PMVSET within the range of -1 and 1 indicates occupants' comfort (Chen et al. 2017).

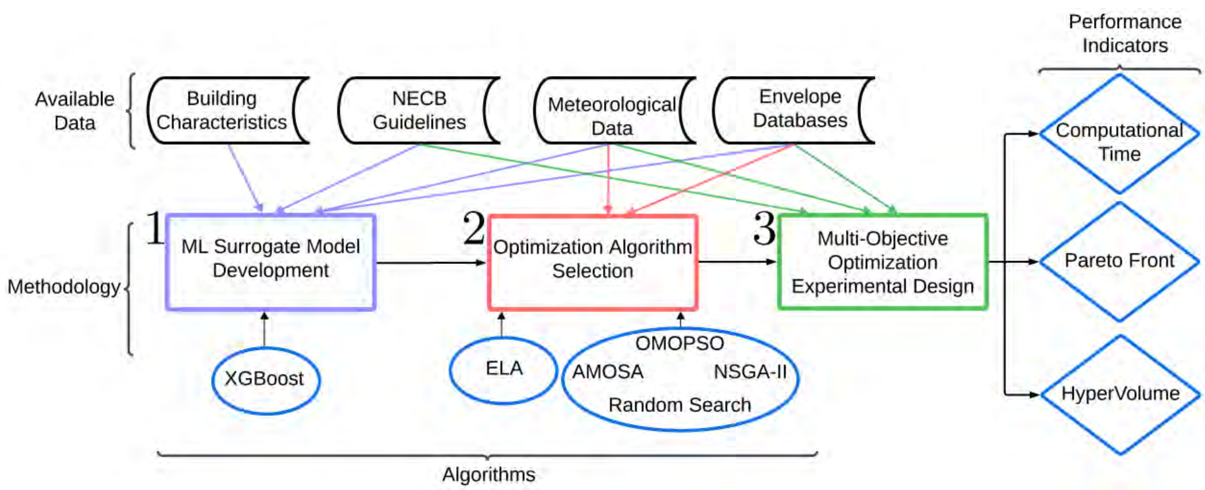

Figure 1: Methodology flowchart.

The list and the range of parameters to vary were determined based on previous literature (Hashempour et al. 2020). A Latin Hypercube Sampling (LHS) method was employed to draw samples from the specified ranges, which were then fed into EnergyPlus to generate hourly results. Overall, the dataset has 11 input features, including both weather and building-related features and three target variables, with a total of 876,000 instances (Markarian et al. 2024).

Extreme Gradient Booting (XGB) is selected to develop the ML surrogate model due to its high level of accuracy in the building energy performance context (Ali et al. 2023; Sánchez-Zabala and Gómez-Acebo 2024; Sun et al. 2020). The Multioutput Regressor from the scikit-learn library is utilized with XGB from the xgboost library as its estimator, allowing for the simultaneous prediction of the three target variables. The model is trained on 80% of the dataset selected randomly and tested on the remaining 20%. To further improve the model performance, a two-step hyperparameter tuning involving "HalvingRandomSearchCV" and "RandomizedSearchCV" is conducted. "HalvingRandomSearchCV" uses a successive halving technique which progressively allocates more resources to promising configurations while discarding poor-performing ones (Soper 2023). This technique is executed multiple times to find the approximate location of the hyperparameters and narrow down the search space, as it is computationally less intensive than a random grid search method. A random grid search is then conducted to find the optimal hyperparameters. Finally, the predictive accuracy is evaluated using adjusted R^2, mean absolute error (MAE) and coefficient of variation of root mean square error (CV(RMSE)). For more details on input selection, sampling, and surrogate model development, please refer to Markarian et al. (2024).

2.2 Optimization Algorithm Selection

Algorithm selection is key to optimization. In general, the fitness of an algorithm is highly dependent on the optimization problem, with Table 1 detailing the families of algorithms as classified by Muñoz et al. (2015). A review done by Nguyen et al. (2014) on optimization of BPS models found that most researchers utilized stochastic algorithms to avoid problems with finding gradients of functions that may not be differentiable across their domain, which are potential challenges for black-box models in general.

Table 1: Algorithm classification.

Class	Family	Common algorithms
Deterministic	Line search methods	Gradient descent
	Trust region methods	Levenberg-Marquardt
	Pattern search methods	Nelder-Mead
Stochastic	Random search methods	N/A
	Simulated annealing methods	AMOSA
	Population based algorithms	PSO, GA

One method of aiding algorithm selection for black-box models that has become increasingly prevalent in recent years is exploratory landscape analysis (ELA). ELA is an umbrella term for calculations that can help gauge a deeper understanding of a black-box model (Malan 2021; Mersmann et al. 2011). While the problem being tackled is multi-objective in nature, most literature on ELA is restricted to single-objective problems. Some work — see Seiler et al. (2024); Liefooghe et al. (2020) — has recently begun to tackle the usage of ELA for multi-objective problems, but the field still lacks maturity. For this reason, ELA is used here to analyze each objective function as a standalone problem, to aid with algorithm selection.

While many ELA techniques exist, the one utilized for this paper is finding local optima networks (LON) proposed by Adair et al. (2019). This technique gives practitioners the ability to gauge an estimate of the number of local optima and study their distribution across the function landscape. Optima that are close together in space and separated from other clusters of optima are known as funnels. LON analysis also gives an estimated number of funnels within the function. Formally, if we denote $X \in \mathbb{R}^n$ as the set of all possible solutions to the problem, N as the neighborhood of solutions around a single solution x, and $f(x)$ as the fitness function at x, then a local optimum x_* is a solution such that $\forall x \in N(x_*), f(x_*) \leq f(x)$. The method of finding these clusters mirrors the logic of basin-hopping, first introduced by Wales and Doye (1997), which translates the landscape of a function into points of different potential energy, in which the algorithm is tasked with finding the highest potential energy (which indicates to the global minima). For this study, the implementation of LON on the Python package pflacco is utilized (Prager and Trautmann 2024). The algorithm is run in order to understand the modality of each objective function. To do this, the decision variables are sampled using an LHS scheme before being passed to the main algorithm.

2.2.1 Algorithms Utilized

Following the classification presented in Table 1, one algorithm from each family of stochastic algorithms was chosen, apart from population-based algorithms; seeing as they are the most popular for similar applications (see (Mousavi et al., 2023)), two algorithms were chosen to represent this family. Each of the four was given an upper limit of 300 possible undominated solutions to find. The first is random search, which is considered in order to judge how well other algorithms perform with respect to some baseline. The second algorithm is an implementation of multi-objective particle swarm optimization algorithms (MOPSO) commonly called OMOPSO, first formulated by Sierra and Coello Coello (2005). OMOPSO is a population based algorithm that simulates particles moving in a swarm, and utilizes a crowding factor in which leading candidate solutions are given a crowding value to represent the distance between it and other candidate solutions. Candidates with fewer crowding neighbors are favored in selection and stored in an archive, promoting diversity of solutions. OMOPSO also utilizes multiple fitness values for each objective, making it well suited for multi-objective problems. Both of the preceding algorithms are implemented through jMetalPy, a Python library created by Benitez-Hidalgo et al. (2019).

The third algorithm implemented is non-dominated sorting genetic algorithm II (NSGA-II), which is a widely used algorithm for optimization problems with 2 or 3 objective functions (Ma et al. 2023). Originally proposed by Deb et al. (2002), NSGA-II creates a parent population and generates an offspring population from the parent, mimicking natural selection. The resultant of these two populations is combined and sorted according to dominance. This is repeated until termination. NSGA-II was implemented through the Python package Pymoo (Blank and Deb 2020b). A stricter convergence than the one originally proposed in Blank and Deb (2020a) was implemented, with a tolerance of the design space set to 10^{-10} and a tolerance of the objective space set to 0.000025. For both OMOPSO and NSGA-II, the mutation parameters (uniform and nonuniform for the former, and polynomial for the latter) were set to a probability of $1/11$ as per the comparative study conducted by Godínez et al. (2010).

The final algorithm implemented was archived multi-objective simulated annealing (AMOSA), first proposed by Bandyopadhyay et al. (2008). This algorithm is similar to basin-hopping (discussed in subsection 2.2) but extends to multi-objective problems. The inspiration for simulated annealing algorithms comes from statistical mechanics, whereby the temperature of matter is increased then slowly decreased to a minimum, where the low-energy states of the matter represent the global minima. For multi-objective problems, this concept is expanded to contain an archive of solutions representing the Pareto front of the problem. This algorithm was implemented using the pyAMOSA package (Barone 2021). An initial temperature of 200 degrees was taken, with the final (termination) temperature taken to be 10^{-7}, as per the original paper's recommendation.

2.2.2 Performance Indicators

Correctly benchmarking the suitability of algorithms requires some measure of performance. Over the years, many different indicators have come to be used, including simple measures such as time and iterations and more sophisticated measures dealing with convergence to a known Pareto front (Audet et al. 2021). The indicators utilized in this study are the number of Pareto efficient solutions found, the hypervolume of the found solutions, as well as the computational time required for each algorithm.

On the first of the mentioned metrics, it should be noted that the number of solutions found is not an indication of the quality of said solutions. This is instead better represented by the number of Pareto efficient solutions found by each algorithm when compared to the union of all four solution sets. Formally, a solution x^* is Pareto efficient in set X if there is no solution x' that dominates it, i.e. if x^* is not worse than x' in all objectives and is strictly better in one. Therefore, the set P of Pareto optimal solutions contains all non-dominated solutions, i.e. solutions in which there exists no other solution that is simultaneously more efficient in all three objectives.

The second metric is hypervolume, which is a widely used metric as it does not require existing knowledge of the Pareto front. A reference point of $(r_i, \dots, r_n)$ is given by the user such that this point falls outside the set of possible solutions. Assuming normalized variables, $r_i > 1$, the volume bounded by the found Pareto front and the reference point is called the hypervolume of the solution, and the aim of the optimization is to maximize the hypervolume (Cao et al. 2015). Given that for the implementation of OMOPSO and random search, there was no other suitable termination criterion, the termination of these two algorithms was done either after a predetermined runtime or if a hypervolume of 0.9 was achieved. The reference point for hypervolume is taken to be (r, r, r), with $r = 1 + 1/23$, which is suitable given the population size of 300 chosen, according to the recommendation given in Ishibuchi et al. (2017). For the termination criterion, the variables are normalized as $x/(1 + x)$ due to incomplete information about the maximum and minimum of each range. For the final hypervolume calculation, maximum and minimum values are extracted from the solution sets to be used, and hypervolume is re-calculated for all four solution sets using Pymoo's hypervolume function to ensure fairness of comparison. In this algorithm, solutions are stored in a list L according to their z coordinate (assuming three objective functions), and projects each solution onto the (x, y) plane relative to the provided reference point. This area is multiplied by the difference in z between the current solution and the one stored below it in L. This is repeated for all solutions in L to provide the hypervolume of the found Pareto front, as detailed in Fonseca et al. (2006).

The final metric, computational time, becomes increasingly important as problems scale up, either in terms of decision variables or number of objectives. It is used here in conjunction with the number of Pareto efficient points to gauge whether the algorithm found a wide array of efficient solutions in an acceptable time. To ensure fairness, all algorithms are run on the same computer, equipped with an AMD Ryzen 7 5800HS processor. All algorithms that do not terminate within 11 hours are manually terminated at the 11-hour mark. Seeing as computational efficiency is a large motivator behind the usage of surrogate models, the need to get results in a reasonably short timeframe outweighs the importance of finding the true Pareto front, especially if some algorithms terminate before 11 hours and perform well in other metrics.

2.3 Multi-Objective Optimization Experimental Design

The optimization problem at hand contains 11 decision variables, with Table 2 listing the variables and the reference material used to generate realistic input values. Unlike the ML training and testing phase in Section 2.1, the design variables are bounded by tighter upper bounds to ensure code compliancy. The optimization utilizes hourly weather conditions to simulate the performance of the building under typical weather conditions in Ottawa over one year. For some variables, such as infiltration, it is difficult to quantify possible inputs, therefore the minimum was taken to be approximately half of the initial condition. These variables represent possible envelope retrofits as well as operational changes, to best capture different retrofit scenarios that practitioners could undertake to improve building performance.

Table 2: Optimization search space.

Variable	**Search Space**	**Variable (cont.)**	**Search space**
X1: Wall R-value (m^2K/W)	[4.17,10] (Morrison Hershfield 2021)	**X7: Cooling Setpoint (6:00-21:00)**	[22,28] (Azar and Menassa 2014; Bamdad et al. 2017; Shen and Pan 2023)
X2: Roof R-value (m^2K/W)	[7.69,10] (Morrison Hershfield 2021)	**X8: Cooling Setpoint (21:00-6:00)**	[25,35] (Azar and Menassa 2014)
X3: Roof U-value (W/m^2K)	[0.79,1.69] (Love and Bozoian 2022; National Fenestration Rating Council 2023)	**X9: Chiller COP**	[4.5,5.3] (Love and Bozoian 2022)
X4: SHGC	[0.79,1.69] (Love and Bozoian 2022; National Fenestration Rating Council 2023)	**X10: Heating Efficiency**	[0.72,0.95] (Love and Bozoian, 2022)
X5: Heating Setpoint (21:00-6:00)	[16,20] (Talami et al. 2023)	**X11: Infiltration ($m^3/s/m^2$)**	[0.000125,0.0003]
X6: Heating Setpoint (7:00-21:00)	[18,24] (Bamdad et al. 2017; Shen and Pan 2023)		

The optimization to be performed is multi-objective in nature. For this study, three objective functions are minimized, (i) annual electric consumption (ii), annual natural gas consumption, and (iii) a discomfort metric. The first two objectives are found by summing the hourly outputs of the model, and the third is found by outputting the absolute value of PMVSET at every hour, and calculating the mean across all occupied hours of the year. Due to PMVSET being calculated at an hourly scale, the sub-annual prediction scheme detailed in Subsection 2.1 is deemed necessary.

3 RESULTS AND DISCUSSION

Results are given in two Subsections. The first details the results of the training and testing of the ML model, and the second gives the results of the optimization.

3.1 ML Model Testing

Table 3 demonstrates the predictive performance of the XGB model developed using the testing portion (20%) of the dataset described in Subsection 2.1. The results confirm the predictive performance of the developed model with R^2 values exceeding 0.9 for all three target variables. Additionally, CV(RMSE) values of the target variables are well below the 30% threshold proposed by the ASHRAE 14 guideline for hourly models (ASHRAE 2023).

Table 3: Developed XGB model predictive performance.

Metrics	**Electricity**	**Natural gas**	**PMVSET**
Adjusted R^2	0.97	0.97	0.92
MAE	4.9 kWh	7.56 kWh	0.1
CV(RMSE)	0.13	0.21	0.22

The actual and predicted values of the three target variables are demonstrated in Figure 2. The data points form a relatively even dispersion along the diagonal line, indicating a strong positive correlation between actual and predicted values. The model overestimates the PMVSET for some instances, falling between 0 and 0.7.. Analysis of the input data indicated that the overestimation was primarily due to one specific EnergyPlus run, which acted as an outlier. This, however, does not undermine the overall performance, as confirmed by the performance metrics listed in Table 3.

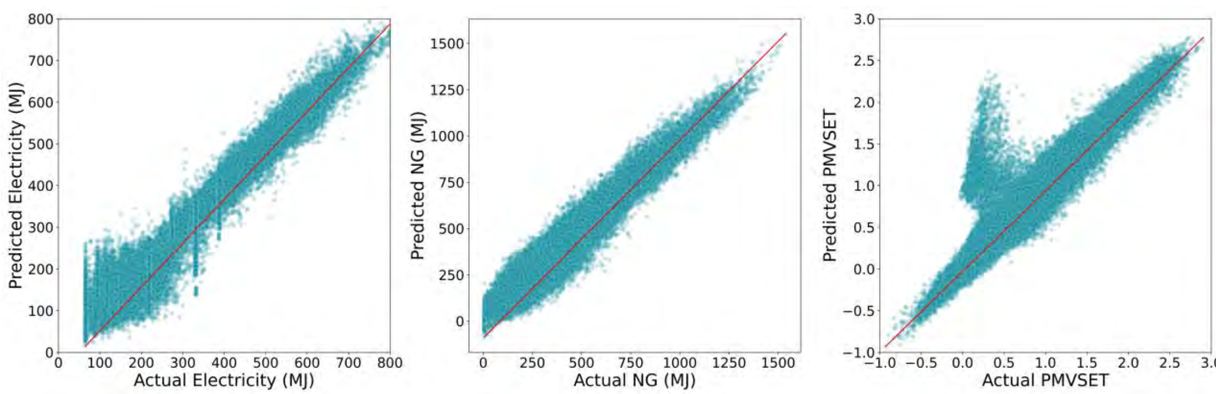

Figure 2: Predicted vs. actual values for each objective. Each point represents one hourly value of electricity (left), natural gas (center), and PMVSET (right).

3.2 Optimization Results

The set of Pareto efficient points across the union of the sets are displayed in Figure 3, showcasing first each possible combination of objective functions in the lower triangle, as well as the distribution of the solutions found along the diagonal, with tags specifying which algorithm corresponds to its distribution.

Different colors and markers are used to represent the algorithm that found its respective solutions. The three scatterplots clearly showcase the conflicting relationship between energy consumption and comfort (lowest row), but the relationship between electricity and natural gas consumption (upper left scatter plot) shows no clear correlation. This can be explained by recognizing that while envelope changes would benefit both objectives, other variables (especially chiller COP and heating efficiency) would have little to no effect on one objective while strongly affecting the other. NSGA-II outperformed the other algorithms in terms of number of efficient solutions found; of the 300 solutions found by the NSGA-II algorithm, 97% were undominated by any other solution found across the four solution sets and constituted 74% of the final set of Pareto efficient points. The distribution of solutions displayed on the diagonal of Figure 3 also shows that NSGA-II was consistently successful in finding low values for all three objectives.

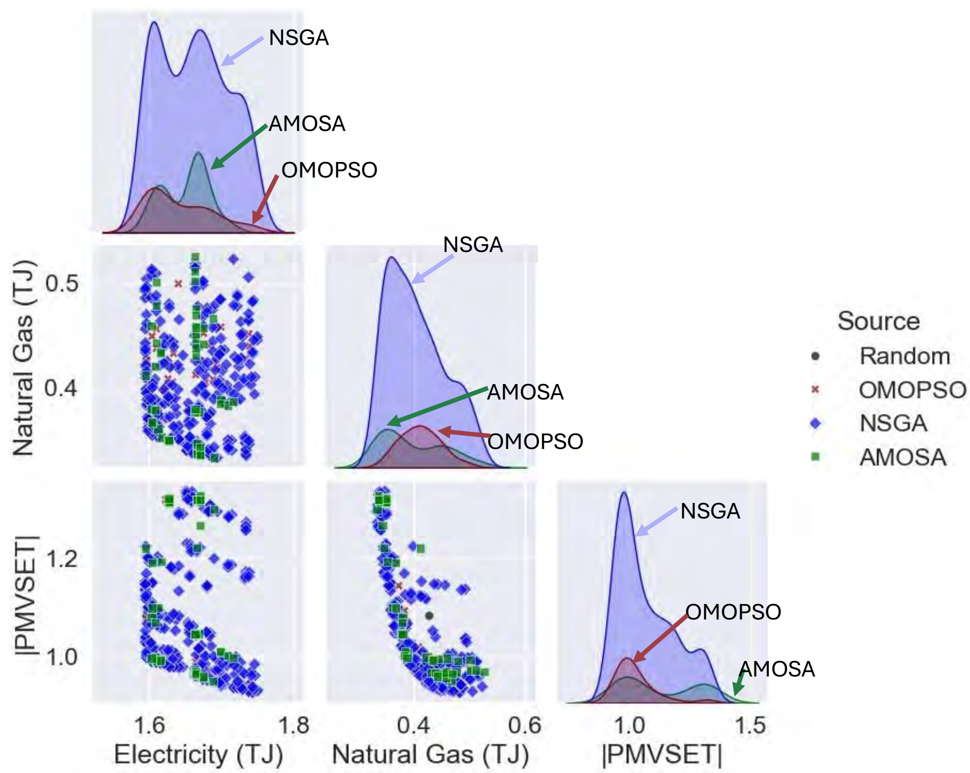

Figure 3: Set of all non-dominated solutions, alongside the distribution of solutions found.

To further understand each algorithm's ability to capture the optimal solution, Table 4 details the optimal solution found by each algorithm for each of the three objectives — annual electricity consumption, annual natural gas consumption, and |PMVSET| respectively. The best-performing solution for each metric is indicated in **bold**. Also listed is the hypervolume of each solution set, as well the number of solutions found out of a predetermined maximum of 300. In this comparison, NSGA-II also found better optimal solutions for each of the three objective functions, and its solution space covered a larger hypervolume.

It is prudent to mention here the results of the LON analysis (detailed in Subsection 2.2). The LON analysis of the objective functions found 100 funnels in each of the objective functions, as well as 1532, 1486, and 1719 local optima in the three objective functions as defined in Subsection 2.3. The high multimodality of the objective functions is further validation of the need to use stochastic optimization algorithms instead of deterministic ones, as the usage of algorithms dependent on derivatives would not be able to traverse the objective space effectively. This density of local optima is showcased through Table 4 as each of the algorithms found a minimum annual electric consumption of ~1.6 TJ, a minimum annual natural gas load of ~0.34 TJ, and a minimum |PMVSET| of ~0.93. The results showcase the many scenarios in which numbers near the absolute minimum could be achieved, and highlights the difficulty of the objective space being traversed. The difficulty in finding good solutions due to the large number of local minima for each function is further evinced by the fact that of the 797 solutions found, only about half were truly non-dominated in the union of the four individual sets. There is also little doubt that there are more non-dominated solutions that were not found by any of the algorithms. NSGA-II handled this complex landscape best in terms of the quality of solutions found.

Table 4: Optimal solutions found by each algorithm, and the hypervolume of their Pareto fronts.

Algorithm	n_{found}	n_{nds}	**Optimal Obj-1**	**Optimal Obj-2**	**Optimal Obj-3**	**HV**	**Time (hr)**
Random Search	224	1	(1.60,0.42,1.17)	(1.72,0.35,1.29)	(1.75,0.51,0.93)	0.64	11
OMOPSO	81	46	(1.60,0.43,1.00)	(1.69,0.34,1.30)	(1.74,0.45,0.93)	0.70	11
NSGA-II	300	**290**	(**1.59**,0.49,1.14)	(1.69,**0.33**,1.29)	(1.74,0.48,**0.93**)	**0.73**	**0.49**
AMOSA	192	55	(1.60,0.41,1.22)	(1.69,0.33,1.30)	(1.74,0.52,0.93)	0.70	5.82

As for the optimization computational cost, the four algorithms were run either until termination, or up to 11 hours provided that the algorithms did not terminate. Of the four, only NSGA-II and AMOSA terminated before the 11-hour mark, at a runtime of 30 minutes and 5.82 hours respectively. The stark contrast in time required to find each solution set also diminishes the possibility that the difference in solution quality was related to small computational inefficiencies in the implementation of each of these algorithms. While the optimal solution quality and hypervolumes do not vary greatly amongst the four solution sets, the time in which NSGA-II achieved the best-performing solution sets its performance apart from the other algorithms tested in this study, having found its superior solution set almost 12 times faster than AMOSA and 22 times faster than OMOPSO and random search.

4 CONCLUSION

Recent studies on building surrogate models with sub-annual prediction schemes do not usually attempt to compare the performance of different optimization algorithms. This paper presented a unique comparison of four optimization algorithms applied to surrogate models of BPS to test retrofit strategies. The presented approach consisted of three main phases: (1) developing an ML surrogate of a validated archetypal BPS office model; (2) detailing the process of algorithm selection and the utilization of ELA techniques to gain a deeper understanding of the black-box model, and (3) conducting a multi-objective optimization to find retrofitting solutions that minimize electricity and natural gas consumption, as well as PMVSET. This study has studied the effectiveness of four different algorithms on such a model to showcase the difficulty of optimizing highly multimodal objective landscapes.

Having developed an accurate ML surrogate of a validated BPS archetype office building, the surrogate was studied using LON within the ELA framework to gauge the modality of the objective functions. This highlighted the need to use stochastic algorithms to efficiently find solutions, and four commonly used algorithms (random search, OMOPSO, NSGA-II, AMOSA) were utilized to compare their performance using three performance indicators, namely the quality of solutions (through comparing the four solution sets and sorting out dominated solutions, as well as judging the best found solution for each objective), the hypervolume of each solution set, and the computational time required to find their respective solution sets.

Of the four algorithms, NSGA-II was found to be the best performer, having yielded the best results for every objective function, as well as finding the largest number of Pareto efficient solutions. Coupled with its superior hypervolume performance, as well as the speed in which these solutions were found, NSGA-II is the most suitable algorithm for such a problem from the ones tested. The efficient solution set found by NSGA-II constituted 74% of the final set of Pareto efficient solutions, and this large volume of efficient solutions was found at a mere fraction of the time needed for other algorithms to find inferior solutions, with NSGA-II finding its solution set in 9% of the time when compared to AMOSA, and 5% when compared to random search and OMOPSO.

The algorithms surveyed in the current paper are commonplace, but new algorithms are being developed regularly that may outperform the algorithms selected here. Due to the generic nature of the proposed methodology, this framework can easily be adapted to test algorithms different from the four selected here. Additionally, the inclusion in ELA is likely novel within the field of building surrogates, and the widespread nature of the algorithms used also gives practitioners an idea of which algorithms could perform best for their respective problems. The ability to efficiently optimize for building surrogates (with the best solution set being found in a mere 30 minutes) further lends credence to the idea that future retrofitting problems can quickly find solutions to complicated designs. This allows for the incorporation of more decision variables and objective functions, thus bypassing one of the main bottlenecks faced when using typical BPS-in-the-loop optimization schemes.

Finally, it is worth noting that a stronger understanding of the problem being optimized is highly recommended for algorithm selection. Tools such as ELA, while not widespread in the field of building surrogates, could be useful to gauge a deeper understanding of the black-box model being utilized. Future work could include more tools implemented from ELA and scaling the comparison to many-objective problems to better inform best practices in the field of surrogate model optimization, with the aim of informing building retrofits. Sub-annual prediction schemes could also be further leveraged for metrics pertaining to Net-Zero Energy Building (NZEB) design, such as renewable energy modeling.

ACKNOWLEDGMENTS

The authors acknowledge the funding support from IC-IMPACTS (India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability) and DST (Department of Science and Technology), Government of India, grant number DST/IC/IC-IMPACTS/2022/P-7, as well as the Canada Research Chairs Program. The authors also thank Adam Wills from the National Research Council Canada for providing reports that helped improve this work.

REFERENCES

Adair, J., Ochoa, G., and Malan, K. M. 2019. "Local Optima Networks for Continuous Fitness Landscapes". In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, edited by M. López-Ibáñez, A. Auger, and T. Stützle 1407–1414. New York: Association for Computing Machinery.

Ali, A., Jayaraman, R., Mayyas, A., Alaifan, B., and Azar, E. 2023. "Machine Learning as a Surrogate to Building Performance Simulation: Predicting Energy Consumption Under Different Operational Settings". *Energy and Buildings 286*:112940.

ASHRAE. 2023. "Measurement of Energy, Demand, and Water Savings". Guideline 14, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Audet, C., Bigeon, J., Cartier, D., Le Digabel, S., and Salomon, L. 2021. "Performance Indicators in Multiobjective Optimization". *European Journal of Operational Research 292*(2):397–422.

Azar, E. and Menassa, C. C. 2014. "A Comprehensive Framework to Quantify Energy Savings potential From Improved Operations of Commercial Building Stocks". *Energy Policy 67*:459–472.

Bamdad, K., Cholette, M., Bell, J., and Guan, L. 2017. "Building Energy Optimisation Using Artificial Neural Network and Ant Colony Optimisation". In *Proceedings of the Australasian Building Simulation Conference 2017*, November 15th-16th, Melbourne, Australia.

Bamdad, K., Cholette, M. E., and Bell, J. 2020. "Building Energy Optimization Using Surrogate Model and Active Sampling". *Journal of Building Performance Simulation 13*(6):760–776.

Bandyopadhyay, S., Saha, S., Maulik, U., and Deb, K. 2008. "A Simulated Annealing-Based Multiobjective Optimization Algorithm: AMOSA". *IEEE Transactions on Evolutionary Computation 12*(3):269–283.

Barbosa, J. D. and Azar, E. 2018. "Modeling and Implementing Human-based Energy Retrofits in a Green Building in Desert Climate". *Energy and Buildings 173*:71–80.

Barone, S. 2021. SalvatoreBarone/pyAMOSA. https://github.com/SalvatoreBarone/pyAMOSA, accessed March 9th 2024.

Bell, N. 2023. The Path to Zero Energy Office Buildings – Energy Retrofitting of Commercial Office Buildings. Ph.D. thesis, School of Photovoltaic and Renewable Energy Engineering, UNSW Sydney, Kensington, Australia. http://hdl.handle.net/1959.4/101322, accessed 20th January 2024.

Benitez-Hidalgo, A., Nebro, A. J., Garcia-Nieto, J., Oregi, I., and Del Ser, J. 2019. "jMetalPy: A Python Framework for Multi-Objective Optimization with Metaheuristics". *arXiv preprint arXiv:1903.02915*.

Blank, J. and Deb, K. 2020a. "A Running Performance Metric and Termination Criterion for Evaluating Evolutionary Multi- and Many-objective Optimization Algorithms". In *2020 IEEE Congress on Evolutionary Computation (CEC)*, July 19th-24th, Glasgow, United Kingdom.

Blank, J. and Deb, K. 2020b. "Pymoo: Multi-Objective Optimization in Python". *IEEE Access 8*:89497–89509.

Cao, Y., Smucker, B. J., and Robinson, T. J. 2015. "On Using the Hypervolume Indicator to Compare Pareto Fronts: Applications to Multi-Criteria Optimal Experimental Design". *Journal of Statistical Planning and Inference 160*:60–74.

Chen, X., Yang, H., and Wang, Y. 2017. "Parametric Study of Passive Design Strategies for High-Rise Residential Buildings in Hot and Humid Climates: Miscellaneous Impact Factors". *Renewable and Sustainable Energy Reviews 69*:442–460.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. 2002. "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II". *IEEE Transactions on Evolutionary Computation 6*(2):182–197.

Fonseca, C. M., Paquete, L., and Lopez-Ibanez, M. 2006. "An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator". In *2006 IEEE International Conference on Evolutionary Computation*, 16th-21st July, Vancouver, Canada, 1157-1163.

Godínez, A. C., Espinosa, L. E. M., and Montes, E. M. 2010. "An Experimental Comparison of Multiobjective Algorithms: NSGA-II and OMOPSO". In *2010 IEEE Electronics, Robotics and Automotive Mechanics Conference*, 28th September - 1st October, Cuernavaca, Mexico, 28–33.

Hashempour, N., Taherkhani, R., and Mahdikhani, M. 2020. "Energy Performance Optimization of Existing Buildings: A Literature Review". *Sustainable Cities and Society 54*:101967.

Ishibuchi, H., Imada, R., Setoguchi, Y., and Nojima, Y. 2017. "Reference Point Specification in Hypervolume Calculation for Fair Comparison and Efficient Search". In *Proceedings of the Genetic and Evolutionary Computation Conference*, 15th-19th July, Berlin, Germany, 585–592.

Kharvari, F., Azimi, S., and O'Brien, W. 2022. "A Comprehensive Simulation-Based Assessment of Office Building Performance Adaptability to Teleworking Scenarios in Different Canadian Climate Zones". *Building Simulation 15*(6):995–1014.

Lee, S. H., Hong, T., Piette, M. A., and Taylor-Lange, S. C. 2015. "Energy Retrofit Analysis Toolkits for Commercial Buildings: A Review". *Energy 89*:1087–1100.

Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H., and Tanaka, K. 2020. "Landscape-aware Performance Prediction for Evolutionary Multiobjective Optimization". *IEEE Transactions on Evolutionary Computation 24*(6):1063–1077.

Love, C. and Bozoian, S. 2022. "Defining Energy Efficiency Measures and their Impacts". RDH Building Science.

Ma, H., Zhang, Y., Sun, S., Liu, T., and Shan, Y. 2023. "A Comprehensive Survey on NSGA-II for Multi-Objective Optimization and Applications". *Artificial Intelligence Review 56*(12):15217–15270.

Markarian, E., Qiblawi, S., Krishnan, S., Divakaran, A., Thomas, A., and Azar, E. 2024. "Evaluating the Capabilities of Machine Learning Surrogate Modeling to Support High-Resolution Building Performance Simulation". In *International Conference on Computing in Civil and Building Engineering (ICCCBE)*, 25th-28th August, Montreal, Canada.

Morrison Hershfield. 2021. Building Envelope Thermal Bridging Guide. https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/power-smart/builders-developers/building-envelope-thermal-bridging-guide-v1-6.pdf, accessed 18th January 2024.

Mousavi, S., Villarreal-Marroquín, M. G., Hajiaghaei-Keshteli, M., and Smith, N. R. 2023. "Data-driven Prediction and Optimization Toward Net-Zero and Positive-Energy Buildings: A Systematic Review". *Building and Environment 242*:110578.

Muñoz, M. A., Sun, Y., Kirley, M., and Halgamuge, S. K. 2015. "Algorithm Selection for Black-Box Continuous Optimization Problems: A Survey on Methods and Challenges". *Information Sciences 317*:224–245.

National Fenestration Rating Council. 2023. NFRC Certified Products Directory for Non-Residential Fenestration Energy Certification and Rating. https://cmast.nfrc.org/?AspxAutoDetectCookieSupport=1, accessed 28th January.

National Research Council Canada. 2020. "Final Report—Alterations to Existing Buildings". Government of Canada.

Natural Resources Canada. 2022. "The Canada Green Buildings Strategy". Government of Canada.

Olu-Ajayi, R., Alaka, H., Owolabi, H., Akanbi, L., and Ganiyu, S. 2023. "Data-Driven Tools for Building Energy Consumption Prediction: A Review". *Energies 16*(6):2574.

Papadopoulos, S., Azar, E., Woon, W.-L., and Kontokosta, C. E. 2018. "Evaluation of Tree-Based Ensemble Learning Algorithms for Building Energy Performance Estimation". *Journal of Building Performance Simulation 11*(3):322–332.

Prager, R. P. and Trautmann, H. 2024. "Pflacco: Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems in Python". *Evolutionary Computation* 32(3):211–216.
Rosso, F., Ciancio, V., Dell'Olmo, J., and Salata, F. 2020. "Multi-objective Optimization of Building Retrofit in the Mediterranean Climate by Means of Genetic Algorithm Application". *Energy and Buildings 216*:109945.
Sánchez-Zabala V. F. and Gómez-Acebo, T. 2024. "Building Energy Performance Metamodels for District Energy Management Optimisation Platforms". *Energy Conversion and Management: X 21*:100512.
Seiler, M. V., Kerschke, P., and Trautmann, H. 2024. "Deep-ELA: Deep Exploratory Landscape Analysis with Self-Supervised Pretrained Transformers for Single- and Multi-Objective Continuous Optimization Problems". *arXiv Preprint arXiv:2401.01192*
Shen, Y. and Pan, Y. 2023. "BIM-Supported Automatic Energy Performance Analysis for Green Building Design Using Explainable Machine Learning and Multi-Objective Optimization". *Applied Energy 333*:120575.
Sierra, M. R. and Coello Coello, C. A. 2005. "Improving PSO-Based Multi-objective Optimization Using Crowding, Mutation and ∈-Dominance". In *Evolutionary Multi-Criterion Optimization,* edited by C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, 505–519. Berlin-Heidelberg, Germany: Springer.
Soper, D. S. 2023. "Hyperparameter Optimization Using Successive Halving with Greedy Cross Validation". *Algorithms 16*(1):17.
Sun, Y., Haghighat, F., and Fung, B. C. M. 2020. "A Review of The-state-of-the-art in Data-driven Approaches for Building Energy Prediction". *Energy and Buildings 221*:110022.
Talami, R., Dawoodjee, I., and Ghahramani, A. 2023. "Quantifying Energy Savings from Optimal Selection of HVAC Temperature Setpoints and Setbacks across Diverse Occupancy Rates and Patterns". *Buildings 13*(12):2998.
Tresidder, E., Zhang, Y., and Forrester, A. I. J. 2011. "Optimisation of Low-Energy Building Design Using Surrogate Models". In *Proceedings of Building Simulation 2011,*14th-16th November, Sydney, Australia*,* 1012-1016.
Tresidder, E., Zhang, Y., and Forrester, A. I. J. 2012. "Acceleration of Building Design Optimisation Through the Use of Kriging Surrogate Models". In *Proceedings of Building Simulation and Optimization*. 10th-11th September, Loughborough, UK, 1-8.
Wales, D. J. and Doye, J. P. K. 1997. "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms". *The Journal of Physical Chemistry A 101*(28):5111–5116.
Zhan, J., He, W., and Huang, J. 2023. "Dual-objective Building Retrofit Optimization Under Competing Priorities Using Artificial Neural Network". *Journal of Building Engineering 70*:106376.

AUTHOR BIOGRAPHIES

SEIF QIBLAWI is a Ph.D. student in the Building Engineering program at Carleton University in Ottawa, Canada. His research interests include building surrogate modeling, data-driven simulation, and computational optimization. His email address is seifqiblawi@cmail.carleton.ca.

ELIN MARKARIAN is a Ph.D. student in the Building Engineering program at Carleton University in Ottawa, Canada. Her research explores the use (and combination) of physics-based and data-driven models to provide fast and reliable building performance predictions. Her email address is elinmarkarian@cmail.carleton.ca.

SHIVRAM KRISHNAN is a Junior Research Fellow in the Department of Civil Engineering at Indian Institute of Technology Bombay, Mumbai, India. His research interests include building energy modelling, data-driven simulations and sustainable building construction. His email address is 30005323@iitb.ac.in.

ANAGHA DIVAKARAN is a Project Associate in the Department of Civil Engineering at Indian Institute of Technology Bombay, Mumbai, India. Her reasearch interests include sustainable building construction, building energy modelling and data-driven simulations. Her email address is 30005356@iitb.ac.in.

ALBERT THOMAS is an Associate Professor in the Department of Civil Engineering at the Indian Institute of Technology Bombay. His research interests include system dynamics, agent-based modeling, discrete event simulation, distributed simulation, life cycle analysis, energy simulation, and various lean construction-based techniques. His primary focus is developing frameworks and tools that help adopt sustainable and lean practices in construction projects and infrastructure systems, thereby reducing waste and environmental impacts. He is also a guest editor for the Sustainable Production and Consumption journal. His email address is albert@iitb.ac.in, and his website is https://www.civil.iitb.ac.in/~albert/.

ELIE AZAR is a Canada Research Chair and Associate Professor in the Department of Civil and Environmental Engineering at Carleton University in Ottawa, Canada. His research interests include human-centered building design and operation, occupant behavior modeling, building performance simulation, agent-based modeling, and data-driven predictive models. He has authored numerous scientific research articles in his field and edited books focused on sustainability, smart cities, and artificial intelligence. His email address is elie.azar@carleton.ca and his website is https://carleton.ca/cee/profile/elie-azar/.