Evolutionary Tuning of Building Models to Monthly Electrical Consumption

Aaron Garrett, PhD　　　　　　　**Joshua New, PhD**　　　　　　　　**Theodore Chandler**

Member ASHRAE

ABSTRACT

Building energy models of existing buildings are unreliable unless calibrated so they correlate well with actual energy usage. Calibrating models is costly because it is currently an "art" which requires significant manual effort by an experienced and skilled professional. An automated methodology could significantly decrease this cost and facilitate greater adoption of energy simulation capabilities into the marketplace. The "Autotune" project is a novel methodology which leverages supercomputing, large databases of simulations, and machine learning to allow automatic calibration of simulations that match measured experimental data. This paper shares initial results from the automated methodology on commodity hardware applied to the calibration of building energy models (BEM) for EnergyPlus (E+) to provide error rates, as measured by the sum of absolute error, for matching monthly load and electrical data from a highly-instrumented and automated ZEBRAlliance research home.

INTRODUCTION

Sustainability, with its connections to energy and climate change, is perhaps the defining challenge of our time. With only 5% of the world's population, the United States (US) consumes 25% of the world's primary energy and contributes 21% of the world's greenhouse gas emissions (DOE Buildings Data Book 2011). The largest sector of energy consumption is the ~119 million buildings in the US which consume 40% of the US primary energy (73% of the electrical energy). By 2030, it is estimated that 60% of the urban building floor space has yet to be built in China, 67% in India. With the economic growth and modernization of highly populous regions of the earth, energy security and sustainability will become increasingly important in a world whose global energy consumption is estimated to increase 50% by 2030. With buildings being the largest single sector of energy consumption, building efficiency is one of the fastest, easiest, and most cost-effective avenues toward achieving a reduction in energy use.

There are many tools available which are used to project how specific policies or retrofit packages would maximize return-on-investment with subsidies through federal, state, local, and utility tax incentives, rebates, and loan programs. Resolving issues such as principle-agent, first cost, return on investment (ROI), and the cost/performance tradeoff between state-of-the-art vs. standard equipment fall in the domain of these tools. Like all software tools, the Achilles heel is often "garbage in, garbage out" (GIGO); these tools rely on an accurate representation of the building being analyzed in order to perform an accurate analysis. A central challenge in the domain of energy efficiency for buildings is being able to realistically model specific building types; however, differences from actual monthly utility bills on the order of 24%-97% (Earth Advantage 2008, Roberts et al. 2012) are currently common. Many measurement and verification (M&V) protocols require accuracy on the order CV(RMSE) of 15% relative to monthly calibration data or 30% for hourly data (Guideline 14 2002). Achieving this accuracy is complicated by the fact that - unlike vehicles or airplanes which are designed and built according to strict schematics - buildings are typically manufactured in the field based on one-off designs. This significant variability from prototypical buildings complicates even further the fact that there are thousands of parameters which must be known about a building to accurately model it using modern simulation engines.

There are several simulation engines, and tools which leverage those simulation engines, that are actively supported (DOE Tools 2012). The major simulation engine supported by the US Department of Energy is EnergyPlus (DOE 2012), with OpenStudio (NREL 2012) serving as the primary interoperability middleware for communication between various tools and file formats. Graphical and other text-based user interfaces for E+ allow a user to provide parameters that fully describe a given building from which E+ calculates the detailed energy usage of the building. These parameters are extensive and their sensitivities are not yet fully explored. It is unrealistic to expect even an advanced user to be capable of providing accurate values for every parameter expected by E+. To mitigate this issue, users often use reference or template buildings already in a tool that are similar to their own as default points for parameter values. These values are then "corrected" to more closely match the actual building under consideration. In addition, average material properties are typically used from the ASHRAE Handbook of Fundamentals which next year will begin to publish the significant variances in material properties from controlled laboratory tests (ASHRAE 2013). However, the corrections currently made to building simulation parameters generally involve user intuition and experience and, thus, are neither scientifically rigorous, repeatable, defensible, nor transferrable. The Autotune project (New et al. 2012) aims to resolve these issues with an automated process.

If a user has accurate sensor information regarding the state of a building, it is possible to perform a search for more accurate parameters whose E+ output most closely matches the corresponding sensor data. The search space in such a problem is extremely large. Even if each parameter were a simple binary value (e.g., a categorical whose values were "yes" or "no"), the search space for a relatively small 3000-parameter building would contain 2^{3000} possibilities, which is unfathomably larger than the number of atoms believed to exist in the observable universe (about 2^{270}). This is a best-case scenario; the actual size of the search space is effectively infinite because many of the parameters are continuous-valued.

Evolutionary algorithms have been shown to efficiently search such extremely large spaces (Michalewicz and Fogel 2004). They generally avoid the problem of local optima by maintaining a population of possible solutions, rather than performing a strictly gradient-based approach. Regardless of the search algorithm employed, the massive size of the search space necessitates that large samples be drawn from which to determine promising areas to explore. This particular requirement is complicated here by the fact that a single E+ simulation (i.e., one sample) requires several minutes of computation time on a single processor core. Once again, however, evolutionary algorithms are uniquely suited to such problems because they are inherently parallel. For instance, each member of the population of solutions can be evaluated independently of the others. It is also possible to allow multiple versions of an evolutionary algorithm to run in parallel and share solutions with one another. These variants are collectively known as *island models* (Eiben and Smith 2007), and they have the potential for each population to explore different regions of the search space in different ways.

A part of the current E+ research effort has been directed towards developing a machine learning model that approximates the E+ system. In other words, given a set of E+ parameters, the model provides an approximation of the energy usage output. Such a model can execute in a fraction of the time of the actual E+ system. An island model evolutionary algorithm could be used to maintain a large population of approximate solutions that are evaluated by the machine learning model, while at the same time maintaining a much smaller population of exact solutions that are evaluated with the actual E+ system. The best solutions found so far in the approximate island could migrate to the exact island where they can drive the search towards promising areas. In this way, cheap and fast approximations can be used to focus the search for accurate solutions.

A final complication of this search problem is that there are likely many different parameter settings that provide accurate matches for the true energy usage, especially when limited sensor data is available and leaves the problem under-determined (i.e., the search space is multi-modal). Any search algorithm must be able to deal with such problems efficiently so that the user can have access to a set of near-optimal solutions from which to choose the most preferred. However, even more than being multi-modal, the problem presented here is also multi-objective. The primary objective is the minimization of the difference between the E+ software state and the real-world (sensed) state of the building. Other objectives might include minimization in the number of parameter changes from a given reference building, minimization of changes to the most sensitive parameters, and/or maximal use of traditionally-used parameters for software calibration. Once again, evolutionary algorithms are an excellent choice in such situations. Several widely used evolutionary multi-objective optimization algorithms exist, such as NSGA-II and SPEA2. These algorithms maintain a repository of candidate solutions that each represents different parameter settings and different objective values. The user can, at any point, stop the

algorithm and choose, from this repository, the near-optimal solution that best meets their needs.

BACKGROUND

This section discusses the EnergyPlus simulation system, highlights previous work in the tuning of building models, and provides an introduction to evolutionary computation.

EnergyPlus

The EnergyPlus whole-building energy simulation engine was consolidated from diverse involvement circa 1996 with functionality traceable back to DOE-2 and the US Department of Energy's Building Loads Analysis and System Thermodynamics (BLAST) from the late 1970s (EnergyPlus 2012). Its original design goals were to provide a more consistent software structure for development and modification, to allow third-party programs and components to easily interface with the core system, and to fully integrate the loads, systems, and plants into the simulation (EnergyPlus 2011). The workflow for a building modeler using a system like EnergyPlus is to create a building's geometry using external software, layer it with detailed metrics encoding material properties, and add equipment currently or expected to be in the building, including anticipated operational schedules. A typical residential building model in EnergyPlus has approximately 3,000 input parameters that must be specified.

Tuning of Building Models

Although even inaccurate models can be tremendously useful, specific business applications require sufficiently accurate building energy models according to established guidelines (Guideline 14 2002). Modification of simulation algorithms can help address this, but theprocess of tuning virtual building models to match real data remains the most tractable way to meet such requirements. Unfortunate for the advancement of the field, this tuning process has remained an "art" which even the practitioners most often do not enjoy. Informal interviews and surveys have indicated less than 1% of people actually enjoy the tuning process and, instead, see it as a necessary and laborious process. These practitioners often indicate the use of infiltration and schedules (of many types) as the primary "knobs" by which to tune the simulation to measured data. However, laboratory research has shown significant variance in the building materials used throughout the envelope of a building and will be incorporated in the upcoming ASHRAE Handbook of Fundamentals. As a starting point, this research paper will focus on tuning these critical envelope material properties before tuning based on properties such as infiltration which is currently very difficult to measure over time and schedules (for which adequate smart-home emulation and sensing is not yet widely prevalent).

It should come as no surprise that, in order to reduce the cost of business, the idea of self-calibrating energy models has been around for decades with initial attempts beginning around the early 1980s. Much of the motivations, history, different levels of thoroughness in calibration, and state-of-the-art approaches (sensitivity analysis, reducing the number of simulations necessary, optimization methods, etc.) are expertly consolidated in ASHRAE report 1051-RP on the subject (Reddy et al. 2006). The Autotune project (New et al. 2012) seeks to facilitate the realization of calibrated energy models by leveraging the world's current fastest supercomputer (299,000-core Jaguar/Titan) and several other supercomputers to increase the number of simulations possible, the latest advances in web-oriented database storage for queryable and publicly-sharable storage of 156 inputs and 96 outputs at 15-minute resolution for 8 million E+ simulations, extensive data-mining for agent-based encapsulation of knowledge for deployment of an automated tuning methodology on a desktop or via the web, simulation comparisons to a robotically emulated-occupancy ZEBRAlliance (Miller et al. 2012, Biswas et al. 2012) 2800 ft² research home with 269+ channels of 15-minute sensor data, and advancing the state-of-the-art in several machine learning algorithms for optimization. The Autotune project, in an effort to promote open science, is making a portion of the 267TB (26.9 trillion data points) of E+ simulation data publicly available at http://autotune.roofcalc.com.

Evolutionary Computation

Evolutionary computation (DeJong and Spears 1993, Spears et al. 1993, Fogel 1994, Fogel 2000) has been shown to be a very effective stochastic optimization technique (Bäck et al. 1997, Michalewicz and Fogel 2004). Essentially, an evolutionary computation (EC) attempts to mimic the biological process of evolution to solve a given problem (DeJong 2006).

Evolutionary computations operate on potential solutions to a given problem. These potential solutions are called *individuals*. The quality of a particular individual is referred to as its *fitness*, which is used as a measure of survivability (DeJong 2006). Most evolutionary computations maintain a set of individuals (referred to as a *population*). During each *generation*, or cycle, of the evolutionary computation, individuals from the population are selected for modification, modified in some way using *evolutionary operators* (typically some type of recombination and/or mutation) to produce new solutions, and then some set of existing solutions is allowed to continue to the next generation (Fogel 2000). Viewed in this way, evolutionary computation essentially performs a parallel, or beam, search across the landscape defined by the fitness measure (Russell and Norvig 2000, Spears et al. 1993). A beam search is simply a search algorithm that maintains k states, rather than just one state, at each iteration.

The particular evolutionary operators used in this work were *heuristic crossover* and *Gaussian mutation*. Heuristic crossover works as follows. Given two parents, p_1 and p_2, where the fitness of p_1 is greater than the fitness of p_2, create two children. The first child is simply p_1. Each element of the second child is created according to the equation $c[i] = p_1[i] + r(p_1[i] - p_2[i])$ where r is a uniform random value between 0 and 1. Gaussian mutation works by randomly modifying each element of a candidate solution using a Gaussian distribution centered on the current element with a variance that becomes a parameter to the algorithm.

According to Bäck et al. (1997), the majority of current evolutionary computation implementations come from three different, but related, areas: genetic algorithms (Holland 1975, Goldberg 1989, Forrest 1993, Vose 1999), evolutionary programming (Bäck et al. 1997, Fogel et al. 1966, Fogel 1994), and evolution strategies (Fogel 1994, Bäck et al. 1991). Each area is defined by its choice of representation of potential solutions and/or evolutionary operators. However, DeJong (2006) suggests that attempting to categorize a particular EC under one of these labels is often both difficult and unnecessary. Instead, he recommends specifying the representation and operators, as this conveys much more information than simply saying that a "genetic algorithm" was used, for example. We adopt DeJong's approach in this work.

METHODOLOGY

The simplest set of data that is common among all energy consumers is that of monthly electrical usage, typically in the form of a utility bill. The work presented here attempts to optimize the match between a model building and actual monthly electricity usage data. The reference building used in this work is house number 1 in the Wolf Creek subdivision (WC1), an Oak Ridge National Labs ZEBRAlliance experimental energy efficient home. This home has a plethora of energy-efficient technologies: (1) standing seam metal roof with infrared reflective pigments to boost solar reflectance, (2) ENERGY STAR appliances, (3) triple-pane low emittance Argon-filled windows, (4) compact fluorescent lighting, (5) horizontal ground loop installation that leverages foundation and utility excavations, (6) high-efficiency water-to-air heat pump for space conditioning, (7) high-efficiency water-to-water heat pump for hot water heating, (8) an energy recovery ventilator for transferring heat and moisture between fresh incoming and outgoing air, and (9) structurally insulated panel (SIP) walls filled with expanded polystyrene insulation. For more information, the interested reader is referred to (Miller et al. 2012, Biswas et al. 2012), The home has been fitted with hundreds of sensors that are capable of collecting sub-hourly data.

In the following experiments, two different model buildings are used. The first was last modified on March 29, 2012 and is stored in an E+ file named "House_1_V7_A2.idf" which matches whole-building annual electric consumption

exactly, but has a sum of absolute errors[1] (SAE) of 1,276.34 kWh for monthly and 6,242.04 kWh for hourly electrical data when summed for the entire year. An earlier version of that same model was completed on July 28, 2010, and is stored in a file named "House_1_07282010.idf" with SAE of annual electric consumption 1,623.36 kWh for monthly and 8,113.69 kWh for hourly electrical data when summed for the entire year. These two baseline models, which we shall henceforth refer to as the "refined" and "primitive" models, are separated by approximately four man-months of effort (consisting of 2 months of laboratory material testing and 2 months manually tuning the input file) over the course of nearly 2 calendar years, with the "refined" model being the recipient of that effort.

In this work, only a subset of the real-valued parameters of the models, as specified by domain experts, was used as a part of the tuning process, and the phrase "tuning parameters" will be used when referencing these variables. While all 156 tunable parameters provided through the project website are too extensive to list here, a majority of the parameters were for building material properties such as thickness, conductivity, density, specific heat, thermal absorptance, solar absorptance, and visible absorptance for materials such as gypsum board, stone, concrete foundation wall, fiberglass insulation, metal roofing, plywood, insulation, gravel, oriented strand board (OSB), and cladding as well as U-factor, solar heat gain coefficient, and visible transmittance for window glazing systems. Other parameters include fraction of latent and radiant for equipment, radiant and visible for lighting, flow coefficients for HVAC, heating and cooling air supply temperatures, building orientation, infiltration, and several others. It should also be noted while these are individual line changes in an E+ *.idf input file, several instances of each material, equipment, etc. may be used throughout a building. The Autotune system currently scales to tune any set of numerical parameters so the particular set of parameters can be customized according to the needs of a particular use case.

In 2010, the average U.S. residential building consumed enough energy to cost the homeowner $2,201 (Building Energy Data Book 2012a). An average of 44.7% of the energy went to space heating and 9.2% toward space cooling (Building Energy Data Book 2012b), totaling 53.9% for space conditioning. However, primarily because of differing costs for various fuel types (Building Energy Data Book 2012c), 28.9% of cost was for heating and 14.0% for cooling, yielding 42.9% and amounting to $944/year for space conditioning. However, the energy-efficient HVAC in WC1 actually consumed $472.62 for Jan. 1 – Nov. 30, 2010. This cost may serve as a point of reference for the tuning results presented throughout the study.

For the all-electric WC1, the actual energy usage data for all HVAC equipment was reliably collected from January 1 through November 28, 2010 at which point a new set of test HVAC equipment was installed. Therefore, in all experiments reported, the "yearly" electrical usage will always refer to the electrical usage from January 1 to November 28. The electrical usage in this work was calculated as the sum of all of the heating and cooling ideal loads for every time period (in kilowatt-hours) divided by their respective unitless coefficients of performance (COPs). In this case, the COPs were 4.1 for heating and 4.62 for cooling in order to derive the anticipated HVAC electrical consumption (kWh). While higher-resolution COP measurements are available, the annual COP measures reported here are accurate for "yearly" electrical usage. The conversion from heating and cooling loads to electrical consumption (load divided by annual COP) and subsequently cost (multiplying by utility rate) is solely for conveying the accuracy in traditional terms. For this reason, application of more advanced utility rate structures or rate-payer limitations to localize model accuracy in terms of US dollars for a given region or occupant is still valid as a post-processing exercise.

In the following experiments, the primary metric used for measuring tuning accuracy is the monthly sum of absolute errors (SAE). The SAE was calculated according to **Equation 1**, where M_i is the monthly heating+cooling load of the model and A_i is the monthly heating+cooling load of the actual ZEBRAlliance WC1 building. This equation only contains 11 months because actual data was not collected for December. The SAE was chosen for this study, instead of the root mean squared error (RMSE), because of its immediate interpretability when converted to electric utility costs.

---

[1] The sum of absolute errors was chosen as the primary metric, rather than root mean squared error, for example, because of its ease of interpretation in terms of relative error or dollar savings.

$$SAE = \sum_{i=1}^{11} |M_i - A_i| \qquad (1)$$

Experiment 1 – Abbreviating the Simulation Schedule

The E+ simulation requires a schedule that specifies temperatures and energy needs for every zone in the building being simulated for each time period. The full schedule for the reference building in this research extends for an entire year at one-hour intervals (8,760 time periods). This type of fine-grained schedule leads to a much more exact, but computationally expensive, simulation. In this case, a single simulation takes approximately eight minutes and cannot currently be efficiently parallelized.

It is possible to reduce the number of time periods in a given schedule by using "representative" time periods, rather than the entire year. In this case, four days were chosen spanning the year—January 1, April 1, August 1, and November 1—which produced only 96 time periods. This amounts to only 1% of the schedule size of the full schedule. Likewise, running E+ on this abbreviated schedule requires only a matter of seconds. However, it is important to determine whether good performance (i.e., low error compared with the actual electrical usage) on the abbreviated schedule correlates with good performance on the full schedule.

Experimental Setup. To determine the correlation between the error rate of the abbreviated schedule and that of the full schedule, the search space of E+ building parameters (using the abbreviated schedule) needed to be sampled and analyzed. However, these samples could not be truly random because random samples would almost certainly produce high error when compared to "actual" electrical usage, which would provide little information. Instead, samples needed to span the range from high error to low error.

In this experiment, the "actual" electrical usage is, in fact, the E+ output from running the refined model, which is modified by each candidate solution. To accomplish the sampling, an evolutionary computation (EC) was created with a population size of 16 individuals and was allowed to run for 1024 simulations (64 generations). This allowed the EC to begin with a random sampling of the parameter space, which could be refined through the generations to produce parameters with lower error. To ensure against statistical anomalies, the EC was run four different times with different initial populations each time.

Though not entirely necessary or relevant for the purposes of this experiment, the full EC parameters are provided for completeness. Tournament selection (tournament size 4) was used for parent selection. Generational replacement with weak elitism (one elite) was used for survivor selection. Heuristic crossover (and Gaussian mutation were used as the variation operators. The mutation usage rate was set to 1.0, and the Gaussian mutation rate was set to 10% of the allowable range of each variable. The fitness of a candidate solution was calculated to be the SAE between its monthly electrical usage and the "actual" monthly electrical usage. For the four-day schedule, this monthly electrical usage is 0 in all months except January, April, August, and November, in which case it is only the total usage for the first day of each month.

Results. The entire set of 1024 individuals for each of the four trials is plotted in **Figure 1**. This plot compares the SAE between the candidate solutions and the "actual" electrical usage, both for the four-day schedule and the yearly schedule. Individuals from each of the four trials are colored black, red, green, and blue, respectively. It is clear from the graph that there is a strong linear correlation between the electrical usage with the four-day schedule and that of the yearly schedule. More precisely, this correlation was 0.9603, 0.9421, 0.9015, and 0.9555 for each of the four trials, respectively, as seen in **Table 1**.

The high correlations between the abbreviated and full schedules reveal a possible solution to the issue of the high computational overhead of E+ simulations. It is possible, at least for the single building currently under investigation, to use a much faster abbreviated schedule as a reasonable surrogate for the more expensive full year schedule. However, these

correlations do deteriorate, sometimes dramatically, as the minimization process progresses. **Table 1** shows the correlations between the four-day and yearly SAE as the data is focused more and more toward the final *n*%. The correlation tends to decrease, and in Trial 3 it decreases substantially. This means that using the four-day SAE may not be as effective as the evolutionary search progresses. However, certainly at the beginning, the four-day SAE is a very good (and fast) approximation.

Table 1. Correlations Between Four-day and Yearly Sum of Absolute Errors of Electrical Usage

Data Used	Trial 1	Trial 2	Trial 3	Trial 4
100%	0.96033	0.94215	0.90148	0.95553
90%	0.86677	0.84380	0.65458	0.87881
80%	0.86830	0.85725	0.64437	0.89219
70%	0.87257	0.85605	0.62313	0.89435
60%	0.86708	0.84889	0.60779	0.89516
50%	0.87030	0.84888	0.55566	0.89711
40%	0.87743	0.84174	0.56627	0.90195
30%	0.87944	0.82078	0.60146	0.89731
20%	0.88406	0.82313	0.59878	0.88600
10%	0.87962	0.82348	0.39849	0.87144
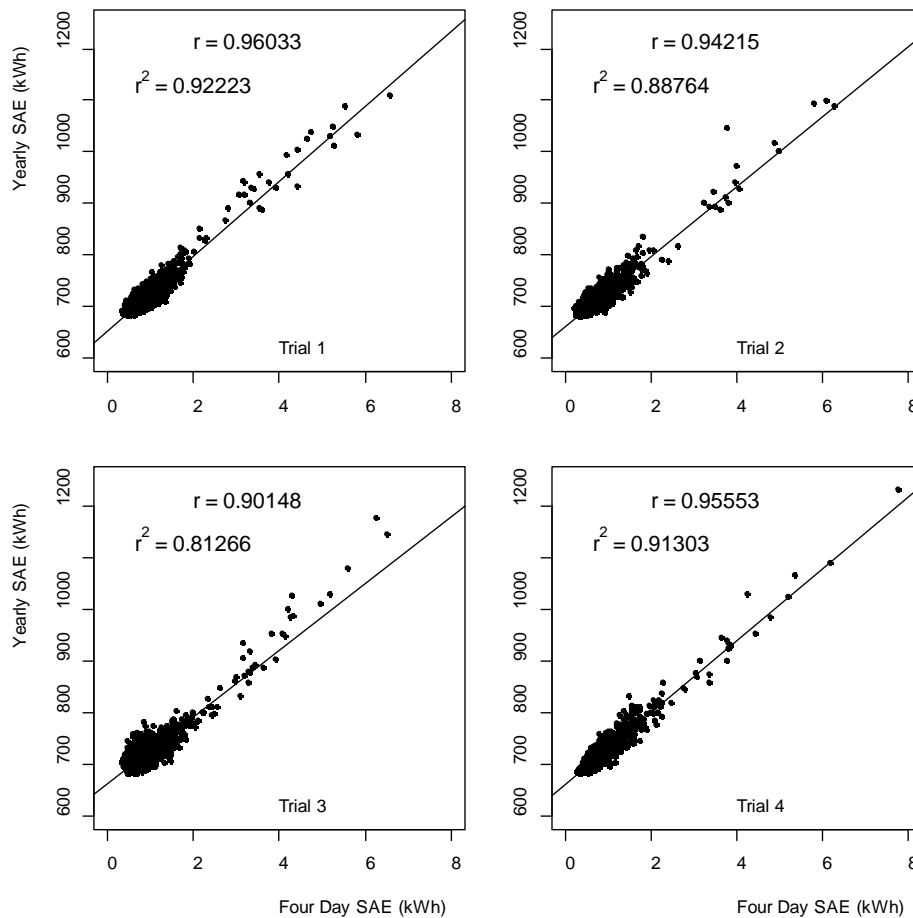
Figure 1 This is a scatterplot of all 1024 individuals from each trial in the abbreviated schedule experiment. The x-axis for each graph plots the sum of absolute errors in daily electricity usage (in kilowatt-hours) for the four-day schedule.

The y-axis plots the sum of absolute errors in monthly electricity usage (also in kilowatt-hours) for the full-year schedule. It is clear from the figure that the correlation between the two is large, which is evidence that the abbreviated schedule can serve as a surrogate for the full schedule.

Experiment 2 –Tuning the Refined Model

In this experiment, the 4-day abbreviated schedule is applied to ascertain the time required and the accuracy of the tuning process between the refined E+ model and actual WC1 data..

Experimental Setup. An EC was created containing candidate solutions composed of real values representing, and bound to the ranges of, the real-valued tuning parameters. This means that each candidate solution was a list of real values that specified those parameters of the building model. The EC used a population size of 16, tournament selection with tournament size 4, generational replacement with weak elitism (one elite), heuristic crossover, and Gaussian mutation with a usage rate of 1.0 and a mutation rate of 10% of the allowable range of each variable. The fitness reported is the sum of absolute errors (SAE) of the monthly electrical usage between the model and WC1 (and load in Megajoules), and the EC was allowed to use only 1024 fitness evaluations (i.e., E+ simulations). Because of the stochastic nature of the evolutionary search, 8 independent trials were performed.

Results. First, it is important to establish a baseline of performance. In this case, the obvious baseline is the SAE between the refined model E+ output and the actual electrical usage. These were 12.813 kWh (46.1268 MJ) for the four-day and 1276.340 kWh (4594.824 MJ) for the yearly. The final average and minimum fitness values for each trial are presented in **Table 2**. The average minimum SAE across the eight trials was 10.259 kWh (36.9324 MJ) and 1078.842 kWh (3883.831 MJ) for the four-day and yearly electrical usage, respectively. This amounts to a 20% reduction in the four-day SAE and a 15% reduction in the yearly SAE. The total computational time for all eight trials on an eight-core machine was approximately 12 hours, so it is likely that a single-core machine could process a single trial in a similar amount of time (overnight).

Table 2. Final Population Statistics in ΔKilowatt-hours (ΔMegajoules) from Tuning Refined Model

Trial	Four-day Average	Four-day Minimum	Yearly Average	Yearly Minimum
Refined	12.813 (46.1268)		1276.340 (4594.824)	
1	10.595 (38.1420)	9.997 (35.9892)	1098.040 (3952.944)	1055.857 (3801.085)
2	10.748 (38.6928)	10.248 (36.8928)	1129.179 (4065.044)	1095.948 (3945.413)
3	10.686 (38.4696)	10.133 (36.4788)	1068.646 (3847.126)	1034.244 (3723.278)
4	10.914 (39.2904)	10.504 (37.8144)	1132.985 (4078.746)	1098.540 (3954.758)
5	10.849 (39.0564)	10.440 (37.5840)	1143.954 (4118.234)	1099.360 (3957.696)
6	11.083 (39.8988)	10.631 (38.2716)	1146.128 (4126.061)	1110.514 (3997.850)
7	10.720 (38.5920)	10.090 (36.3240)	1100.887 (3963.193)	1064.195 (3831.102)
8	10.548 (37.9728)	10.031 (36.1116)	1104.110 (3974.796)	1072.097 (3859.549)

As stated previously, we use utility rates solely to convey the tuning accuracy in dollars; this does not affect our tuning methodology. Throughout this report, we'll use the 2010 national average utility rate of 11.5 cents per kilowatt-hour. In this case, the refined model produced a cumulative error of $147 per year while the average cumulative error across the eight tuned models was $124 per year. This corresponds to a reduction in error from 31% ($147 of $473 actual yearly usage) to 26%.

Experiment 3 –Tuning the Primitive Model

The refined model used in the previous experiment was constructed, as mentioned previously, through approximately four man months over the course of two calendar years. Two of the primary research questions are whether such a model could be automatically created and with what degree of fidelity to the actual energy data. In order to test this, we leverage

the primitive model from which the refined model was created. In this experiment, the primitive model is used as the base in order to determine whether the tuning results are competitive with the refined model.

Experimental Setup. The experimental setup was identical to that defined for the previous experiment.

Results. As before, it is important to establish a baseline for comparison. In this case, there are two building models—primitive and refined. Their (lack of) accuracy (SAE) compared to the actual WC1 data serves as the baseline. In this experiment, the primitive model produced a four-day SAE of 17.995 kWh (64.782 MJ) and a yearly SAE of 1623.364 kWh (5844.11 MJ). The refined model produced a four-day SAE of 12.813 kWh (46.1268 MJ) and a yearly SAE of 1276.340 kWh (4594.824 MJ). The final population statistics for the eight trials are presented in **Table 3**.

The average minimum SAE across all eight trials was 12.800 kWh (46.080 MJ) for the four-day SAE and 1415.213 kWh (5094.767 MJ) for the yearly SAE. Compared to the newer refined model, the four-day SAE of the tuned model is actually slightly better, and its yearly SAE is only about 140 kWh (500 MJ) greater. In contrast, the yearly SAE of the primitive model is about 350 kWh (1260 MJ) greater than the refined model. That corresponds to a 60% reduction from the primitive model toward the refined model. Once again, the four-day SAE, on which the EC was evaluated, shows remarkable performance. The tuned yearly SAE is a definite improvement over the primitive model, but it is likely to perform even better if it is incorporated into the evolutionary process. It should be noted that this 60% reduction in yearly SAE was achieved, as in the previous experiment, in about 12 hours on a single-core machine, rather than four man-months. It should also be noted that the manual refinements between the primitive and refined files changed several properties of the model outside the scope of the envelope material properties being automatically tuned in these experiments, making 100% SAE reduction unachievable. **Table 4** enumerates those changes. It should be noted that, of all of the properties listed in **Table 4**, only the zone infiltration flow coefficient was accessible to the tuning process.

Table 3. Final Population Statistics in ΔKilowatt-hours (ΔMegajoules) from Tuning Primitive Model

Trial	Four-day Average	Four-day Minimum	Yearly Average	Yearly Minimum
Primitive	17.955 (64.7820)		1623.364 (5844.1100)	
Refined	12.813 (46.1268)		1276.340 (4594.8240)	
1	13.085 (47.1060)	12.564 (45.2304)	1423.893 (5126.0148)	1395.917 (5025.3012)
2	13.580 (48.8880)	12.863 (46.3068)	1439.713 (5182.9668)	1415.510 (5095.8684)
3	13.774 (49.5864)	12.875 (46.3500)	1471.857 (5298.6852)	1437.560 (5175.2160)
4	13.407 (48.2652)	12.788 (46.0368)	1444.796 (5201.2656)	1415.741 (5096.6676)
5	13.422 (48.3192)	12.892 (46.4112)	1447.312 (5210.3232)	1416.477 (5099.3172)
6	13.689 (49.2804)	12.877 (46.3572)	1445.754 (5204.7144)	1420.612 (5114.2032)
7	13.504 (48.6144)	12.815 (46.1340)	1451.337 (5224.8132)	1421.126 (5116.0536)
8	13.283 (47.8188)	12.724 (45.8064)	1428.582 (5142.8952)	1398.754 (5035.5144)

Table 4. Modifications between Primitive and Refined Models

Components	Primitive Model	Refined Model
Building solar distribution	Full Exterior	Full Exterior with Reflections
Surface convection algorithm: inside	TARP	Ceiling Diffuser
Surface convection algorithm: outside	DOE-2	Simple Combined
Schedule	Simple, pre-defined schedule	Complex, independent schedule for each major system
Air gap material	Does not exist	Included as construction material
Construction		Additional wall and modified roof construction
Zone inside convection algorithm	TARP	Default
Zone outside convection algorithm	DOE-2	Default
Lights (interior and exterior)	Unspecified	Specified for each zone
Zone infiltration flow coefficient	0.01515 per zone	0.00758 per zone

Experiment 4 – Combining Abbreviated and Full Schedules

The final experiment completely duplicates the methodology of those above, except in regards to the evaluation of the fitness. In this experiment, the first 768 fitness evaluations (E+ simulations) were performed using the four-day schedule and electrical usage. At the end of those evaluations, the best half (8) of the final population's candidate solutions were inserted into a new initial population of 16 individuals that was evolved for the remaining 256 evaluations on the yearly electrical usage and schedule. In this way, the yearly schedule could be incorporated into the evolution after near-optimal candidates had quickly been found using the abbreviated 4-day schedule.

Recall that the yearly SAEs produced by the primitive model and the refined model were 1623.364 kWh (5844.11 MJ) and 1276.340 kWh (4594.824 MJ), respectively. **Table 4** shows the statistics of the final populations of the evolutionary tuning. Recall that each of the trials produces a population of 16 tuned models, of which one will have the minimum yearly SAE (as listed in the "Minimum" columns in **Table 4**). The average of these eight minimum values is the most representative value for the performance of the tuning process for each of the primitive and refined models.

Compared to the refined model with average minimum yearly SAE of 1026.844 kWh (3696.638 MJ), the tuned refined model corresponds to an almost 25% reduction in yearly SAE using the yearly schedule, compared to the 15% reduction when using the 4-day schedule. For the primitive base model, the average minimum yearly SAE was 1392.688 kWh (5013.677 MJ). Recall from above that the difference in yearly SAE between the primitive and refined models was about 350 kWh (1260 MJ). In the previous experiment, the EC was able to reduce that difference to about 140 kWh (500 MJ) (a 60% reduction). Here, the difference was reduced to about 116 kWh (418 MJ), which corresponds to a 67% reduction. It appears that there may be diminishing gains in continuing to tune the primitive model. As mentioned above, this is due to variables that were changed between the primitive and refined models to which the EC has no access. In such a case, no amount of tuning will be able to provide better performance than the refined model.

Table 5. Final Full-year Population Statistics in ΔKilowatt-hours (ΔMegajoules) from Tuning Models

	Refined		Primitive	
Trial	Average	Minimum	Average	Minimum
1	1079.026 (3884.4936)	1043.687 (3757.2732)	1417.083 (5101.4988)	1388.076 (4997.0736)
2	1070.660 (3854.3760)	1035.439 (3727.5804)	1431.135 (5152.0860)	1397.486 (5030.9496)
3	1051.714 (3786.1704)	1014.915 (3653.6940)	1405.782 (5060.8152)	1386.083 (4989.8988)
4	1061.497 (3821.3892)	1035.180 (3726.6480)	1440.669 (5186.4084)	1394.712 (5020.9632)
5	1043.299 (3755.8764)	1014.629 (3652.6644)	1419.470 (5110.0920)	1391.053 (5007.7908)
6	1071.961 (3859.0596)	1035.432 (3727.5552)	1420.174 (5112.6264)	1389.923 (5003.7228)
7	1058.082 (3809.0952)	1026.530 (3695.5080)	1443.183 (5195.4588)	1417.371 (5102.5356)
8	1036.930 (3732.9480)	1008.939 (3632.1804)	1404.181 (5055.0516)	1376.801 (4956.4836)

CONCLUSIONS AND FUTURE WORK

A portion of the results of the experiments in this work are summarized in **Table 5**. This table shows the baseline for each model, which is the yearly SAE produced between the E+ model output and the actual data from WC1. It includes results from tuning with the abbreviated four-day schedule as well as the "serial" four-day schedule followed by the full schedule evaluation.

Table 6. Summary of ΔKilowatt-hours (ΔMegajoules) for Different Approaches

Model	Baseline	Tuning Abbreviated	Tuning Serial
Refined	1276.340 (4594.8240)	1078.842 (3883.8312)	1026.844 (3696.6384)
Primitive	1623.364 (5844.1104)	1415.213 (5094.7668)	1392.688 (5013.6768)

Table 6 provides information about the CV(RMSE) of the baselines and different tuning approaches. Here, the CV(RMSE) is calculated as specified in **Equation 2**. While these values are not competitive with the ASHRAE Guideline 14 standard of 15% for monthly utility CV(RMSE), they do provide a starting point and a context for both the inadequacy of the primitive model and the ability of the tuning process to reduce the error noticeably.

$$CV(RMSE) = \sqrt{\frac{\sum_{i=1}^{11}(M_i - A_i)^2}{11}} \Big/ \frac{\sum_{i=1}^{11} A_i}{11} \qquad (2)$$

Table 7. Summary of CV(RMSE) of Monthly Electrical Usage for Different Approaches

Model	Baseline	Tuning Abbreviated	Tuning Serial
Refined	33.7%	31.7%	30.5%
Primitive	62.5%	52.3%	51.4%

These experiments have shown that an evolutionary search can provide effective tuning to an existing model, in acceptable time (e.g., overnight), when judged on monthly electrical usage. The primitive and refined models under consideration were both significantly improved upon by the EC in comparison to actual measured data. In the first experiment, justification was given and experimentation verified the use of an accelerated tuning methodology using abbreviated schedules. In the second experiment, operating only on the four-day surrogate simulations, the EC was capable of reducing the SAE by 15%. In the third experiment, the tuned primitive model's SAE was reduced by 13%, replacing ~60% of the ~4 man-months with an overnight computation on a single core. In the final experiment, the yearly simulations were brought into the evolutionary process directly by allowing the final quarter of the search process to specifically use the yearly, rather than the four-day, simulations as part of a "serial" process to measure fitness. The tuned primitive model reduced SAE by 14%, which was a 67% move toward the refined model (with 100% unachievable since properties outside the scope of tuning material properties were used to change the files between the primitive and refined models). Tuning the refined model resulted in a 25% reduction in SAE.

These experiments do not address central tuning questions regarding tuning of properties other than electrical data for HVAC consumption using envelope material properties, whether the modifications made to the input file are physically realistic, and prioritization of which properties should be explored first. These subjects will be partially addressed in publications for previously conducted work whereas the following subjects may be addressed in future work. The Autotune process is simulation-engine agnostic and is being designed to ease its general application; it is currently being tested as a calibration procedure for DOE's Weatherization Assistance Program's National Energy Audit Tool (NEAT) used to weatherize 1 million homes as part of the American Recovery and Reinvestment Act (ARRA). More importantly, the Autotune project (New et al. 2012) is designed to be domain-agnostic as a way of bridging the gap between the virtual world and the real one, in holding software accountable, and automating the process of empirical science..

ACKNOWLEDGMENTS

REFERENCES

ASHRAE. 2013. *Handbook of Fundamentals (chapter 26): Heat, Air, and Moisture Control in Building Assemblies – Material Properties*.

Bäck, T., F. Hoffmeister, and H.-P. Schwefel. 1991. A survey of evolution strategies. In R. K. Belew and L. B. Booker, editors, *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 2–9. Morgan Kaufman.

Bäck, T., U. Hammel, and H.-P. Schwefel. 1997. Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation* 1(1):3–17.

Biswas, K., A. Gehl, R. Jackson, P. Boudreaux, and J. Christian. Comparison of Two High-Performance Energy Efficient Homes: Annual Performance Report, December 1, 2010 – November 30, 2011. Oak Ridge National Laboratory report ORNL/TM-2011/539. Available: http://info.ornl.gov/sites/publications/files/Pub34163.pdf

Briggs, R. S., R. G. Lucas, and Z. T. Taylor. 2003. Climate classification for building enrgy codes and standards: Part 1---development process and Part 2---zone definitions, maps, and comparisons. *ASHRAE Transactions* 109(1):109–130.

Building Energy Data Book. 2012a. Average Annual Energy Expenditures per Household, by Year ($2010). Available: http://buildingsdatabook.eren.doe.gov/docs/xls_pdf/2.3.9.pdf

Building Energy Data Book. 2012b. Residential Sector Energy Consumption. Table 2.1.5. Available: http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=2.1.5

Building Energy Data Book. 2012c. 2010 Residential Energy End-Use Expenditure Splits, by Fuel Type ($2010 Billion). Table 2.3.5. Available: http://buildingsdatabook.eren.doe.gov/TableView.aspx?table=2.3.5

DeJong, K. A., and W. Spears. 1993. On the state of evolutionary computation. In Stephanie Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 618–623, San Mateo, CA: Morgan Kaufman.

DeJong, K. A. 2006. *Evolutionary Computation: A Unified Approach*. MIT Press.

Department of Energy. 2008. Energy Efficiency Trends in Residential and Commercial Buildings (Figure 18). Oct. 2008. Available: http://apps1.eere.energy.gov/buildings/publications/pdfs/corporate/bt_stateindustry.pdf

Department of Energy. 2010. *Building Energy Data Book*. Available: http://buildingsdatabook.eren.doe.gov

Department of Energy Tools 2012. Building Energy Software Tools Directory. Available: http://apps1.eere.energy.gov/buildings/tools_directory/subjects_sub.cfm

Deru, M., K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, and M. Rosenberg. 2011. US Department of Energy commercial reference building models of the national building stock.

Earth Advantage Institute and Conservation Services Group. 2009. Energy Performance Score 2008 Pilot: Findings and Recommendations Report.

EIA, U.S. Energy Information Administration. 2012. Frequently Asked Questions. Available: http://www.eia.gov/tools/faqs/faq.cfm?id=97&t=3.

Eiben, A. E. and Smith, J. E. 2007. *Introduction to Evolutionary Computing*. New York: Springer.

EnergyPlus. 2012. Getting Started with EnergyPlus: Basic Concepts Manual. Available: http://energyplus.gov

Fogel, D. B. 1994. An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks* 5(1):3–14.

Fogel, D. B. 2000. What is evolutionary computation? *IEEE Spectrum* 37(2):26–32.

Fogel, L. J., A. J. Owens, and M. J. Walsh. 1966. *Artificial Intelligence through Simulated Evolution*. New York: Wiley.

Forrest, S. 1993. Genetic algorithms: principles of natural selection applied to computation. *Science*, 60:872–878.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.

Guideline 14-2002, ASHRAE. Measurement of Energy and Demand Savings.

Holland, J. H. 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI.

IECC 2009 and ASHRAE 90.1-2007. Energy Code Climate Zones.

Michalewicz, Z., and D. B. Fogel. 2004. *How to Solve It: Modern Heuristics*. Springer.

National Renewable Energy laboratory. 2012. OpenStudio: Commercial Buildings Research and Software Development. Available: http://openstudio.nrel.gov

Miller, W.A. et al. 2012. ZEBRAlliance: An Alliance Maximizing Cost-Effective Energy Efficiency in Buildings. Available: http://www.zebralliance.org.

New, J. R., J. Sanyal, M. Bhandari, and S. Shrestha. 2012. Autotune E+ Building Energy Models. *Proceedings of the 5th National SimBuild of IBPSA-USA*, International Building Performance Simulation Association (IBPSA), Aug. 1–3, 2012. Available: http://www.ibpsa.us/simbuild2012/Papers/SB12_TS05a_1_New.pdf

Reddy, T.A., I. Maor, S. Jian, and C. Panjapornporn. 2006. Procedures for Reconciling Computer-Calculated Results with Measured Energy Data. ASHRAE Research Project RP-1051, 2006.

Roberts, D., N. Merket, B. Polly, M. Heahey, S. Casey, and J. Robertson. 2012. Assessment of the U.S. Department of Energy Scoring Tool. National Renewable Energy Laboratory.

Russell, S., and P. Norvig. 2002. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition.

Spears, W. M., K. A. DeJong, T. Bäck, D. B. Fogel, and H. deGaris. 1993. An overview of evolutionary computation. In *Proceedings of the 1993 European Conference on Machine Learning*.

Vose, M. D. 1999. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press.