AUTOTUNE E+ BUILDING ENERGY MODELS

Joshua New, Jibonananda Sanyal, Mahabir Bhandari, and Som Shrestha
Oak Ridge National Laboratory, Oak Ridge, TN

ABSTRACT

This paper introduces a novel “Autotune” methodology under development for calibrating building energy models (BEM). It is aimed at developing an automated BEM tuning methodology that enables models to reproduce measured data such as utility bills, sub-meter, and/or sensor data accurately and robustly by selecting best-match E+ input parameters in a systematic, automated, and repeatable fashion. The approach is applicable to a building retrofit scenario and aims to quantify the trade-offs between tuning accuracy and the minimal amount of “ground truth” data required to calibrate the model. Autotune will use a suite of machine-learning algorithms developed and run on supercomputers to generate calibration functions. Specifically, the project will begin with a de-tuned model and then perform Monte Carlo simulations on the model by perturbing the “uncertain” parameters within permitted ranges. Machine learning algorithms will then extract minimal perturbation combinations that result in modeled results that most closely track sensor data. A large database of parametric EnergyPlus (E+) simulations has been made publicly available. Autotune is currently being applied to a heavily instrumented residential building as well as three light commercial buildings in which a “de-tuned” model is autotuned using faux sensor data from the corresponding target E+ model.

INTRODUCTION

In 2006, the US consumed $220 billion in annual energy costs with 39% of primary energy (73% of total electrical energy) being consumed by buildings; with less than 2% of this energy demand being met by renewable resources, the US constituted 21% of worldwide CO₂ emissions in 2005 with an annual growth rate of 1.2% from 1990-2005 (U.S. Dept. of Energy, 2010) (Figure 1). For reasons financial, environmental, and social, the United States Department of Energy (DOE) has set aggressive goals for energy efficiency, which constitutes the low-hanging fruit for slight to moderate energy savings in the US buildings sector.

A central challenge in the domain of energy efficiency is being able to realistically model specific building types and scaling those to the entire US building stock (Deru et al., 2011) across ASHRAE climate zones (IECC 2009 and ASHRAE 90.1-2007; Briggs et al., 2003a,b), then projecting 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. Nearly all energy efficiency projections are reliant upon accurate models as the central primitive by which to integrate the national impact with meaningful measures of uncertainty, error, variance, and risk. This challenge is compounded by the fact that retrofits and construction of buildings happen one at a time and an individual building is unlikely to closely resemble its prototypical building. Unlike vehicles and aircraft, buildings are generally manufactured in the field based on one-off de-
signs and have operational lifetimes of 50-100 years; each building would need to be modeled uniquely and more precisely to determine optimal energy efficiency practices.

This challenge has been partially addressed through the many software packages developed for energy modeling and software tools which leverage them. There are over 20 major software tools with various strengths and weaknesses in their capability of realistically modeling the whole-building physics involved in building energy usage (Crawley et al., 2008). The major software supported by DOE is EnergyPlus (E+), constituting approximately 600,000 lines of FORTRAN code. There are many tools which use similar simulation engines, such as the National Renewable Energy Laboratory’s (NREL) BEopt (Christensen et al., 2006) and Lawrence Berkeley National Laboratory’s (LBNL) Home Energy Saver (HES) (Mills, 2008), in order to determine a set of optimal retrofit measures. There are many other use cases for energy simulation engines and tools, some of which are becoming required by law such as the progressive California Legislature Assembly Bills AB1103 (California Energy Commission, 2010a) and AB758 (California Energy Commission, 2010b) which require energy modeling anytime commercial property changes owners. The increasing application of energy software and the accuracy of projected performance is entirely contingent upon the validity of input data: a sufficiently accurate input model of an individual building and its use is required.

One of the major barriers to DOE’s Building Technology Program (BTP) goals and the adoption of building energy modeling software is the user expertise, time, and associated costs required to develop a software model that accurately reflects reality (codified via measured data). The sheer cost of energy modeling makes it something that is primarily done by researchers and for large projects. It is not a cost that the retrofit market or most use cases would absorb in the foreseeable future without drastic reductions in the cost of having cheaper and more accurate model generation. This weak business case, along with concerns regarding the cost for upkeep, maintenance, and support of the very capable E+ simulation engine, has driven DOE sponsors to investigate facilitating technologies that would enable the energy modeler and retrofit practitioner in the field.

The business-as-usual approach for modeling whole-building energy consumption involves a building modeler using the software tool they have most experience with to create the geometry of a building, layer it with detailed metrics encoding material properties, adding equipment currently or expected to be in the building, with anticipated operational schedules. An E+ building model has ~3,000 inputs for a normal residential building with very specific details that most energy modelers do not have the sources of data for. Experimentation has established that even the ASHRAE handbook and manufacturer’s label data are not reliable due to substantial product variability for some materials (DeWit, 2001). This is compounded by the fact that there is always a gap between the as-designed and as-built structure (e.g., contractors may neglect to fill one of the corner wall cavities with insulation). Due to the sources of variance involved in the input process, it should come as no surprise that building models must often be painstakingly tuned manually to match measured data. This tuning process is highly subjective and repeatable across neither modelers nor software packages. An automated self-calibration mechanism capable of handling intense sub-metering data is called for.

The development of an autotuning capability (Figure 2) to intelligently adapt building models or templates to building performance data would significantly facilitate market adoption of energy modeling software, aid in accurate use cases such as the effective retrofit strategies for existing buildings, and promote BTP’s goals of increased market penetration for energy modeling capabilities. The idea of self-calibrating energy models has been around for decades and expertly consolidated in an ASHRAE report on the subject (Reddy et al., 2006), but is generally lacking in its employ of machine learning algorithms or similar autonomous application of modern technology. In this initial paper, we discuss the general methodology behind the Autotune project, specific technologies enabling its implementation, and preliminary data generation results including a large database of parametric E+ simulations available for general use.

**SIMULATION/EXPERIMENT**

The goal of the Autotune project is to save building modelers time spent tweaking building input parameters to match ground-truth data by providing an "autotune"
Figure 3: A virtual building model (software space) and a real building (sensor space), when viewed as vectors of numbers, allows a mathematical mapping between vector spaces for direct comparison between simulation state and sensed world state.

easy button for their computer which intelligently adjusts model inputs. In order to achieve this, the Autotune project entails running millions of parametric E+ simulations on supercomputers, multi-objective optimization of E+ variables via sensitivity analysis, using machine learning systems to characterize the effect of individual variable perturbations on E+ simulations, and adapting an existing E+ model to approximate sensor data. The system will be demonstrated using an E+ building model automatically matched to a subset of the 250+ sensors in a heavily instrumented residential research building as well as to DOE’s commercial reference buildings (Field et al., 2010) for a medium office, stand-alone retail, and warehouse in which 3 customized buildings will provide faux sensor data for tuning the original models. This paper will summarize the Autotune methodology focusing primarily on the definition of parametric simulations and accessibility of the public database.

Parametric Analysis

Sensitivity analysis is a standard statistical technique (Bradley et al., 1977) in which a large parametric sweep of possible values for each input variable in a simulation is altered and then mathematically classified as contributing the variance in the final simulation result. This technique has been the hallmark mathematical technique for several analyses regarding energy efficiency. In fact, the oft-referenced Building Energy Data Book (U.S. Dept. of Energy, 2010) does not use direct measurements of the reported data, but relies upon ratios developed in earlier reports (Huang et al., 1987), some of which can be traced back to reports from the Energy Crisis in the late 1970s. In Huang et al. (1987), the authors used thousands of DOE-2 simulations to establish sensitivities and develop look-up tables for practitioners in the field since energy modeling, particularly in a mobile fashion, was inaccessible at that time. As a potential use case, DOE sponsors have considered forming a new basis consisting of hundreds of millions of E+ simulations, rather than thousands of DOE-2 runs, to develop more modern and robust data for use in a reconstruction project. As such, we are using the latest version of E+ and OpenStudio to run millions of simulations, store those in a database, and make that database publicly accessible for anyone to mine for relevant knowledge.

The computational space for this search problem is one crucial aspect of the project. While a database of millions of simulations would be a boon to the energy analysis community, it would not be sufficient for the success of this project. Domain experts have defined a set of parameters for a building model that it would be preferable to vary; however, all combinations of these variables would require $5 \times 10^{52}$ E+ simulations. There are many techniques to be utilized in an effort to effectively prune and intelligently sample the search space. First, domain experts have identified $\sim 156$ parameters typically used by energy modelers that need to be varied and ranked them in several importance categories. Second, building experts have realistic (minimum, maximum, and step size) ranges for those variables. Third, researchers have defined meta-parameters that allow several individual parameters to be varied as a function of a single variable. Fourth, low-order Markov simulations are being conducted to de-
termine variables with a monotonic effect on sensor data that could reliably be interpolated to estimate impact of a given variable. Fifth, sources of variance for individual variables in the initial results will be used to guide higher sampling rates for more sensitive variables. Sixth, an expert in multi-parameter optimization will be investigating computational steering algorithms to determine the optimal sampling strategy for the remaining space beyond the brute-force sampling of higher order Markov chains of Monte Carlo simulations.

Mapping Mechanism

In order for autotuning to work, there must be a mapping from the measured data to the corresponding state variables within the simulation (Figure 3). By defining a mathematical mapping between measurements in sensor space and simulation variables in software space, a Euclidean or similar vector-distance approach can be used to identify “how close” the software simulation is to the measured performance.

This mapping must be performed by domain experts initially, but the expert-defined mapping will be mined to discover labeling patterns used by the domain experts. The final result will be a data dictionary in which other field experiments can easily have their sensor data mapped to internal software state using labels (i.e. Temperature °F, north wall, 3' above grade). We also plan to investigate automating the mapping for new sensor data using machine learning techniques. This general mapping mechanism is necessary for widespread use of the autotune technology.

While vector-distance is used as an error metric, it should be pointed out that the search space is so large that there most likely exists a large multitude of feasible solutions (buildings which match the measured data within some threshold). We anticipate eventually using clustering to present unique/representative solutions. However, as additional outputs are added (e.g. room temperatures), the problem becomes more difficult to find a robust match, thereby reducing the number of potential solutions and allowing quantification of the tradeoffs between vector size and tuning accuracy. While the commercial buildings discussed in the Commercial Building Simulation section were selected to allow direct comparison of “actual” building properties to the tuned models, it is important to realize that approaches employed by Autotune offer the capability of compensating not only for input errors, but for the unavoidable algorithmic approximations required by software modeling algorithms on computing devices.

Suite of Machine Learning Algorithms

Machine learning allows the autonomous generation of algorithms by iteratively processing empirical data in order to allow repeatable detection of patterns (Figure 4). More importantly, cross-validation techniques ensure that each instance of a machine learning technique (agent) learns only from a small portion of the data and then its classification accuracy is tested on data which it has not seen before. This process of validation is crucial to the generalized learning necessary for properly capturing BEM dynamics without over-fitting for a specific building. This process is rarely used by energy modelers in the manual tuning process and is the primary culprit for post-retrofit measurements not matching a model that was expertly tuned.

Each type of learning system has its own strengths and
weaknesses, making it particularly suited for solving a particular type of problem. Moreover, a given type of learning system can vary in its performance based upon its own internal variables (learning rate, etc.). We have previously developed a suite of machine learning algorithms, called MLSuite, that allows general XML-file based definition of jobs to run on supercomputers and was published previously for testing “sensor-based energy modeling” (sBEM) in which whole building electrical usage was predicted as a function of sensor data (Edwards et al., 2012). MLSuite currently allows various types of parameter-settings for multiple learning systems, input orderings, cross-validation techniques, and accuracy metrics to analyze the patterns in simulation data. It includes the following 8 machine learning algorithms: linear regression, genetic algorithms, feed forward neural networks, non-linear support vector regression, hierarchical linear regression experts, hierarchical least-squares support vector regression experts, hierarchical feed forward neural network experts, and Fuzzy C-means with local models of feed forward neural networks.

A massive amount of data will be generated during the parametric sensitivity analysis, and mapped to the sensor data. This data captures dynamics that can quickly inform the role multiple simulation input variables have on the simulation output to inform the Autotuning process. There are three primary learning tasks that have been defined for MLSuite which constitute novel and promising data mining use cases for the building community: pattern detection, simulation approximation, and inverse modeling (Kissock et al., 2003).

Pattern detection of single-variable parametric simulations (all other variables constant) can be used to determine the sensitivity and pattern changes evoked by that “knob” of the simulation. By detecting the patterns for every pre-computed combination of parameters, a set of “knob turns” can be defined which is expected to push the simulation results into alignment with sensor data.

The primary problem and focus of development effort in the latest E+ 7.0 was to address the long simulation runtime. E+ simulations vary with the amount of temporal resolution required in reporting, algorithms used to model certain properties, the amount of equipment included, and many other properties. While an envelope-only simulation takes 2 minutes, one with ground loops and additional equipment currently takes ~ 9 minutes. The parametric database stores a compressed and vectorized version of the E+ input file (*.idf) and 15-minute data for 82 E+ report variables (*.csv). By applying MLSuite to process the IDF as the input feature vector to learn and reliably match the CSV output feature vector, machine learning agents can be developed which require kilobytes (KB) of hard drive space to store and can give approximate E+ simulation results for a given input file in seconds rather than minutes. Tradeoffs between storage, runtime, and accuracy are currently undergoing study.

Inverse modeling (Kissock et al., 2003) is a method of working backwards from observed sensor data to information about a physical object/parameter; this method works even if the physical parameter is not directly observable. In the context of BEM, inverse modeling often works backwards from utility bill data and use mathematical models (primarily statistics and model assumptions) to identify more specific breakdown of energy use within a building. By using CSV data as the input feature vector and IDF as the output feature vector, machine learning algorithms can be used to predict E+ input files as a function of sensor data and is the primary autotuning technique currently being tested.

DISCUSSION AND RESULT ANALYSIS

Residential Building Simulations

A three-level highly energy efficient research house, with a conditioned floor area of 382 m$^2$, was selected for the initial phase of this project. This house is one of the four energy efficient ZEBRAAlliance houses (http://zebralliance.com) built using some of the most advanced building technology, products, and techniques available at the time of construction. The main reasons for this house selection was to eliminate the uncertainties of input parameters and schedules (such as lighting, plug loads and occupancy) through emulated occupancy and since it was very heavily instrumented for validation studies allowing investigations into the tuning capabilities with intense submetering. In this unoccupied research house, human impact on energy use is simulated to match the national average according to Building America benchmarks (Hendron and Engebrecht, 2010) with showers, lights, ovens, washers and other energy-consuming equipment turned on and off exactly according to schedule. This house uses a structurally insulated panel (SIP) envelope with a thermal resistance of 3.7 m$^2$K/W, with
very low air leakage (measured $\text{ACH}_{50} = 0.74$) and thus has very low heat gain and loss through the building envelope. The details of this house’s envelope and other characteristics are described in Miller et al. (2010) (Figure 5).

This E+ model was created and carefully iterated and compared to sensor data by domain experts, but many discrepancies still exist. This is compounded by the fact that there are many input parameters for which a precise value cannot be attained; as examples: the conductivity of all materials used, radiant fraction of all lighting fixtures, submetering of all individual plug loads, or heat dissipated by the dryer to the conditioned space. Various studies, including Hopfe and Hensen (2011), highlight the danger in combining multiple uncertainties in input parameters due to their different source of nature (climatic, structural, or serviceability parameters), controllability, etc.; therefore, during the first part of this project the main focus is on the building envelope related input parameter uncertainties. A set of 156 parameters was selected for the initial variation. Since many of the characteristics for this house were identified through lab tests, experts decided to specify a realistic range for the uncertain parameters manually instead of assigning a fixed percentage variation as used in several calibration and uncertainty analyses (O’Neill et al., 2011). A base, minimum and maximum value was assigned to each of the 156 parameters. This approach allows greater specificity over the parameter values while reducing the number of parameter variations.

**Commercial Building Simulations**

While the residential application allows connection to real-world sensor data and tests for practical deployment decisions, the commercial buildings were chosen to allow a cleaner approach to the testing of multiple autotuning methodologies. DOE’s reference buildings for warehouse, medium office, and stand-alone retail were selected due to their predominance in either number of buildings or square footage in the US. In an approach similar to signal-processing, we have made changes to the original models to create 3 “golden” models, added noise by permuting random variables to create 3 “de-tuned” models, and then use internal E+ variables from simulation runs of the golden models as “sensor data” for tuning the “de-tuned” models back to the golden” models.

The warehouse, retail, and office golden models have been defined to use approximately 5%, 10%, and 20% more electrical energy than the original models, respectively. These changes were created using overlapping subsets of input variables and show, in agreement with previous sensitivity analysis studies, that small changes add up quickly.

**Open Research Buildings Database**

In order to deploy Autotune as a desktop program in which a limited number of E+ simulations can be run, several mechanisms are required to speed up the process. In addition to the application of machine learning, pre-computing E+ simulations using supercomputers are necessary to explore the combinatorial search space of E+ input parameters. Time on several supercomputers have been competitively awarded or used to demonstrate the ability to scale software and algorithms for these resources. Systems include the 1024-core shared memory Nautilus, 2048-core Frost, and 224,256-core Jaguar which is currently the 3rd fastest supercomputer in the world at 2.3 petaflops and is in transition to become the 299,008-core Titan. Frost is being used as a staging area to verify large computational parameter sweeps before running on Jaguar and both are used primarily for embarrassingly parallel compute-bound E+ simulation jobs. Nautilus unique shared-memory architecture allows every core to access the 4TB (terabytes) of Random Access Memory (RAM) for processing of memory-bound jobs common in machine learning.

The parametric simulations run by desktops and supercomputers has been uploaded to a centralized database to allow public access to this data (Figure 6). It is anticipated that this data would be of interest to researchers at univer-
Figure 8: Screenshot of EplusCleaner showing desktop client upload of simulations.

The data storage and access software has been architected as a distributed, heterogeneous, client-server framework. The embarrassingly parallel nature of the independent simulations allows us to exploit computational resources that are remote and disparate leading to an architecture capable of collecting simulation results from individual desktop systems as well as supercomputing resources. Our experiments indicate that the system functions efficiently and has been found to be bound primarily by the network bandwidth connecting the resources or local hard disk access. The database engine currently in use is the MyISAM relational MySQL database, although tools have been designed in a general manner so as to allow easy interchange as database storage technologies continue to evolve. The database has been created in a manner that allows data compression and efficient retrieval. Data access patterns are being studied to allow re-architecting the database and load-balancing for higher efficiency. The internal data storage format is not tied to the format of input or output E+ variables but instead uses its own generic internal naming scheme. Depending on the current set of variables and preferences of the users, a custom view of the data is provided that can be easily queried, summarized, and analyzed, providing the full benefits of a relational database system. Figure 7 shows the various software components of the Autotune database illustrating the independence of the data storage mechanism from the user view of the data, the software components on the server, and the remote web-based clients. There are several methods for accessing the data: a web-based IDF reconstructor, command-line access for MySQL queries, phpMyAdmin for GUI-based data interaction, a webpage for uploading simulation data, and EplusCleaner.

One of the components of this framework is an application named EplusCleaner (Figure 8) which has been developed using the Qt Software Development Kit (SDK) (Nokia) for platform-independent support. It has been architected to provide a powerful and intuitive interface capable of cleaning up after E+ simulation runs, compressing the input idf and the E+ output, and sending it to the server for database entry while waiting on the server for success/error status messages. It concurrently, continually, and remotely consolidates the parametric simulation data. Options for compression or deletion on the local machine keep E+ from flooding the local hard drive with storage of simulation results. The EplusCleaner client was run simultaneously on several machines and exhibited proper cleaning, compressing, and parsing with no observable slow-down in the simulation process indicating a bottleneck at the client machines access to the local hard drive. The database server keeps track of submissions from clients and a comprehensive log of the data provenance such that back-traces for troubleshooting may be performed if necessary. Upon receiving a compressed unit of E+ data, the server decompresses the data, creates a vector representative of the input, and commits the entire unit to a database.

A web-based method for reconstructing an IDF file from the database vector is provided which allows users to retrieve an IDF file from a stored vectorized set of input parameters. A web-interface is also available for uploading external E+ simulations input and output files to the database. External access to this database can be provided upon request using several user validation and access methods including a command line interface, password-protected phpMyAdmin for interactive queries, drill-down, and analysis of the simulation database. As of the time of this writing, the server currently hosts tens of thousands of parametric E+ simulations in 136GB from multiple distributed workstations and supercomputers, but millions of simulations (trillions of data points) are anticipated by time of publication. The latest Autotune project information, including database size and access methods, can be found at http://autotune.roofcalc.com.
CONCLUSION

A heavily instrumented residential building has been selected to leverage intense submetering for the autotuning process while eliminating variability due to occupant behavior through emulated occupancy. An E+ model of this house has been iteratively refined by experts to model the house. Experts have identified 156 input parameters to be varied with min, max, and step-sizes for undermined properties.

DOE’s reference buildings for warehouse, stand-alone retail, and medium office have been selected for creating “golden” models that use 5%, 10%, and 20% more electrical energy, respectively. “De-tuned” models have been created by permuting an undisclosed number of overlapping subsets of E+ input parameters. E+ variables from runs of the “golden” models will be used for autotuning “de-tuned” models back to “golden” models.

The database and submission/retrieval software tools for Autotune have been developed with generalizability and scalability in mind. Capabilities developed include a platform-independent Qt application named EplusCleaner for continual curation, compression, and upload of simulation data to a centralized MyISAM server storing tens of thousands of E+ parametric simulations with many mechanisms allowing public access. The software system is distributed, heterogenous, scalable and could potentially evolve into a full-fledged simulation, curation, and data assimilation framework.

ACKNOWLEDGMENT

This work was funded by field work proposal CEBT105 under the Department of Energy Building Technology Activity Number BT0201000. We would like to thank Amir Roth for his support and review of this project. We would like to thank our collaborators which include Mr. Richard Edwards and Dr. Lynne Parker from The University of Tennessee, Dr. Aaron Garrett from Jackson-ville State University, and Mr. Buzz Karpay from Karpay Associates. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Our work has been enabled and supported by data analysis and visualization experts, with special thanks to Pragnesh Patel, at the NSF funded RDAV (Remote Data Analysis and Visualization) Center of the University of Tennessee, Knoxville (NSF grant no. ARRA-NSF-OCE-0906324 and NSF-OCE-1136246).

Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Dept. of Energy under contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC, under Contract Number DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

REFERENCES


