Quantitative Methods for Comparing Different HVAC Control Schemes[*]

Anil Aswani
EECS, Berkeley, CA 94720
aaswani@eecs.berkeley.edu

Neal Master
EECS, Berkeley, CA 94720
neal.m.master@
berkeley.edu

Jay Taneja
EECS, Berkeley, CA 94720
taneja@cs.berkeley.edu

Andrew Krioukov
EECS, Berkeley, CA 94720
kriokov@cs.berkeley.edu

David Culler
EECS, Berkeley, CA 94720
culler@cs.berkeley.edu

Claire Tomlin
EECS, Berkeley, CA 94720
tomlin@eecs.berkeley.edu

ABSTRACT
Experimentally comparing the energy usage and comfort characteristics of different controllers in heating, ventilation, and air-conditioning (HVAC) systems is difficult because variations in weather and occupancy conditions preclude the possibility of establishing equivalent experimental conditions across the order of hours, days, and weeks. This paper is concerned with defining quantitative metrics of energy usage and occupant comfort, which can be computed and compared in a rigorous manner that is capable of determining whether differences between controllers are statistically significant in the presence of such environmental fluctuations. Experimental case studies are presented that compare two alternative controllers (a schedule controller and a hybrid system learning-based model predictive controller) to the default controller in a building-wide HVAC system. Lastly, we discuss how our proposed methodology may also be able to quantify the efficiency of other building automation systems.

Categories and Subject Descriptors
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*control theory, performance, measurement*; J.7 [**Computer Applications**]: Computers in Other Systems—*command and control*

General Terms
Experimentation,Measurement,Performance

Keywords
HVAC, energy-efficiency, comfort, metrics, control

---

[*]Open source software that implements our methodology in the MATLAB language can be found at the following link `http://hybrid.eecs.berkeley.edu/~NEDE/compEng.zip`

1. INTRODUCTION
Heating, ventilation, and air-conditioning (HVAC) systems contribute to a significant fraction of building energy usage. As a result, these systems have seen an increasing amount of research towards their modeling and efficient control (e.g., [16, 10, 17, 13, 3, 14]). The primary challenge is ensuring comparable levels of occupant comfort, in relation to existing HVAC controllers, while achieving reductions in energy consumption. Simulations and experiments indicate that this goal is achievable for a large variety of systems.

Experimentally comparing the energy efficiency and comfort of different control schemes is difficult because of the large temporal variability in weather and occupancy conditions. Furthermore, energy usage of HVAC equipment does not typically scale linearly in the relevant variables. For example, a simplistic model would state that the amount of energy E required to maintain a set point of T_s for the supply air temperature (SAT) of an HVAC system with a warmer outside air temperature (OAT) of $T_o(> T_s)$ is proportional to $E \sim T_o - T_s$. We have observed on our building-wide testbed [5] that such models do not capture the complex energy characteristics of the equipment. One reason for this is that some equipment is designed to operate most efficiently at certain temperatures or settings.

There is an additional difficulty in experimentally comparing the efficiency of different control schemes. Not all buildings have equipment to directly measure the energy consumption of only HVAC equipment. It is common for energy measurements to include appliances, water heating, and other sources of energy consumption that are difficult to disaggregate from HVAC energy usage. Such disaggregation is hard because the energy signature and characteristics of HVAC equipment is often similar to appliances like refrigerators or water heaters. Another reason that disaggregation is difficult is that installing separate meters to measure energy usage of an HVAC system that is physically and electrically distributed throughout a building can be cost-prohibitive. As a result, experimental comparison may require comparing measurements for which the absolute differences are a small percentage of the total. Determining whether such differences are statistically significant can be challenging.

The purpose of this paper is to propose a set of quantita-

Type	Description
Option A	Measuring key variables that affect HVAC energy consumption
Option B	Directly meausring HVAC energy consumption
Option C	Measuring whole building or sub-building energy consumption
Option D	Estimating HVAC energy consumption using simulation models

Table 1: The International Performance Measurement and Verification Protocol (IPMVP) [6] categorizes approaches to comparing energy usage into four groups.

tive methods for comparing different HVAC control schemes. We define quantitative metrics of efficiency and comfort, and we introduce a mathematical framework for computing these metrics and then determining whether differences in these metrics are statistically significant. For the benefit of the sustainability community, we have provided open source software that implements our comparison methodology in the MATLAB programming language. A case study of experiments on our building-wide testbed shows the utility of the proposed techniques.

2. EXISTING COMPARISON METHODS

The International Performance Measurement and Verification Protocol (IPMVP) [6] classifies approaches to tracking energy usage into four general classes, and it is summarized in Table 1. Though IPMVP is a standard protocol for evaluating energy usage of different building components, here we restrict our discussion to HVAC equipment. One class in IMMPV is called Option A and refers to measuring key variables that affect HVAC energy consumption. An example is measuring the OAT and the SAT settings. Another class is called Option B and involves directly measuring energy usage of the HVAC. Option C is the situation in which the whole building or sub-building energy usage is measured, while Option D involves utilizing simulation models that estimate HVAC energy consumption.

The approach in [3], where the energy consumption of two control schemes on a single-room testbed with central air conditioning were compared, was a combination of Option B and Option D. Mathematical models of the temperature dynamics of different control schemes and their energy characteristics were constructed in order to allow comparisons of experimentally measured HVAC energy usage to simulations over identical weather and occupancy conditions. Unfortunately, this method does not easily scale from the small system considered in [3] to building-sized HVAC systems because of additional complexities of building-sized systems.

In [18], the energy usage of two controllers for a ceiling radiant heating system was compared using an Option A approach. The temperature difference between the OAT and the set point for a loop of hot water was used to evaluate the energy efficiency of different control methods. The general challenge with using Option A is that, depending on the particular HVAC equipment, the energy usage can scale in

complex ways as a function of the key parameters that are measured. Moreover, auxiliary equipment in an HVAC system (e.g., pumps, fans) often significantly contribute to energy usage; such parameters and their relationship to energy consumption are not usually measured or well understood.

Optimization of thermal storage for campus-wide building cooling was considered in [14], and the energy usage of different controllers was compared using an Option B approach coupled with a regression model of baseline performance: This indicates large energy savings when using novel controllers. Such an analysis is only possible when direct energy measurements of the equipment are available; when this is not the case, the differences in measurements of two controllers can often be a small percentage of the total building-wide energy consumption. A regression model alone cannot determine whether such differences are statistically significant.

We believe that a more unbiased approach for comparing HVAC is Option C combined with simple models. Note that our proposed method also applies to Option B, when the HVAC energy usage is directly measured. Experiments on our testbeds indicate a significant impact of OAT on the energy characteristics of HVAC systems. Our proposed method is to use nonparametric modeling methods to compute quantitative metrics of energy usage and occupant comfort, as they relate the OAT. This framework allows for additional nonparametric tools that can determine whether the differences in energy usage and occupant comfort of different HVAC controllers are statistically significant.

3. EXPERIMENTAL SETUP

In order to ensure fair comparisons between different control methods, it is imperative that the experiments be conducted using identical building and HVAC configurations. For the sake of argument, suppose that controller 1 is the manufacturer's configuration, and controller 2 is identical to controller 1 except for that it turns the HVAC off during the night time. Then controller 2 can achieve substantial energy savings, but this does not reflect savings due to control schemes: The energy savings are due to having different configurations of comfort levels.

When comparing HVAC controllers, settings related to occupant comfort should be kept constant because the majority of energy is used in maintaining comfort. The specific settings that need to be kept equal will vary depending upon the building, but important variables to consider include desired temperatures in different building zones, allowable amounts of temperature deviation in these zones, and minimum and maximum amounts of air flow.

There is another experimental issue that is subtle: It is incorrect to keep repeating the analysis as more experimental data is measured, without making corrections to the methodology. The reason is that the probability of making errors accumulates each time the hypothesis testing methodology is used with additional data. Corrections would need to be made to ensure that the accumulated probability of error does not grow too large. This can be done in a principled manner using techniques from sequential hypothesis testing (e.g., [20, 15, 8]); though, we do not discuss in our
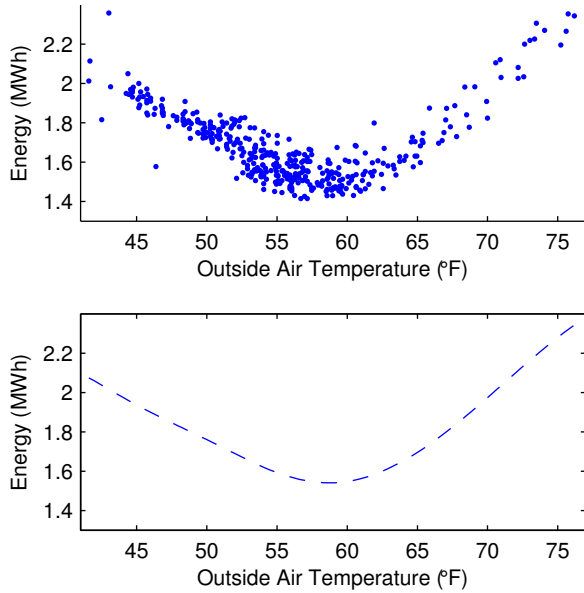
Figure 1: The scatter plot shows the data points $(T_0[i], E[i])$, and the smooth curve is the estimated energy characteristic $\hat{E}_{o,x}(T_0)$ for an HVAC controller.

paper on how to do so. Here, we assume that the analysis is conducted once with a fixed amount of data.

4. MEASURING ENERGY EFFICIENCY

Without loss of generality, we assume that building-wide measurements of energy usage are available at hourly intervals. These values will be denoted as $E[i]$ for $i = 1, \ldots, N$ measurements. Again, note that this framework also applies to the situations where energy usage is measured at daily intervals and where the HVAC energy usage is directly measured. Furthermore, we assume the availability of OAT measurements that correspond to the energy usage measurements: $T_o[i]$ for $i = 1, \ldots, N$.

4.1 Quantifying Energy Consumption

The general model describing the relationship between energy usage, OAT, occupancy O, and other factors (e.g., solar effects, equipment, etc.) X is

$$E = f(T_o, O, X), \qquad (1)$$

where $f(\cdot, \cdot, \cdot)$ is a nonlinear relationship that is unknown. However, occupancy and other factors are generally not directly measured (though their effects on the thermal dynamics and energy usage can be estimated using semiparametric regression [3, 5]). As a result, it is typically only possible to consider the energy usage averaged over occupancy and other factors

$$E_{o,x}(T_o) = \mathbb{E}\big[f(T_o, O, X)\big|T_o\big]. \qquad (2)$$

Equation (2) can be estimated using nonparametric regression [9]. (We use local linear regression in the provided code.) It is not a single value, but is rather a curve that
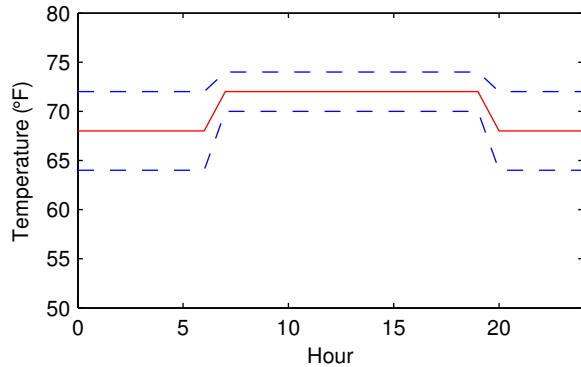

Figure 2: The solid line indicates the set point of a zone throughout a single day, and the extent of the dashed lines describes the band of comfort.

describes the relationship between the OAT and the average energy consumption. Intuitively, if the data points $(T_o[i], E[i])$ for $i = 1, \ldots, N$ represent a scatter plot of energy usage versus OAT; then the estimated curve $\hat{E}_{o,x}(T_o)$ represents the smoothed version of the scatter plot. An illustration of this can be seen in Fig. 1.

The average amount of energy used in one hour is therefore

$$E_g = \mathbb{E}_g\big(E_{o,x}(T_o)\big) = \mathbb{E}_g\big[\mathbb{E}\big[f(T_o, O, X)\big|T_o\big]\big], \qquad (3)$$

where $g(T_o)$ is a probability distribution of OATs. This notation allows us to define the average amount of energy used in one day as $E_{day} = \sum_{i=1}^{24} E_{g_i}$, where $g_i(T_o)$ is the probability distribution of OAT during the i-th hour of the day. For simplicity of calculations, the source code we provide assumes a uniform distribution for the OAT.

4.2 Interpretation of Energy Characteristics

It is well known (e.g., [11]) that the energy characteristics of HVAC have the following qualitative features: Energy usage is lowest at moderate OATs, and the consumption of energy increases as the OAT increases or decreases. Physically, the minimum in the energy characteristic corresponds to the OAT at which the building switches between predominantly heating and predominantly cooling. This can be seen in Fig. 1, which uses measurements taken from our building-wide HVAC testbed. The quantitiative features of the energy characteristics vary depending upon the HVAC controller and other specific characteristics of the building and its weather microclimate.

4.3 Comparing Energy Consumption

We make the assumption that exactly two different control schemes are being compared; such an assumption is not restrictive, because we can do a set of pairwise comparisons with appropriate corrections made for multiple hypothesis testing. The two controllers being compared will be referred to as controller 1 and controller 2, and they will be denoted using a superscript 1 and 2, respectively. Though we could compute the average amount of energy used in one day for each controller, there is no guarantee *a priori* that the difference of these estimated energy characteristics $\hat{E}^1_{o,x}(T_o), \hat{E}^2_{o,x}(T_o)$ is statistically significant
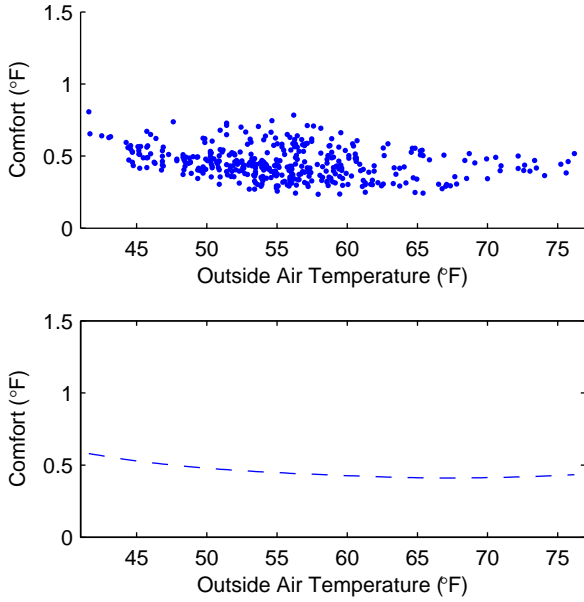
Figure 3: The scatter plot shows the data points $(T_0[i], C[i])$, and the smooth curve is the estimated comfort characteristic $\hat{C}_{o,x}(T_0)$ for an HVAC controller.

Our approach is to use a hypothesis test to quantify the evidence for the statement that the estimated characteristics for the two controllers are equal: $\hat{E}^1_{o,x}(T_o) = \hat{E}^2_{o,x}(T_o)$. If the p-value of this test is less than a significance level α, then we can say that the difference is statistically significant. Otherwise, the difference is not statistically significant if the p-value is greater than α. The code we provide uses a significance level of $\alpha = 0.01$, and note that the intuition is that α gives the probability of incorrectly concluding that a difference is statistically significant.

Because the curves $\hat{E}^1_{o,x}(T_o)$, $\hat{E}^2_{o,x}(T_o)$ are computed using nonparametric methods, it is not possible to use common methods like the t-test to compare them. Nonparametric hypothesis tests based on bootstrap procedures are an attractive alternative for our situation [19]. However, there is a temporal correlation between the measured energy usage $E[i]$ for different values of i. This precludes the use of resampling residuals bootstrap, which is a standard bootstrap method.

The methodology we propose makes use of a nonparametric hypothesis test that utilizes the moving block bootstrap. This form of bootstrap is designed to handle dependent data in which there is a temporal correlation between measurements. The interested reader can refer to [19] for details about this hypothesis test, and this is the technique that is implemented in our code. This is used to determine whether $\hat{E}^1_{o,x}(T_o)$, $\hat{E}^2_{o,x}(T_o)$ are statistically different. We can also use the same methodology to determine whether the difference in the estimated average energy usage over one day $\Delta\hat{E}^{2,1} = \hat{E}^2_{day} - \hat{E}^1_{day}$ is not equal to zero to a statistically significant level. Note that because two hypotheses are being tested, we must make appropriate adjustments [15]:

Figure 4: Sutardja Dai Hall is 141,000-square-foot building, and our BRITE-S testbed is within this building.

Our code uses a Bonferroni correction to generate adjusted p-values [21].

4.4 Confidence Intervals of Energy Consumption

If the hypothesis test in Sect. 4.3 indicates that $\Delta\hat{E}^{2,1} \neq 0$ to a statistically significant level, then we can interpret both the value and magnitude of this quantity. If $\Delta\hat{E}^{2,1} < 0$, then this means that controller 2 uses less energy than controller 1 on average, and the absolute value $|\Delta\hat{E}^{2,1}|$ is the average amount of energy savings over a day due to controller 2 as compared to controller 1. The opposite statement holds if $\Delta\hat{E}^{2,1} > 0$.

Since $\Delta\hat{E}^{2,1}$ is an estimated quantity, it will itself have uncertainty. In order to better characterize the difference between the energy consumption of controllers 1 and 2, it is useful to also compute a confidence interval for $\Delta\hat{E}^{2,1}$. For a confidence level of $\beta \in (0, 1]$, the confidence interval contains the true quantity (e.g., $\Delta E^{2,1}$) for β fraction of the experiments. The same moving block bootstrap methodology used in Sect. 4.3 can be used to estimated a bias-corrected bootstrap confidence interval [7].

5. MEASURING OCCUPANT COMFORT

Quantifying the comfort levels of occupants is difficult because it is a function of many variables: metabolic rate, clothing insulation, air temperature, radiant temperature, air speed, and humidity [1]. This problem is further complicated by the fact that most buildings are not instrumented to measure these variables.

In order to define a quantification of comfort that is both tractable and will scale to many buildings, we focus on a measure that is only dependent on temperature; temperature measurements are usually available through existing thermostats. We assume that the average temperature for each of the $j = 1, \ldots, Z$ zones are measured at hourly intervals $T_j[i]$ for $i = 1, \ldots, N$.

5.1 Band of Comfort

Controller 2	Energy Characteristics	$\Delta \hat{E}^{2,1}$	Comfort Characteristics	$\Delta \hat{C}^{2,1}$
Temperature and Reheat Schedule	Statistically Different ($p = 0.002$)	-1.16MWh ($p = 0.08$)	Statistically Different ($p = 0.002$)	$0.35^\circ F$ ($p = 0.9$)
LBMPC SAT Control	Statistically Different ($p = 0.002$)	-1.53MWh ($p = 0.002$)	**NOT** Statistically Different ($p = 0.8$)	$-0.75^\circ F$ ($p = 0.5$)

Table 2: Summary of case studies on the BRITE-S testbed. Controller 1 is the default, manufacturer-provided controller, and controller 2 is as indicated. Also, $\alpha = 0.05$ was taken to be the significance level of the hypothesis tests.

The j-th zone of a building will have a temperature set point S_j that can vary throughout the day, and we assume that we have recorded the average value of this set point at hourly intervals: $S_j[i]$ for $i = 1, \ldots, N$. Furthermore, we assume that comfort is defined as maintaining zone temperature within $\pm B_j[i]$ for $i = 1, \ldots, N$ of the set point. It can be visualized as a band of comfort, as shown in Fig. 2. This band is allowed to vary because the configured comfort levels often depend on the time of day or week; an example is allowing greater temperature fluctuations in an office building at night.

Our justification for this notion of comfort is that it corresponds to the Class A, Class B, and Class C notions of comfort as defined in [1]. Furthermore, this quantification is implicitly making the assumption that the set point indicates the true occupant preference. It represents a compromise between tractability of quantifying comfort and maintaining a reasonable metric.

5.2 Quantifying Occupant Comfort
Let $(x)_+$ be the thresholding function, which is defined so that $(x)_+ = 0$ if $x < 0$ and $(x)_+ = x$ otherwise. We define our quantification of comfort using soft thresholding as

$$C = 1/Z \cdot \sum_{j=1}^{Z} \int_0^1 (|T_j - S_j| - B_j)_+ dt, \qquad (4)$$

where the integral with respect to dt is over one hour of time. The intuition is that this quantity increases whenever a zone temperature leaves the band of comfort, and the amount of increase in this quantity is proportional to the amount and duration of temperature deviation.

Because T_j is implicitly a function of the outside temperature T_0, occupancy O, and other factors X, we can abstractly represent the comfort quantification as

$$C = h(T_0, O, X), \qquad (5)$$

where $h(\cdot, \cdot, \cdot)$ is a nonlinear relationship that is unknown. Just as in the case of energy, we consider the comfort averaged over occupancy and other factors

$$C_{o,x}(T_0) = \mathbb{E}[h(T_0, O, X)|T_0], \qquad (6)$$

because they are not measured. We can again generate a scatter plot of the data points $(T_o[i], C[i])$ for $i = 1, \ldots, N$, where $C[i] = \sum_j (|T_j[i] - S_j[i]| - B_j[i])_+$. And a smoothed version can be estimated $\hat{C}_{o,x}(T_0)$ using nonparametric regression. An example of this is seen in Fig. 3.

The average occupant comfort over one hour is given by

$$C_g = \mathbb{E}_g(C_{o,x}(T_0)) = \mathbb{E}_g[\mathbb{E}[h(T_0, O, X)|T_o]], \qquad (7)$$

where $g(T_o)$ is a probability distribution of OATs. We can then define the average amount of comfort in one day as $C_{day} = \sum_{i=1}^{24} C_{g_i}$, where $g_i(T_o)$ is the probability distribution of OAT during the i-th hour of the day. Our source code assumes a uniform distribution for the OAT.

5.3 Interpretation of Comfort Characteristics
The purpose of an HVAC system is to provide uniform levels of occupant comfort regardless of external and internal conditions, and so the ideal HVAC controller will provide a comfort characteristic that is constant across all OATs. However, in practice it is more difficult for the HVAC to maintain the building environment when the weather is more extreme. This leads to slight decreases in comfort (i.e., the comfort characteristic is higher) at low and high OATs. This can be seen in Fig. 3, which uses measurements taken from our building-wide HVAC testbed. The quantitative features of the comfort characteristics vary depending upon the HVAC controller and other specific characteristics of the building and its weather microclimate.

5.4 Comparing Occupant Comfort
Using the hypothesis testing methodology outlined in Sect. 4.3, we can determine (a) whether the estimated comfort characteristics $\hat{C}_{o,x}^1(T_o), \hat{C}_{o,x}^2(T_o)$ for the two controllers are statistically different, and (b) whether the difference in the estimated average comfort over one day $\Delta \hat{C}^{2,1} = \hat{C}_{day}^2 - \hat{C}_{day}^1$ is not equal to zero to a statistically significant level. Like in Sect. 4.3, we must correct for multiple comparison effects.

5.5 Confidence Intervals for Occupant Comfort
If the hypothesis test in Sect. 4.3 indicates that $\Delta \hat{C}^{2,1} \neq 0$ to a statistically significant level, then we can interpret both the value and magnitude of this quantity. Because lower values of C correspond to greater comfort levels, we can interpret the quantity $\Delta \hat{C}^{2,1}$ in the same way as $\Delta \hat{E}^{2,1}$ is interpreted in Sect. 4.4. And because $\Delta \hat{C}^{2,1}$ is an estimated quantity, the difference between the comfort levels of controllers 1 and 2 can be better characterized by also computing its confidence interval as in Sect. 4.4.

6. CASE STUDY: BRITE-S TESTBED
The Berkeley Retrofitted and Inexpensive HVAC Testbed for Energy Efficiency in Sutardja Dai Hall (BRITE-S) platform [12, 4] is a building-wide HVAC system that maintains the indoor environment of a 141,000-square-foot building, shown in Fig. 4, that is divided between a four-floor
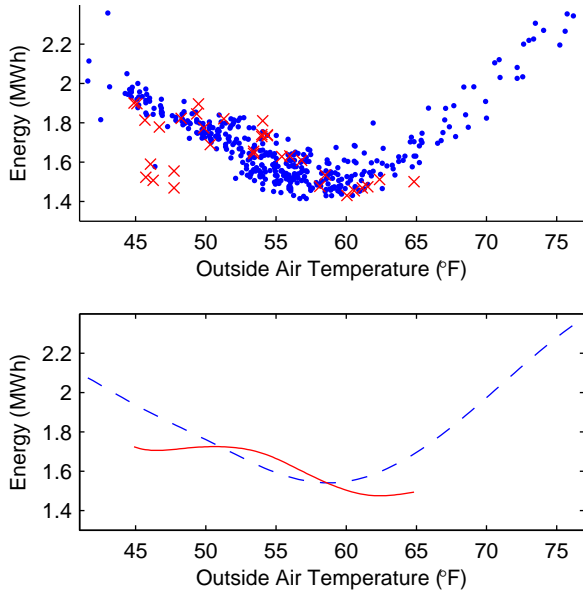
Figure 5: The cross marks and solid line (points and dashed line) denote the energy characteristics of the schedule (default) controller.

Figure 6: The cross marks and solid line (points and dashed line) denote the comfort characteristics of the schedule (default) controller.

nanofabrication laboratory (NanoLab) and seven floors of general space (including office space, classrooms, and a coffee shop). The building automation equipment can be measured and actuated through a BACnet protocol interface.

The HVAC system uses a 650-ton chiller to cool water. Air-handler units (AHUs) with variable-frequency drive fans distribute air cooled by the water to variable air volume (VAV) boxes throughout the building. Since the NanoLab must operate within tight tolerances, our control design can only modify the operation of the general space AHUs and VAV boxes, with no modification of chiller settings that are shared between the NanoLab and general space.

The default, manufacturer-provided controller in BRITE-S uses PID loops to actuate the VAV boxes and keeps a constant SAT within the AHUs. Conventional SAT reset control is not possible because several VAV boxes for zones with computer equipment provide maximum air flow rates throughout the entire day for all, except the coldest SATs.

Here, we present case studies that compare this default controller (controller 1) with two different controllers. Controller 2 will refer to the controller being compared to controller 1. Note that the building is configured to provide a band of comfort of $\pm B_j[i] = \pm 1°F$ for all times and zones. The results of these case studies are summarized in Table 2, and details are given below.

6.1 Temperature and Reheat Schedule

The first controller to be compared to controller 1 turns off all heating in the building during the night time, and it matches the SAT to the OAT in an effort to reduce the amount of energy used to cool the air. We call this the schedule controller. An Option A analysis that neglects auxiliary
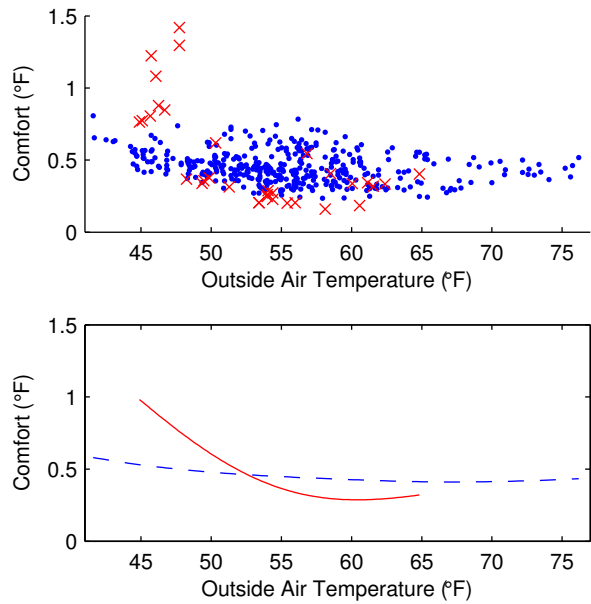
equipment like fans will estimate nearly zero energy usage. However, our comparison methodology shows that this is not correct.

The estimated energy characteristics $\hat{E}_{o,x}^1(T_0), \hat{E}_{o,x}^2(T_0)$ are shown in Fig. 5, and their differences are statistically significant ($p = 0.002$). However, the estimated difference in average energy usage over one day $\Delta\hat{E}^{2,1} = -1.16$MWh is not statistically significant ($p = 0.08$), meaning that there is not enough evidence to exclude that $\Delta E^{2,1} = 0$MWh.

The estimated comfort characteristics $\hat{C}_{o,x}^1(T_0), \hat{C}_{o,x}^2(T_0)$ are shown in Fig. 6, and their differences are statistically significant ($p = 0.002$). However, the estimated difference in average comfort over one day $\Delta\hat{C}^{2,1} = 0.35°F$ is not statistically significant ($p = 0.9$), meaning that there is not enough evidence to exclude that $\Delta C^{2,1} = 0°F$.

Though the schedule controller saves energy by turning off heating at night, these savings are somewhat negated by having to heat the building to a comfortable temperature during the day (i.e., the hump in $\hat{E}_{o,x}^2(T_0)$ at $55°F$). As compared to the default controller, the schedule controller substantially degrades comfort at colder temperatures and provides slight improvements in comfort at moderate temperatures.

6.2 LBMPC Control of Supply Air Temperature

The second controller to be compared to controller 1 uses a hybrid system learning-based model predictive controller [2, 4] to determine a sequence of SATs amongst the values of $52°F$, $58°F$, and $64°F$. An Option A analysis provides overly optimistic estimates of the energy savings due to this con-
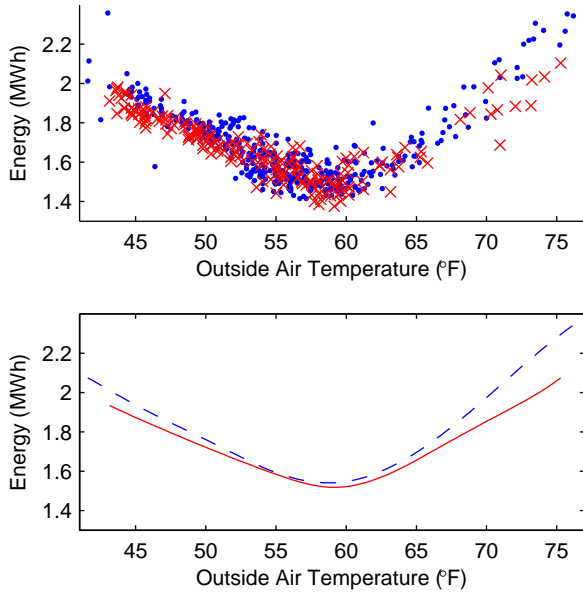
Figure 7: The cross marks and solid line (points and dashed line) denote the energy characteristics of the LBMPC (default) controller.

Figure 8: The cross marks and solid line (points and dashed line) denote the comfort characteristics of the LBMPC (default) controller.

trol method, while the methodology described in this paper shows modest energy savings.

The estimated energy characteristics $\hat{E}_{o,x}^1(T_0), \hat{E}_{o,x}^2(T_0)$ are shown in Fig. 7, and their differences are statistically significant ($p = 0.002$). The estimated difference in average energy usage over one day $\Delta\hat{E}^{2,1} = -1.53$MWh is statistically significant ($p = 0.002$). And the 95% confidence interval is $\Delta\hat{E}^{2,1} \in [-2.07, -1.02]$MWh.

The estimated comfort characteristics $\hat{C}_{o,x}^1(T_0), \hat{C}_{o,x}^2(T_0)$ are shown in Fig. 8, and their differences are not statistically significant ($p = 0.8$). Furthermore, the estimated difference in average comfort over one day $\Delta\hat{C}^{2,1} = -0.75°$F is not statistically significant ($p = 0.5$), meaning that there is not enough evidence to exclude that $\Delta C^{2,1} = 0°$F.

The LBMPC controller provides modest energy savings at most OATs, which sum up to significant savings over a day. And because the difference in comfort characteristics is not statistically significant, this suggests that the LBMPC and default controllers provide comparable levels of comfort.

7. CONCLUSION
We have presented quantitative metrics for comparing the energy usage and comfort of different HVAC controllers. These metrics can be computed relatively easily for a wide variety of buildings and are amenable to methods that can determine whether experimental differences between controllers are statistically significant. Though our focus here has been on energy-efficient HVAC control, this methodology can be used to compare the efficiency of different building components such as water heating, lighting, and even changes in the comfort level of HVAC.
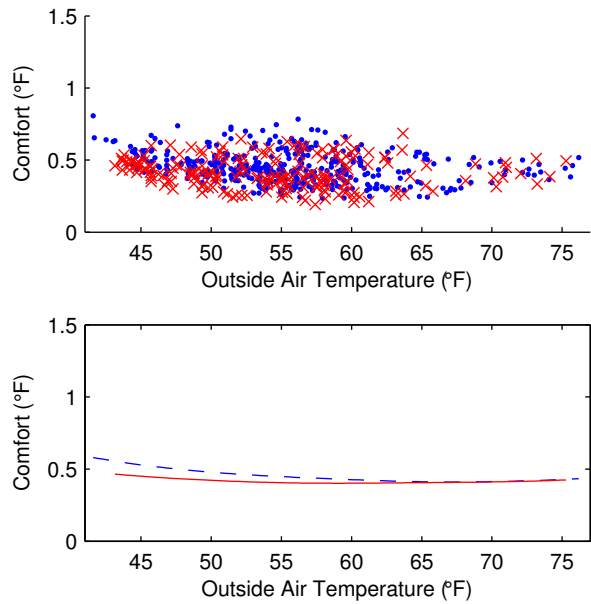
8. ACKNOWLEDGMENTS

9. REFERENCES
[1] ASHRAE 55. Thermal environmental conditions for human occupancy. ANSI approved, 2010.
[2] A. Aswani, H. Gonzalez, S. Sastry, and C. Tomlin. Provably safe and robust learning-based model predictive control. Submitted, 2011.
[3] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin. Reducing transient and steady state electricity consumption in HVAC using learning-based model predictive control. *Proceedings of the IEEE*, 100(1):240–253, 2011.
[4] A. Aswani, N. Master, J. Taneja, D. Culler, and C. Tomlin. Energy-efficient building HVAC control using hybird system LBMPC. Submitted, 2012.
[5] A. Aswani, N. Master, J. Taneja, V. Smith, A. Krioukov, D. Culler, and C. Tomlin. Identifying models of HVAC systems using semiparametric regression. In *Proceedings of the American Control Conference*, 2012.
[6] Efficiency Valuation Organization. International performance measurement and verification protocol. Concepts and Options for Determining Energy and

Water Savings, 2012.

[7] B. Efron. Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics*, 9(2):139–158, 1981.

[8] B. Ghosh and P. Sen. *Handbook of Sequential Analysis*. CRC Press, 1991.

[9] L. Györfi. *A distribution-free theory of nonparametric regression*. Springer, 2002.

[10] A. Kelman and F. Borrelli. Bilinear model predictive control of a HVAC system using sequential quadratic programming. In *IFAC World Congress*, 2011.

[11] J. Kissock, J. Haberl, and D. Claridge. Development of a toolkit for calculating linear, change-point linear, and multiple-linear inverse building energy analysis models. Technical Report ASHRAE Research Project 1050-RP, Energy Systems Laboratory, Texas A&M University, 2002.

[12] A. Krioukov, S. Dawson-Haggerty, L. Lee, O. Rehmane, and D. Culler. A living laboratory study in personalized automated lighting controls. In *ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2011.

[13] C. Liao, Y. Lin, and P. Barooah. Agent-based and graphical modeling of building occupancy. *Journal of Building Performance Simulation*, 5(1):5–15, 2012.

[14] Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves. Model predictive control for the operation of building cooling systems. *IEEE Transactions on Control Systems Technology*, 20(3):796–803, 2012.

[15] R. Miller. *Simultaneous statistical inference*. Springer-Verlag, 1981.

[16] T. Nghiem and G. Pappas. Receding-horizon supervisory control of green buildings. In *Proceedings of the American Control Conference*, pages 4416–4421, 2011.

[17] F. Oldewurtel, A. Parisio, C. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27, 2012.

[18] J. Siroky, F. Oldewurtel, J. Cigler, and S. Privara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88:3079–3087, 2011.

[19] J. Vilar-Fernández, J. Vilar-Fernández, and W. González-Manteiga. Bootstrap tests for nonparametric comparison of regression curves with dependent errors. *TEST*, 16:123–144, 2007.

[20] A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

[21] S. Wright. Adjusted p-values for simultaneous inference. *Biometrics*, 48:1005–1013, 1992.