NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Control Limits for Building Energy End Use Based on Engineering Judgment, Frequency Analysis, and Quantile Regression

Gregor P. Henze, Shanti Pless, Anya Petersen, and Nicholas Long
National Renewable Energy Laboratory

Alexander T. Scambos
University of Colorado at Boulder

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Technical Report
NREL/TP-5500-60020
February 2014

Contract No. DE-AC36-08GO28308

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Control Limits for Building Energy End Use Based on Engineering Judgment, Frequency Analysis, and Quantile Regression

Gregor P. Henze, Shanti Pless, Anya Petersen, and Nicholas Long
National Renewable Energy Laboratory

Alexander T. Scambos
University of Colorado at Boulder

Prepared under Task No. 2940.1404

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

Technical Report
NREL/TP-5500-60020
February 2014

Contract No. DE-AC36-08GO28308

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at http://www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Acknowledgments

This report was prepared by the Commercial Buildings Research Group of the National Renewable Energy Laboratory (NREL).

The authors extend their thanks to the following NREL colleagues:

Larry Brackney for initiating the project and providing scope and context, Luigi Gentile Polese for providing a wide range of project support and oversight, Michael Sheppy for preparing and furnishing the energy consumption data, Alex Swindler for assisting with Ruby programming, Jennifer Scheib for contributing to previous versions of this effort, and to Stefanie Woodward for editorial assistance.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Executive Summary

Background

Commercial buildings need approaches that continuously characterize energy performance to (1) provide timely responses to excess energy use by building operators; and (2) enable building occupants to develop energy awareness and to actively reduce energy use. Energy information systems, many of which involve graphical dashboards, are gaining popularity for presenting (near) real-time energy performance metrics to occupants and operators. NREL researchers developed Building Agent, which incorporates a dashboard, for these purposes.

Each building is, by virtue of its purpose, location, and construction, unique. Thus, building energy performance can be assessed in a relative sense only, as comparison of absolute energy use out of context is not meaningful. In some cases, performance can be judged relative to the average performance of comparable buildings. However, in high-performance building designs, such as NREL's Research Support Facility (RSF), relative performance is meaningful only when it is compared to the facility's historical performance or to its theoretical maximum performance as estimated through detailed energy modeling.

Development Process and Results

The report presents three different but related approaches to determine performance bounds on building energy performance for display on an energy information system:

- Engineering judgment-informed curve fitting
- Frequency of occurrences
- Quantile regression.

Electricity consumption for cooling provided by chilled water was considered the energy end use of choice to illustrate the process. Yet, other energy end uses such as heating, lighting, plug loads, mechanical systems (fans and pumps), and even photovoltaic power generation may be analyzed identically.

In the first approach, using engineering judgment and manual curve fitting, end-use control limits were derived using static lookup-based performance targets, combined with polynomial curve-fit models. These control limits were developed based on a combination of historical end-use data and as-built energy modeling; single main effect variables were selected for each end use, such as solar irradiance for lighting performance.

The RSF uses hot water and chilled water from the campus district energy plant. Hot water and chilled water meters measure the RSF's thermal energy use and chilled water efficiency (kW/ton). Natural gas boiler efficiencies are applied to the metered thermal energy use to calculate system energy end uses. For heating and cooling end uses, the outdoor air dry-bulb temperature is the dominant environmental variable. NREL researchers compared the modeled hourly cooling energy use to the outdoor dry-bulb temperature to develop minimum and maximum control limits through visual inspection for occupied and unoccupied hours. Linear curve fits based on the modeled cooling performance were then provided to the Building Agent (Figure ES–1) and cooling acceptance ranges were calculated and compared to measured cooling energy use. The heating control limits were derived with similar methods based on outdoor dry-bulb temperature as the single main effect variable.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

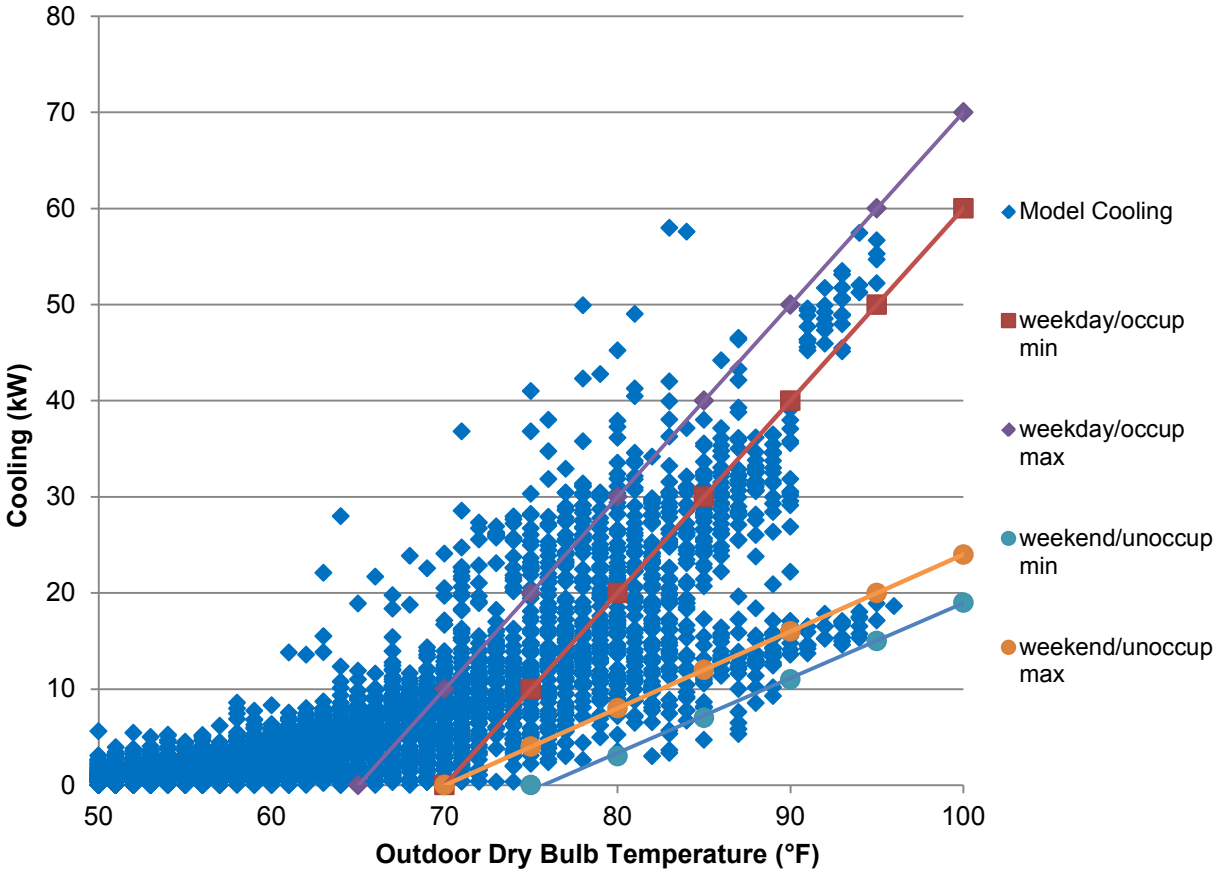

Figure ES–1 Modeled hourly cooling with weekday and weekend control limits

These first-order curve fits for acceptable control limits (based on the as-built model prediction for cooling energy use) successfully predicted actual cooling energy use (Figure ES–2) for RSF I for 2011. Setting end-use budgets, and tracking actual energy use in real time against these budgets, have enabled us to realize net-zero design expectations for the RSF.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

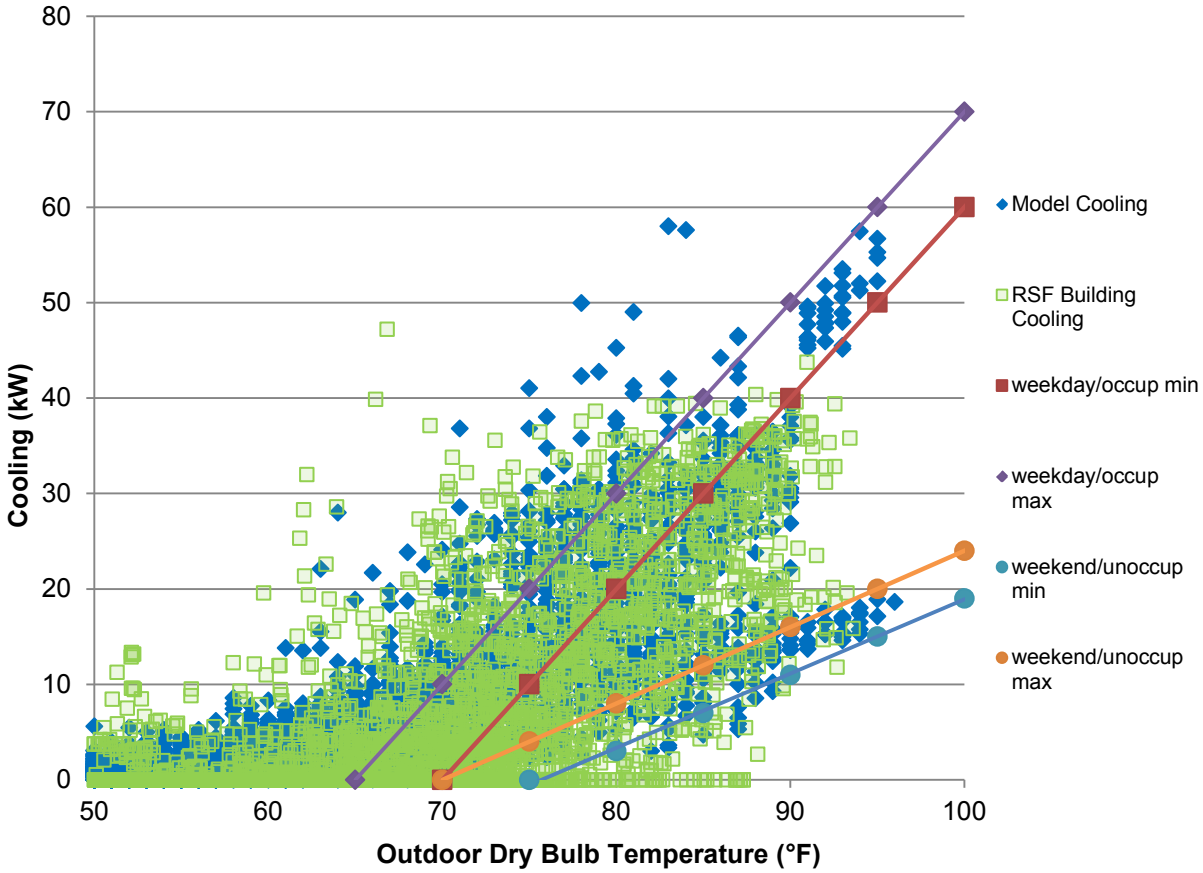

Figure ES–2 Measured and modeled hourly cooling

This manual curve fit process provided first-order control limits; however, this visual inspection method has limitations and an automated process is valuable. Cooling energy use is not always best predicted based on occupancy state and outdoor dry-bulb temperature; outdoor humidity and solar gains can also contribute significantly to cooling energy use. These limitations suggest the need for developing better methods for automating these control limit curve fits and including multiple dominant variables. Descriptions of these two enhanced methods and the improvements are presented here.

In the second frequency-based approach, measured building end use energy data are used to develop empirical frequency distributions for each end use as a function of the dominant independent variable, which is determined from visual inspection and a pairwise correlation analysis. Appropriate control limits were defined as ranges in which a certain percentage of end-use consumption—here the central 50% of observations between the 25th and 75th percentiles—has been observed historically as a function of the dominant independent variable (Figure ES–3. The box plots shown in this report adopt the common notation that the box occupies the interquartile range (IQR) from the lower (25th percentile) to the upper (75th percentile). The whiskers extend to the minimum and maximum values if these are less than 1.5 times the IQR below the lower or 1.5 times the IQR above the upper quartile. Larger and smaller values, respectively, are shown as outliers.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

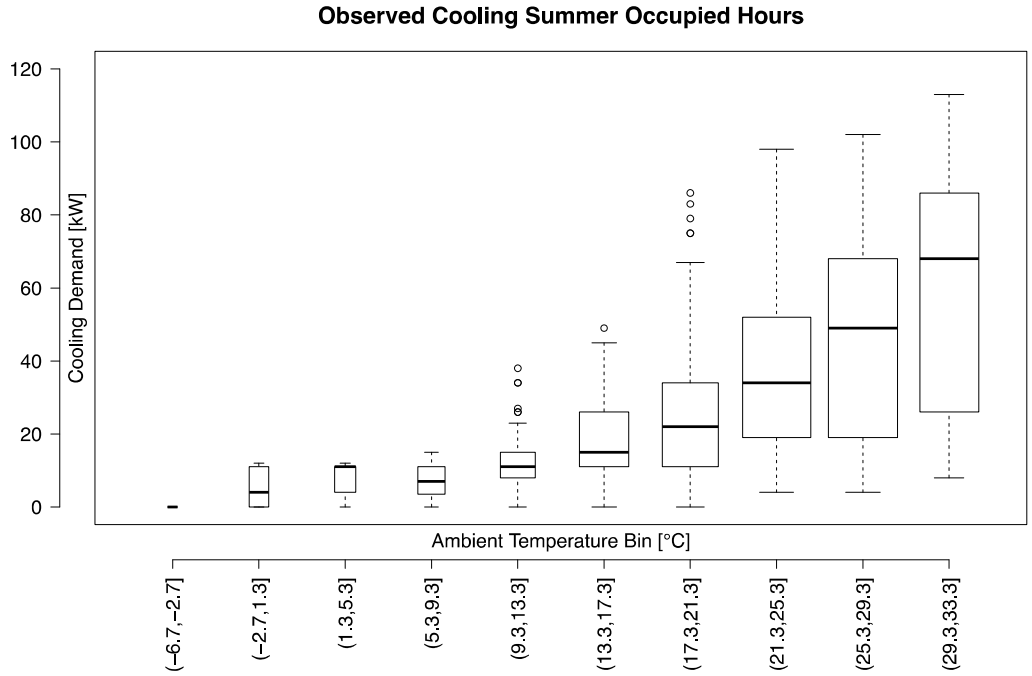

Figure ES–3 Measured cooling energy end use versus ambient temperature for occupied summer hours

The same process is then applied to simulated energy end-use data sourced from a calibrated building energy model (Figure ES–4.

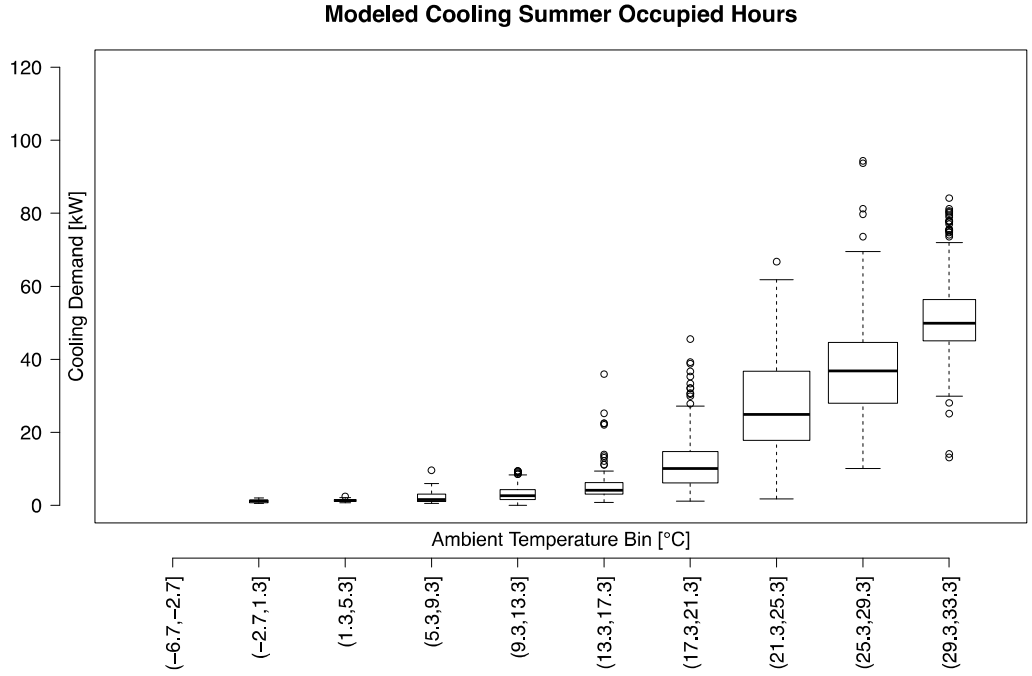

Figure ES–4 Modeled cooling energy end use versus ambient temperature for occupied summer hours

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Two sets of upper control limits (UCLs) and lower control limits (LCLs) are generated from measured and simulation-based data. An energy-conservative control logic is then applied to select one set of control limits in each situation: The model's suggested range is considered invalid whenever the lower observed limit is greater than the upper model limit or when the upper observed limit is smaller than the lower model limit. The model is considered valid in all other cases. If the model is invalid, the observed range is applied. Presuming the model can be considered valid, the LCL is the lesser of either the lower model or lower observed limit and the UCL is the lesser of either the upper model or upper observed limit.

This set of LCLs and UCLs for each energy end use was adopted in Building Agent for presentation to building operators and occupants (see Table ES–1) for the case of cooling energy end use.

Table ES–1 Cooling Energy End-Use LCLs and UCLs (kW) Versus Ambient Temperature (°C) for Occupied Hours

Bin	(–17, –13)	(–13, –9)	(–9, –5	(–5, –1)	(–1, 3)	(3, 7)	(7, 11)	(11, 15)	(15, 19)	(19, 23)	(23, 27)	(27, 31)	(31, 35)
1st Quartile	0	0	0	0	0	0	0	0	2	7	11	22	20
3rd Quartile	0	0	0	0	0	2	2	3	7	20	40	49	78

The control limit tables shown in Table ES–1 were implemented in the Building Agent's dashboard system. Figure ES–5 shows the instantaneous cooling energy end use and Figure ES–6 shows the historical development. The gray band in Figure ES–5 and Figure ES–6 is defined by the LCLs andUCLs developed in this report. Although the realities of actual building control are not represented in the mode (Figure ES–6), the general performance trends are captured sufficiently well.

Figure ES–5 Dashboard view of instantaneous total cooling power (kW) in the RSF along with LCLs and UCLs shown as a gray band above

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

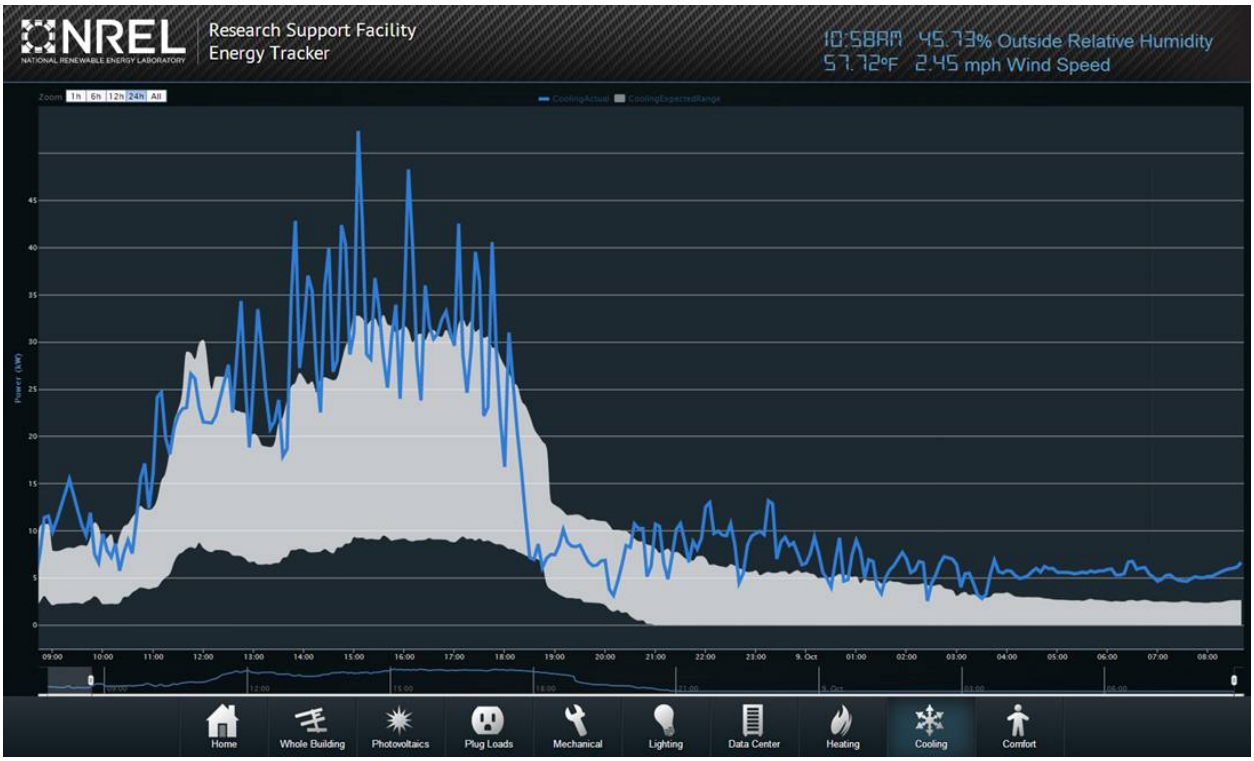

Figure ES–6 Dashboard view of historical total cooling power (kW) (blue) development in the RSF along with control limits

A more universal methodology that could harness *all* the available covariates of an energy end use was desired to predict UCLs and LCLs as a function of all the relevant independent variables instead of just the dominant one. In essence, a methodology was needed that can be applied to any situation where control limits on energy end use were needed, whether a sophisticated monitoring system with dozens of available variables or a system with only a few monitored points.

Therefore, the third quantile regression-based approach, the building energy end use, was modeled linearly in terms of other end uses, exogenous variables, and weather data, all of which were transformed to orthogonal principal components. The linear models were used to generate control limits using quantile regression of measured and simulated data. In this report, quantile regression was used to predict the 40^{th} and 60^{th} quantiles. The same energy-conservative control logic was then applied to the two sets of quantile regression-based control limits (Figure ES–7) for a summer week, where LCL is the lower control limit, UCL is the upper control limit, and "Data" is the observed cooling consumption. Figure ES–7 shows that the actual consumption is often higher than the 60^{th} percentile, indicating that the actual consumption is on the high side of the expectation.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

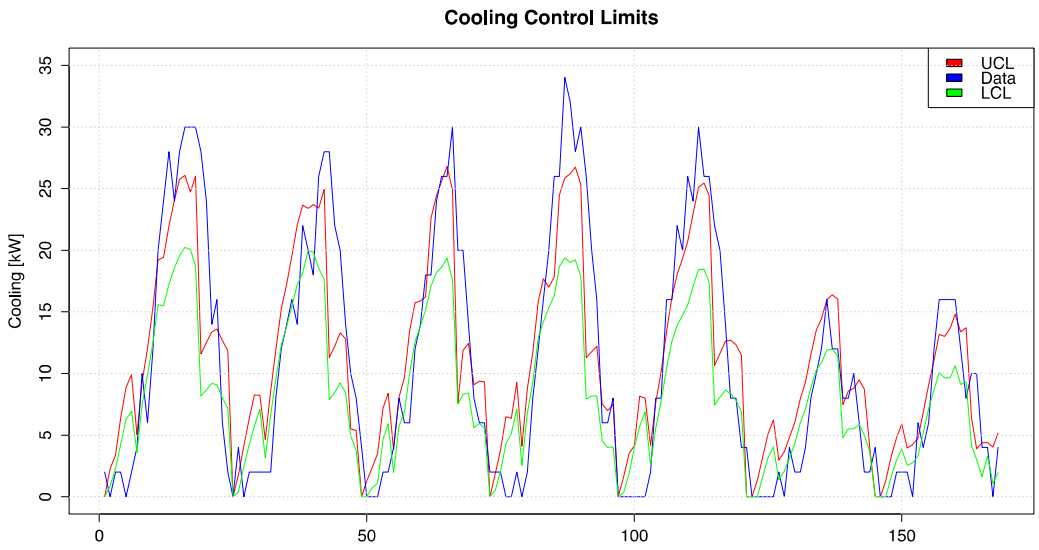

Figure ES–7 Summer week cooling UCLs (red) and LCLs (green) along with measured cooling end use versus hour of the week

Because dashboard users can view an expected range of energy end use consumption in addition to the current instantaneous value, they can draw conclusions about the relative energy intensity. If the current value falls within the expected range, the building energy end use can be considered acceptable, meaning it requires no action from facility operators or occupants. Alternatively, if the energy use exceeds the UCL, the dashboard indicates opportunities for energy savings. The greater the departure, the more significant the savings potential will likely be.

The first approach shows engineering judgment-based control limits; the second and third approaches present more energy-conservative control limits based on comparisons of modeled and observed data. Each method encourages occupants and operators to maintain the building's energy performance in a more energy-conservative state.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Nomenclature

BIC	Bayesian Information Criterion
COP	coefficient of performance
CSV	comma separated values file
DOE	U.S. Department of Energy
DW	Durbin-Watson statistic
ESIF	Energy Systems Integration Facility
EUI	energy use intensity
LCL	lower control limit
LPD	lighting power density
LQM	lower quartile modeled
LQO	lower quartile observed
NREL	National Renewable Energy Laboratory
OLS	ordinary least squares
PC	principal component
PCA	principal component analysis
PV	photovoltaic
Q-Q plot	quantile-quantile plot
RSF	Research Support Facility
TMY3	Typical Meteorological Year 3 weather data
UCL	upper control limit
UQM	upper quartile modeled
UQO	upper quartile observed
WE	weekend
WD	weekday

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contents

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Figures and Tables

Figures

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Tables

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

1.0 Introduction

1.1 Motivation

As net-zero energy building procurement matures, the endeavor expands from realizing aggressive energy goals in design to maintaining designed performance over the life of the building. Changes in occupancy patterns, miscellaneous loads, space use type, and installed equipment performance can significantly impact year-to-year energy performance. In one possible building lifetime scenario, energy use increases over time as equipment is added or operational schedules change. In another, energy use decreases over time as equipment degrades or fails, typically at the expense of occupant comfort. The ideal scenario is that equipment is maintained and upgraded appropriately. These actions reduce energy consumption and maintain occupant comfort for each year of operation (accounting for variations in weather and other external factors).

The underlying principal shortcoming is a disconnect between the performance expectations and predictions created during the building's design and the reality that unfolds during building operation. Assuming realistic expectations for energy consumption were set during the design process—based on the emerging understanding of predictive energy modeling—the charge at hand is to first establish and then continuously monitor budgets for each end use. To accomplish this charge, an interface between the building, the occupants, and the operational staff is needed for fact-based, proactive decision-making.

Like energy models in design, energy dashboards have surfaced as the primary interfaces between the building and the decision-makers. Many dashboard visualizations are paired with a single energy goal, such as net-zero energy, which is sufficient if no problems are encountered; however, this approach does not provide direction for corrective action if the goal is not met. The premise of this research is to develop a dashboard that does not oversimplify the design-based energy model results but rather disentangles them for ongoing analysis by building decision-makers. The result is a more complex dashboard architecture that provides a deeper understanding of the building's operation, as illustrated through the U.S. Department of Energy's (DOE) National Renewable Energy Laboratory (NREL) campus dashboard and occupant interface, the Building Agent.

1.2 NREL Campus Background

The DOE/NREL goal for the campus is to expand its leadership as a state-of-the-art laboratory that supports innovative research, development, and commercialization of renewable energy and energy efficiency technologies that address the nation's energy and environmental needs. Its recent growth has resulted in a significant increase in employees and facilities on its 327-acre (1.32 km^2) main campus in Golden, Colorado.

To support this growth over the last 5 years, NREL developed and demonstrated new construction procurement methods that proved cost effective and showed that 50% energy savings over typical building code are possible when design-build teams integrate to achieve specific and measurable whole-building energy goals. NREL facility growth provided an opportunity to demonstrate the integrated approach in real projects by incorporating energy performance specifications into the projects' contracts, as documented through predictive modeling. NREL developed and piloted this energy performance-based design-build process in 2008 with the Research Support Facility (RSF I), an 824-occupant, 220,000-ft^2 (20,500 m^2)

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

office building. The process has since been replicated on other campus projects such as an office wing expansion, the Research Support Facility II (RSF II), the Energy Systems Integration Facility (ESIF), a parking structure, and a site entrance building.

Each project employed contractual energy use requirements in the design-build contracts to incorporate world-class efficiency strategies, all on typical DOE construction budgets. In addition to general energy reduction goals such as a 50% reduction versus typical building code and sustainability goals such as Leadership in Energy and Environmental Design Platinum certification, the contracts included specific and directly measurable site energy use requirements:

- RSF I: 36 kBtu/ft^2/yr (11 kWh/m^2/yr) site energy use intensity (EUI) and net-zero energy, including the data center
- RSF II: 33 kBtu/ft^2/yr (10 kWh/m^2/yr) site EUI and net-zero energy
- ESIF: 27 kBtu/ft^2/yr (9 kWh/m^2/yr) site EUI for the office space, and 1.06 power usage effectiveness for the data center
- Parking structure: 175 kBtu/parking space/yr (51 kWh/parking space/yr)
- Site entrance building: 32 kBtu/ft^2/yr (10 kWh/m^2/yr) EUI and net-zero energy.

Each project's design-build team successfully designed and constructed these buildings to reach the contractual energy goals, substantiated by energy models and submetering during the first year of operation. Based on the successful first year of operation vis-à-vis the performance specification, additional monetary rewards were paid to the design-build team in excess of the base contract value as an incentive for the commitment to this novel delivery process. NREL terms this contractual energy goal substantiation effort, combined with energy goal maintenance over the life of the building, *energy performance assurance*.

1.3 Energy Performance Assurance

In each of NREL's recent campus construction projects, whole-building energy models were started in the predesign phase by each proposing design-build team to prove that its design concepts were likely to meet the contractual goals (Pless et al. 2011). The design-based energy models evolved over the design phases and were updated based on constructed reality. This energy performance assurance task differs from a typical design process in that information from the design-based energy model was used to make iterative decisions about the building design and to understand cost and energy performance tradeoffs versus simply verifying that design concepts perform as expected at the end of design.

The first step in energy performance assurance is to advance energy model applications throughout the design and construction industry. Specifically, this means continually refining building geometry, internal loads, weather input, occupancy schedules, etc. For the NREL campus, energy modeling reports were provided at each phase of design and as energy-cost tradeoff questions arose, with clearly identified variables such as input schedules that the integrated owner, design, and construction team could discuss. The final advancement of the energy modeling process resulted in an as-built energy model that forms the basis of the operational energy performance expectations.

The second step is submetering. This requires the team to consider the electrical design in order to lay out meters appropriately to make end-use disaggregation available for validating the

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

design-based energy models. For each NREL project, submetering was an explicit contractual requirement, beyond the implicit contractual need to verify the energy goal at occupancy.

These first two steps are critical to setting up a building-human interface such as a dashboard that allows for ongoing, proactive decision-making and energy end use budget tracking.

The final step—and the focus of this work—is to set up such a building-human interface. This effort includes defining the system architecture for data collection and designing system visualizations that allow humans—including occupants, facility managers, information technology managers, and owners—to make evidence-based and goal-driven decisions.

NREL's building-human interface, developed in tandem with the recent campus construction, is called Building Agent. The following Building Agent description focuses on the RSF (combination of RSF I and II with a total of 360,000 ft^2 or 33,600 m^2), because the building has been operational for more than 1 year and provides a fuller picture of the effort required to maintain energy performance. The RSF was the living laboratory for Building Agent development and continues to be the platform for further enhancements.

This work builds on Long et al. (2013) in that a wider range of approaches to determining the energy end-use control limits is developed and implemented.

1.4 Comments on Units

Please note that SI units are used for all data analysis and IP units are used for public presentation of the dashboard.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

2.0 Building Agent System Design

The overarching aim for the Building Agent (Schott et al. 2012) is to create cohesion between building automation systems, local energy use measurements, and occupant feedback to provide visualizations that empower occupants, facility managers, and building engineers to take diagnostic, proactive, energy-saving, and comfort-improving actions. The Building Agent architecture consists of central server, measurement, and distributed application and visualization layers.

The visualization scope of the Building Agent extends beyond a typical energy dashboard: It collects and displays energy performance data and allows facility managers and typical occupants to analyze and investigate the balance of energy use, energy expenditures, and comfort. The later portion of this report emphasizes the analysis underlying the Building Agent visualization.

2.1 Building Agent Architecture

The Building Agent architecture consists of four layers: hardware and protocols, databases, applications, and visualizations. Figure 2–1 defines the layer components and color-codes them according to those developed uniquely for Building Agent ("Building Agent" and "Visualizations") and shared campus resource ("Databus").

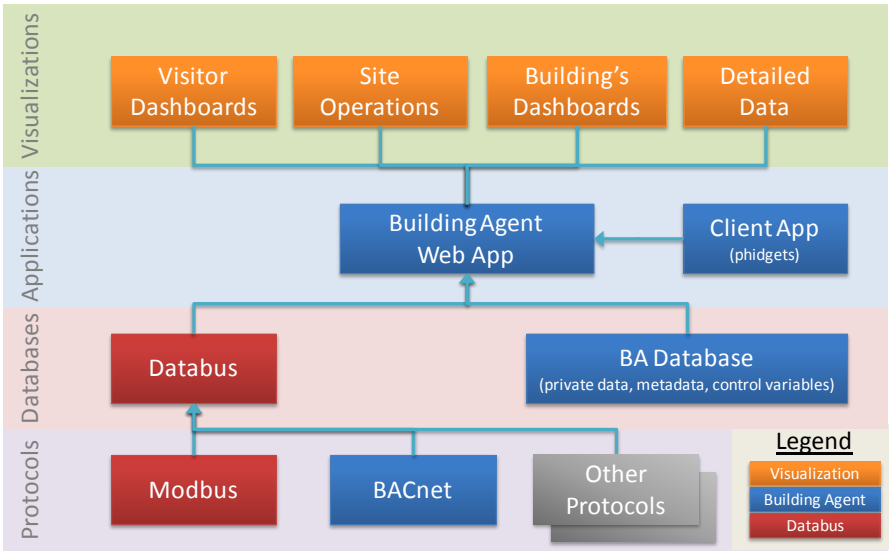

Figure 2–1 Building Agent architecture

The first application layer addresses the spatially distributed devices and varied protocols for monitoring energy- and comfort-related events on campus. Elements of this layer include (but are not limited to):

- Power quality meters such as lighting, mechanical, miscellaneous loads, elevators, and photovoltaic (PV) panels (*Modbus* protocol) on each building subsystem
- Flow meters on hot and chilled water, temperature, and carbon dioxide gauges in the building (*BACnet* protocol)

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

- Spatially distributed *environmental sensors* (called phidgets), such as illuminance, temperature, and humidity sensors at individual workstations
- *Other protocols* such as weather station sensors, including global horizontal and vertical irradiance, dry-bulb temperature, and wind speed and direction
- *Client applications* deployed to collect occupant feedback. Occupants are considered meters with many sensors in the Building Agent architecture.

All the devices either send data or are polled by a second database layer, at a time interval appropriate to the rate of change of the condition being measured. For example, environmental sensors send information either every 10 minutes or when occupants at workstations provide feedback about their comfort. The power meters are set to send data at 1-minute intervals. All these data are collected, aligned, cleansed, aggregated, and tagged in two databases for use by the Building Agent applications. One database, Databus, houses raw building performance data. These data are open and available for a variety of NREL research and assessment needs. A second database, specific to and named for Building Agent, houses metadata for the raw Databus data used for campus dashboards. The Building Agent database also segregates private occupant data such as comfort feedback.

The third and fourth layers of the Building Agent architecture go hand-in-hand. The relevant data are organized and manipulated and then displayed via a dashboard. The up-to-date status report from the building to the occupant allows for a human-in-the-loop control scenario. For the RSF, the interface takes the form of data visualizations through a desktop client application, a campus display, and a website. Each is meant to connect with occupants in different contexts, but all communicate the RSF's current performance and clearly indicate when the building is not meeting expectations.

The design team was contractually required to deliver real-time public visualizations for energy consumption by end use for the RSF (see Figure 2–2). It used the submetered data from the power quality and flow meters, but the applications layer was not included in the architecture. This means there was no point of comparison for the displayed energy performance; the data were simply shown as they were, leaving the viewer unsure of the building's health. Further research into historical information was necessary for energy performance assurance tasks such as verifying that the building's contractual energy goal was being met or determining the appropriate course of action if the building was not meeting the goal. The limitations of this original visualization prompted the development of an improved design.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

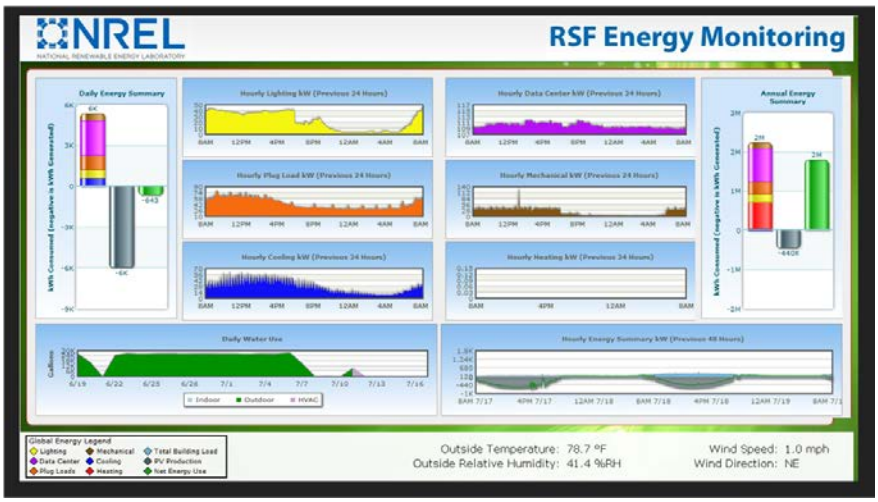

Figure 2–2 Original, control vendor-provided dashboard for the RSF (July 18, 2011, sunny day)

The second iteration of the RSF dashboard design added the data analysis layer. Over the first year of building operation, the metered data were collected and compared to the energy model results at the whole-building and end-use scales, as is done in an enhanced measurement and verification scope. For some end uses, monitored data were either not required or not needed; in these cases, the control limits relied solely on model-derived data. The second iteration of the dashboard was successful in that it enabled us to easily identify and implement improvements to building systems. This allowed the RSF's energy performance to meet the contractual energy goal. A new interactive dashboard (see Figure 2–3) was developed that displayed the building's power profile as a breakdown of the whole-building energy goals.

Figure 2–3 Dashboard for the RSF

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

The interface enables users to "drill down" to inspect individual end-use time series data (see Figure 2–4). Here, the mechanical loads are outside the expected range, especially during unoccupied periods. The cause of the deviation was an air handler that was manually overridden to run all night. After this inappropriate operating schedule was corrected, the agreement between measured and modeled data improved greatly.

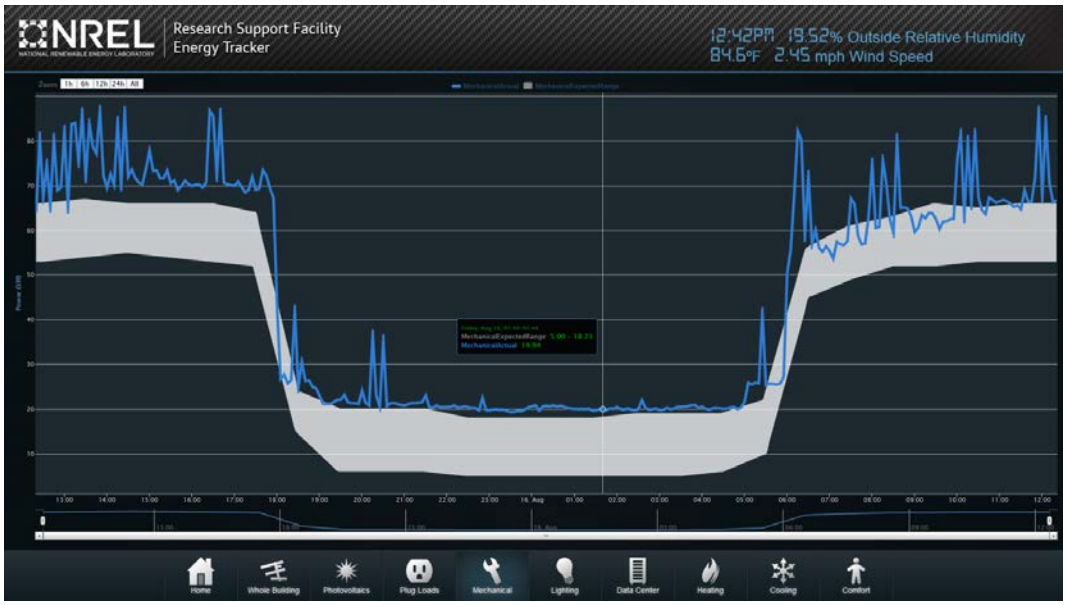

Figure 2–4 Building Agent dashboard for the RSF mechanical loads, August 16, 2013

These examples use a real-time snapshot and a historical analysis. Both are necessary depending on the metric for comparison; multiple dashboard views provide the diversity of information occupants and building engineers need to make informed decisions about comfort and energy use. Although the fourth layer (building-occupant interface, display selection, and design) is an important discussion point for meaningful and actionable visualizations (Schott et al. 2012), the remainder of this report addresses the third, applications layer.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

3.0 Methodology

The report describes three different but related approaches to calculating performance bounds, called *control limits*, on building energy performance, divided by energy end use.

1. Univariate Control Limits Based on Engineering Judgment
2. Univariate Control Limits Based on Quantile Analysis
3. Multivariate Control Limits Based on Quantile Regression.

The goal of developing these control limits is to provide intuitive, visual contexts for displayed performance data without requiring users to conduct an in-depth building energy data analysis.

In the first approach, which uses engineering judgment and manual curve fitting, end-use control limits are derived using static lookup-based performance targets, combined with polynomial curve-fit models. These control limits are developed based on a combination of historical end-use data and as-built energy modeling. Single main effect variables are selected for each end use, such as solar irradiance for lighting performance.

In the second approach, measured building end use energy data are used to develop empirical frequency distributions for each end use as a function of the main effect variable determined from visual inspection and a pairwise correlation analysis. Appropriate control limits are defined as ranges in which a certain percentage of end-use consumption has been observed historically; e.g., the central 50% of observations between the 25^{th} and 75^{th} percentiles, as a function of the main effect variable. The same process is then applied to simulated (modeled) energy end use data sourced from a calibrated building energy model.

In the third approach, the building energy end use is modeled linearly in terms of other end uses, exogenous variables, and weather data, all of which are then transformed to orthogonal principal components (PCs), and the linear models are used to generate control limits using quantile regression for the measured and simulated data. Here, quantile regression is used to predict the more stringent 40^{th} and 60^{th} quantiles. Different percentile ranges are chosen between the second and third approaches to show different perspectives on control limit stringency.

For the first approach, control limits are developed for all end uses, including heating, cooling, lighting, mechanical, plug loads, data center, and PV system output. For the sake of brevity, in case of the second and third approaches, electricity consumption for cooling provided by chilled water is considered the energy end use to illustrate the process.

In all three approaches, two sets of upper control limits (UCLs) and lower control limits (LCLs) are generated from historical measured and simulation-based data. An energy-conservative logic is applied in all approaches to select one set of control limits in each situation. This set of LCLs and UCLs for each energy end use is then adopted in the Building Agent for presentation to building operators and occupants.

The energy-conservative logic, described in greater detail below, selects the lower value of the two UCLs and the lower value of the two LCLs for display on the Building Agent as long as the model-sourced control limits can be considered valid. Conversely, if the modeled control limits are deemed invalid, the control limits derived from the measured data are adopted. In effect, this logic presents to the viewer the more conservative control limits, encouraging occupants and operators to maintain the building's energy performance in a more energy-conservative state. In

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

the examples presented in this report, the modeled control limits are at times much lower than the measured data. If the modeled control limits significantly deviate from the measured data results, users may not be motivated to take action because the targets are not realistically attainable. Although the model is meant to represent end-use targets, large differences may mean either that the model is aggressive but attainable or that it is unrealistically aggressive. In the latter case, indeed, occupants will be demotivated. Therefore, a control limit development process that balances aggressive, yet attainable targets is a key motivation for our efforts.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

4.0 Univariate Control Limits Based on Engineering Judgment

4.1 Predicted Range Approaches

Unique to the RSF dashboard (Figure 2–3) are the acceptable ranges around the subsystem goals that dynamically update based on measured main effect variables. This concept of providing acceptable ranges, or control limits, is critical for creating a useful dashboard that compares real-time end-use performance with model expectations. In the first approach, which uses engineering judgment and manual curve fitting, end-use control limits are expressed by static performance tables and polynomial curve-fit models. These control limits are developed based on a combination of historical end-use data and as-built energy modeling. Single main effect variables, such as solar irradiance for lighting performance, were selected for each end use.

4.1.1 Development of Curve Fits for Control Limits

The following process was executed to develop the univariate (i.e., single main effect) control limits for each end use as displayed on the Building Agent. To begin the process a minimum of 1 year of hourly end use predictions from the as-built energy model are needed, along with corresponding hourly weather data. For each end use, engineering judgment and past experience were used to determine the main effect or dominant driving variable. Common examples included occupancy type, time of day, outdoor dry-bulb temperature, and solar radiation on a horizontal or an inclined surface. Internal gains from people and equipment differ greatly between unoccupied and occupied states, so establishing separate curve fits for each occupancy state helped build more accurate control limits. The researchers then used the dominant driving variable and hourly end-use data to develop curve fits that approximated the upper and lower bounds to identify the "acceptable" hourly end-use range versus the dominant driving variable. They incorporated the general philosophy for developing upper and lower curve fits in an attempt to exclude visually identified outliers. The curve fits by end use were then translated either to lookup tables or to polynomial curve fit equations; the final goal was to measure each dominant driving variable and determine the acceptable ranges for any point in time.

The fidelity of this method improves as end-use performance data are collected and integrated into the control limit calculations and the engineering analysis. As measured performance data from the first year of RSF operations were evaluated, the purely model-based control limits were compared to the measured data. For hours when the measured end uses exceeded the expectations, commissioning actions were identified and improvements implemented in an attempt to reduce end uses to better align with expectations. For modes of operation where measured end uses were significantly below the expectations, the expectation curve fits were manually updated to align with past measured performance.

This first approach to developing control limits successfully aligned operations with expectations; however, significant expert analysis was needed to develop the curve fits and identify the end-use dominant driving variables. A key characteristic of this approach was that dynamic control limit ranges are available both in real time and over historical trends. This insight was necessary for building operators and energy use decision-makers to understand how to align operations with expectations, especially when unrealistic ranges were updated based on improved operations. However, a disadvantage was that an experienced energy engineer had to analyze an initial as-built energy model to identify the dominant driving variables. The manual curve fit development process was based on engineering judgment, which limited the replication

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

potential. Subsequent development efforts for the control limits focused on automating both the acceptable range development process and the identification of the dominant driving variables.

4.2 Application

The reduced-order model approach to end use energy range predictions initially implemented in the Building Agent is discussed in more detail in this section, showing the development process, minimum data requirements, and positive outcomes of using a predictive interface in energy performance assurance in the RSF.

4.2.1 Photovoltaic Panels

More than 2 MW of PV panels are installed on or near the RSF to offset the office building's annual energy use to achieve net-zero energy. Although renewable energy was a last consideration in the design process (after system efficiency measures), it is presented first because of the relative ease in determining a dashboard representation with acceptable control limits.

A PVWatts PV model was used to predict energy production based on the final panel specification and configuration. Figure 4–1 shows that the continuous curve fit model was determined by first plotting the modeled production against global horizontal irradiance, a known dominant variable for PV production. Minimum and maximum control limit boundaries were then added from visual inspection and fitted with curve-fit equations that were then implemented in the application layer of the Building Agent.

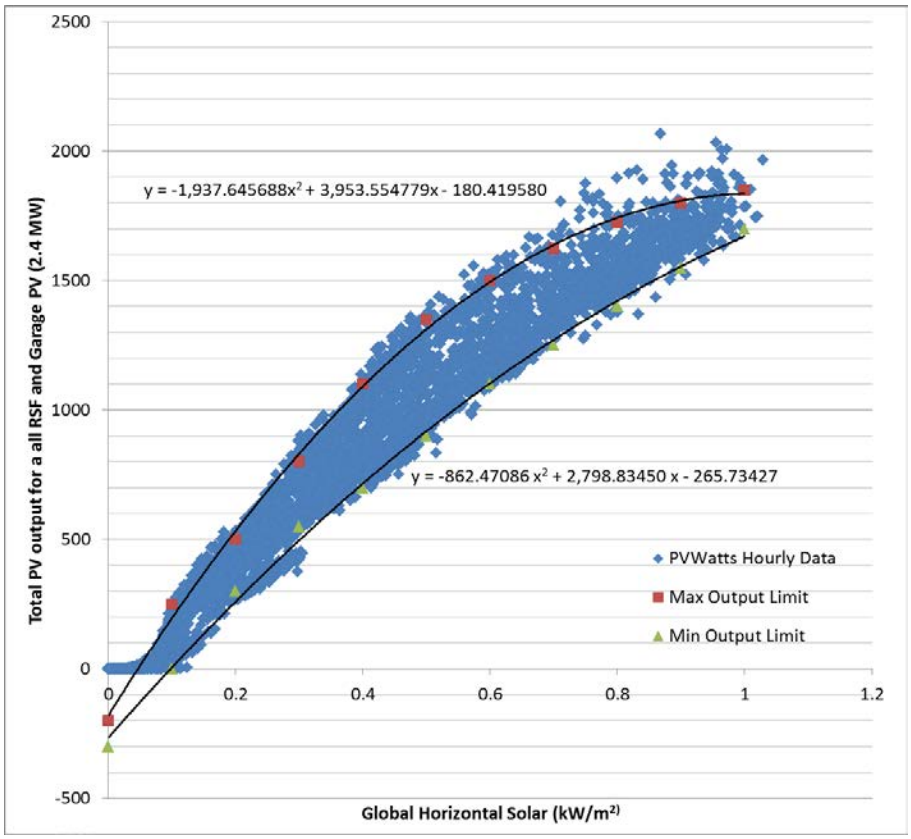

Figure 4–1 PV production versus global horizontal irradiance

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

The visualization layer of the PV predicted range is shown in Figure 4–2, with the measured PV output compared to the expected control limits, as calculated based on the measured global solar insolation (the dominant variable for PV performance). The measured PV output exceeded the expected range for many hours, suggesting the temperature during those hours was lower and thus the PV output higher than for the average temperature used to establish the control limits. This observation reveals the limitation associated with selecting only one dominant variable and suggests more advanced approaches that account for all explanatory variables. The quantile regression approach presented below is one possible pathway to accounting for all explanatory variables, as would be a Monte Carlo-based execution of the PVWatts simulation model using actual data to the extent available and probability distributions for the unobserved input variables.

Figure 4–2 PV production predicted range visualization, blue line (measured) and silver band (control limits) for September 25, 2013

4.2.2 Lighting

The RSF occupants have some manual control of the electric lighting. Daylight and occupancy sensors and timed sweeps automatically control the lights to dim and turn off. Daylighting in a net-zero energy office building is critical, so a detailed daylighting model was produced in design. Figure 4–3 shows the results of the daylighting model, as expressed in hourly lighting power density (LPD) (w/ft^2) as plotted against the available solar daylighting resource. Based on visual inspection of general dependent variable impacts on daylighting performance, the initially developed UCLs and LCLs were generated and are also shown.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

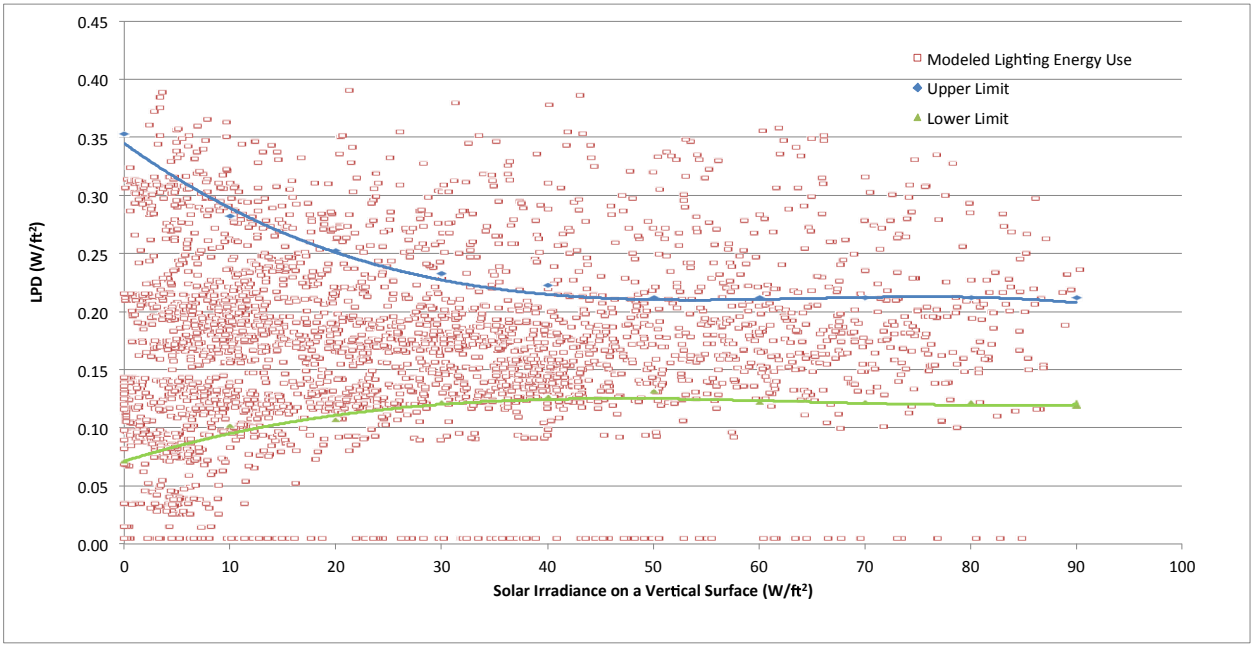

Figure 4–3 Modeled LPD versus solar irradiance on a vertical south-facing surface, with UCL and LCL curve fits

Figure 4–4 shows that the daytime lighting energy use is largely dependent on exterior south vertical irradiance in actual operations, with the UCLs and LCLs for lighting energy use during occupied hours. UCLs and LCLS based on the measured data also provided a better visual fit than the modeled control limits suggested. Therefore, we used the measured lighting control limits for implementation in the Building Agent. The data scatter in the modeled and measured lighting energy uses at low to no solar availability demonstrate the occupant control impacts on lighting energy use when daylighting is not available.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

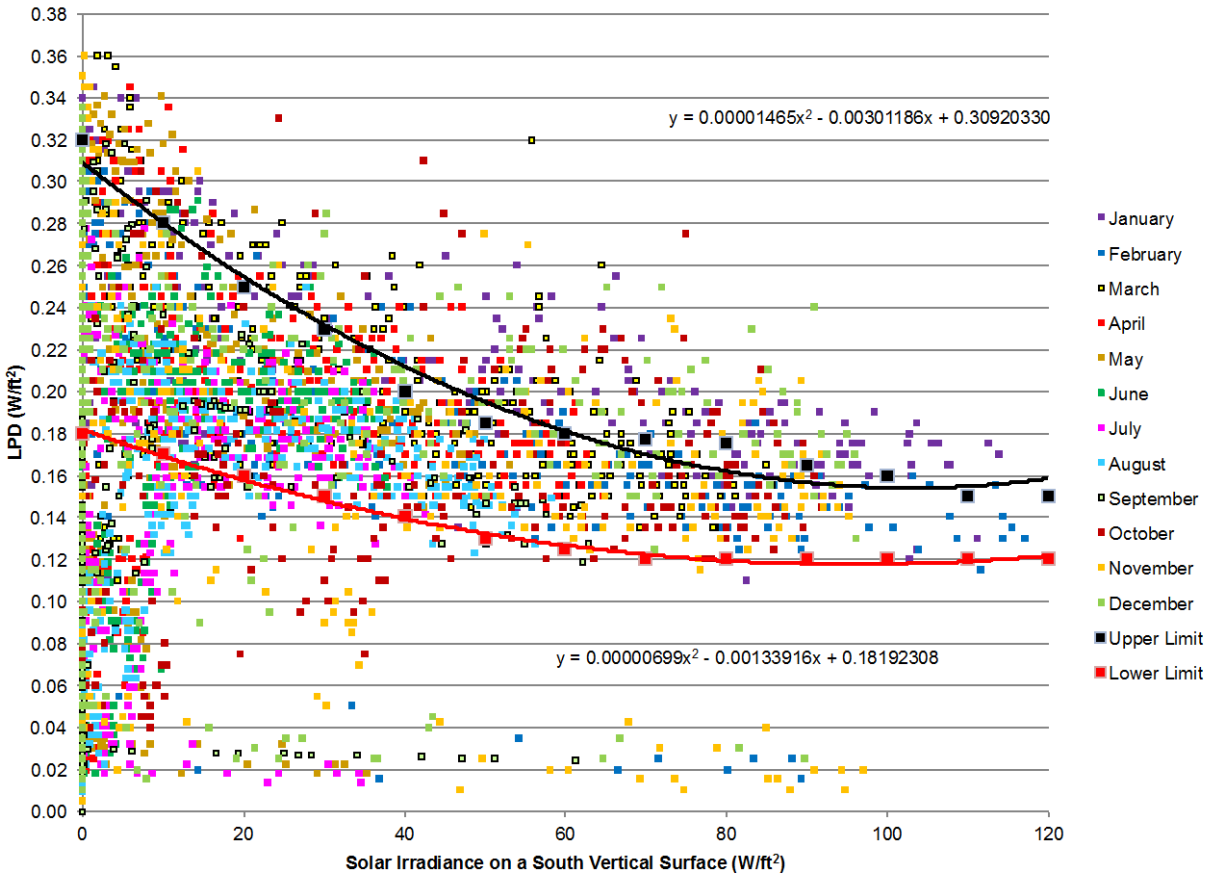

Figure 4–4 Weekday daytime LPD versus solar south vertical insolation

Figure 4–5 shows the importance of an operational model. Unlike the daytime lighting energy use, the nighttime lighting energy use is much higher than the design model predicted because the custodial crew's system use was not considered during design. The annual average lighting energy use during occupied hours matches the model well, and was used to develop control limits. The energy model expectations for unoccupied hours were also used to deploy the Building Agent dashboard to continue to show the opportunity for turning off lights after hours.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

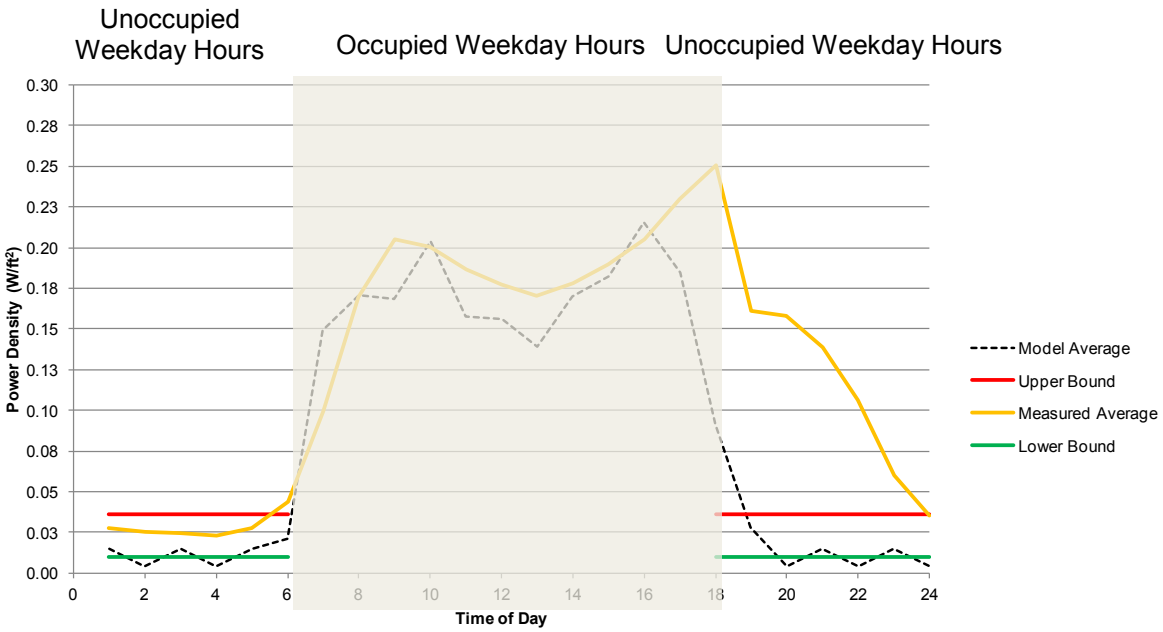

Figure 4–5 Weekday unoccupied LPD control limits

The lighting predicted ranges as initially applied in Building Agent show a continuous curve fit control limit combined with the unoccupied hours lookup table (see the dashboard implementation in Figure 4–6).

Figure 4–6 Lighting predicted range visualization with measured use (blue line) and control limits (silver band) for October 10, 2013

4.2.3 Plug Loads

Plug loads are one of the largest end uses in the RSF, and are subject to aggressive load reductions (Lobato et al. 2011); each occupant has a 55-W allowance for a laptop, monitors, a phone, a task light, and miscellaneous items. The energy model accounted for this load but could only assume a diversity factor of occupant use throughout the day.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

The first step in operational energy performance assurance is to compare the first-year measured data, or initial operational model, to the design modeled data. Figure 4–7 shows that the hourly average daytime operational load is much lower than the hourly average design prediction, but the nighttime load is higher than expected. The model average is the same as the upper bound during unoccupied hours. The model suggests the end-use budget; thus, the lesser of both datasets at any given hour is used to determine the dashboard control limits.

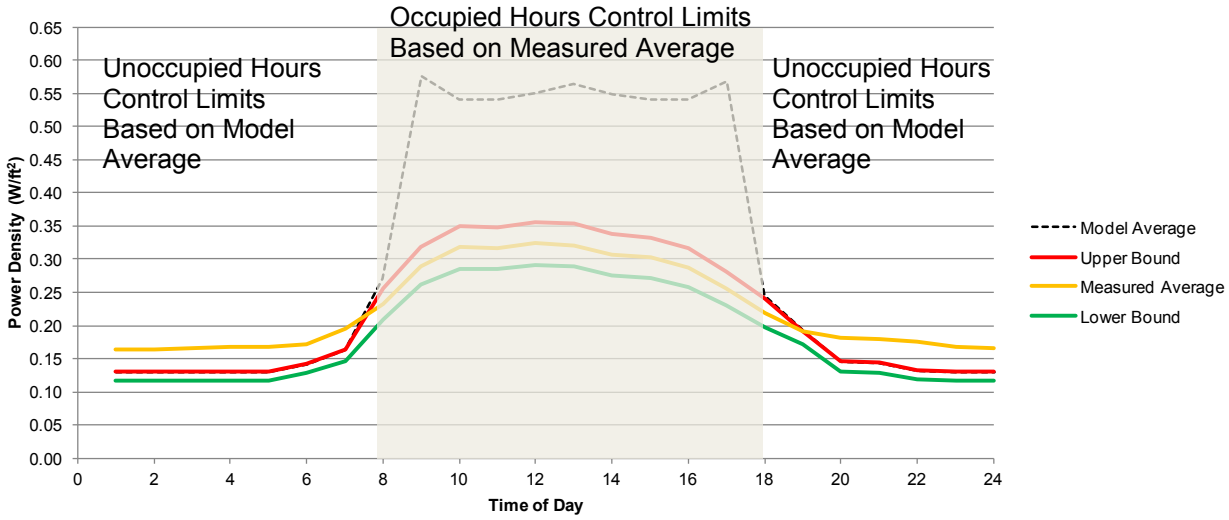

Figure 4–7 First-year plug load operation hourly average power profile

The UCLs and LCLs for occupied hours were determined by visually bounding the first-year data range (excluding outliers) and implemented in the dashboard using a lookup table, because time of day is the most obvious factor in the primarily occupant-driven load. The resulting visualization is shown in Figure 4–8. The data center energy use acceptable control limits were developed in the same manner.

Figure 4–8 Plug loads predicted range visualization with measured use (blue line) and control limits (silver band) for September 24, 2013 14:00 to September 25, 2013 14:00

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

4.2.4 Mechanical

The RSF's mechanical loads consist primarily of ventilation air fans and pumps for the radiant heating and cooling systems.

The energy performance assurance and predicted range evaluation for the mechanical systems mirror the process for developing plug load control limits, with two exceptions: (1) the design model predictions overestimate energy use during the day instead of the night, so the first-year metered data are used as an upper bound at night and the design model results are used as an upper bound during the day; and (2) like the plug loads, a lookup table based on superior measured performance during unoccupied hours is used in place of the modeled expectation. This process is demonstrated in Figure 4–9.

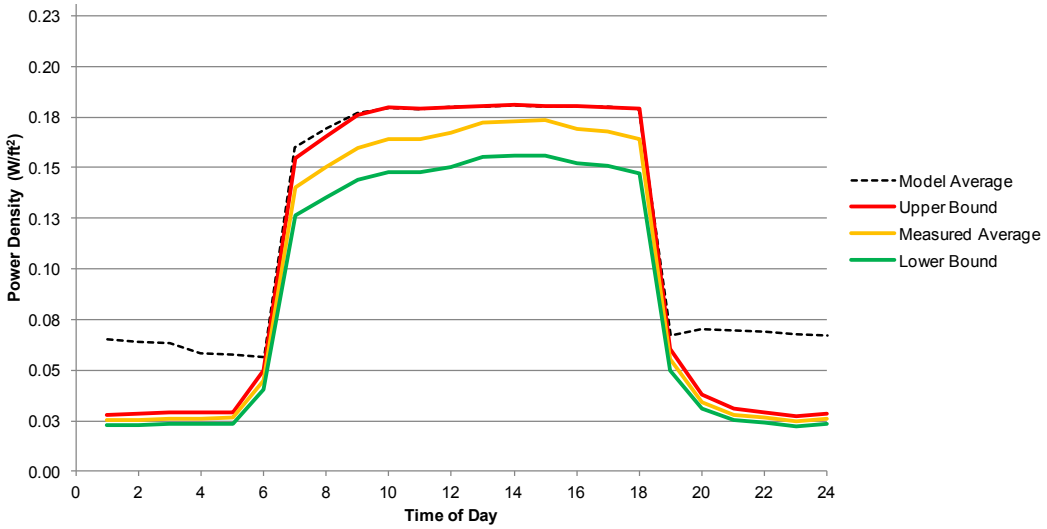

Figure 4–9 First-year mechanical systems operation hourly average power profile

Figure 4–10 shows the implementation of the control limits for the mechanical end uses in the Building Agent. The measured data for this implementation example exceed the daytime and nighttime control limits, suggesting a possible fault. Investigation into the mechanical systems revealed a manually overridden air handler schedule, keeping the system 100% on during unoccupied and partially occupied hours. Once this was discovered and corrected, the actual use coincided more closely with the control limits (see Figure 4–11).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

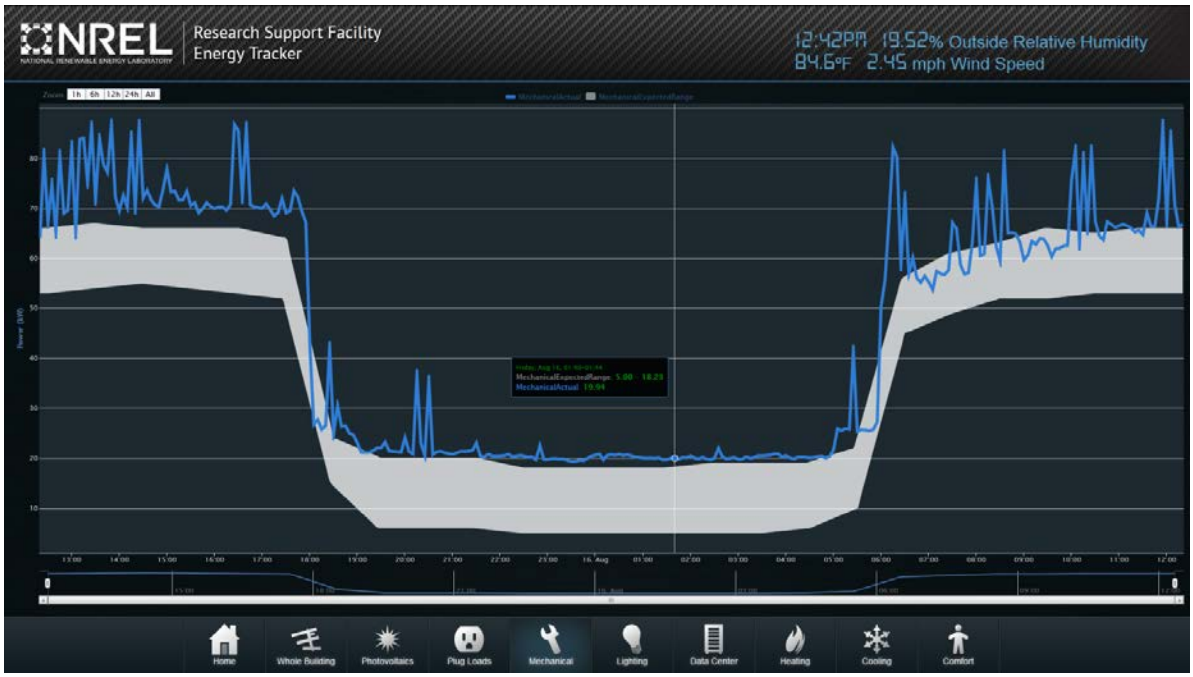

Figure 4–10 Mechanical systems predicted range visualization with measured use (blue line) including faulty air handler schedule and control limits (silver band) on August 15 and 16, 2013

Figure 4–11 Mechanical systems predicted range visualization with measured use (blue line) including corrected air handler schedule and control limits (silver band) on October 7 and 8, 2013

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

4.2.5 Cooling and Heating

The RSF uses hot water and chilled water from the campus district energy plant. Hot water and chilled water meters measure the RSF's thermal energy use and chilled water efficiency (kW/ton), and natural gas boiler efficiencies are applied to the metered thermal energy use to calculate system energy end uses. For heating and cooling end uses, the outdoor air dry-bulb temperature is the dominant environmental variable. By comparing the modeled hourly cooling energy use to the outdoor dry-bulb temperature, LCLs and UCLs were developed for occupied and unoccupied hours through visual inspection. Linear curve fits based on the modeled cooling performance were then provided to the Building Agent (Figure 4–12, and cooling acceptance ranges were calculated and compared to measured cooling energy use. This overly simplistic approach is limited, as it fails to capture important features in the data. The heating control limits were derived with similar methods based on outdoor dry-bulb temperature as the single main effect variable.

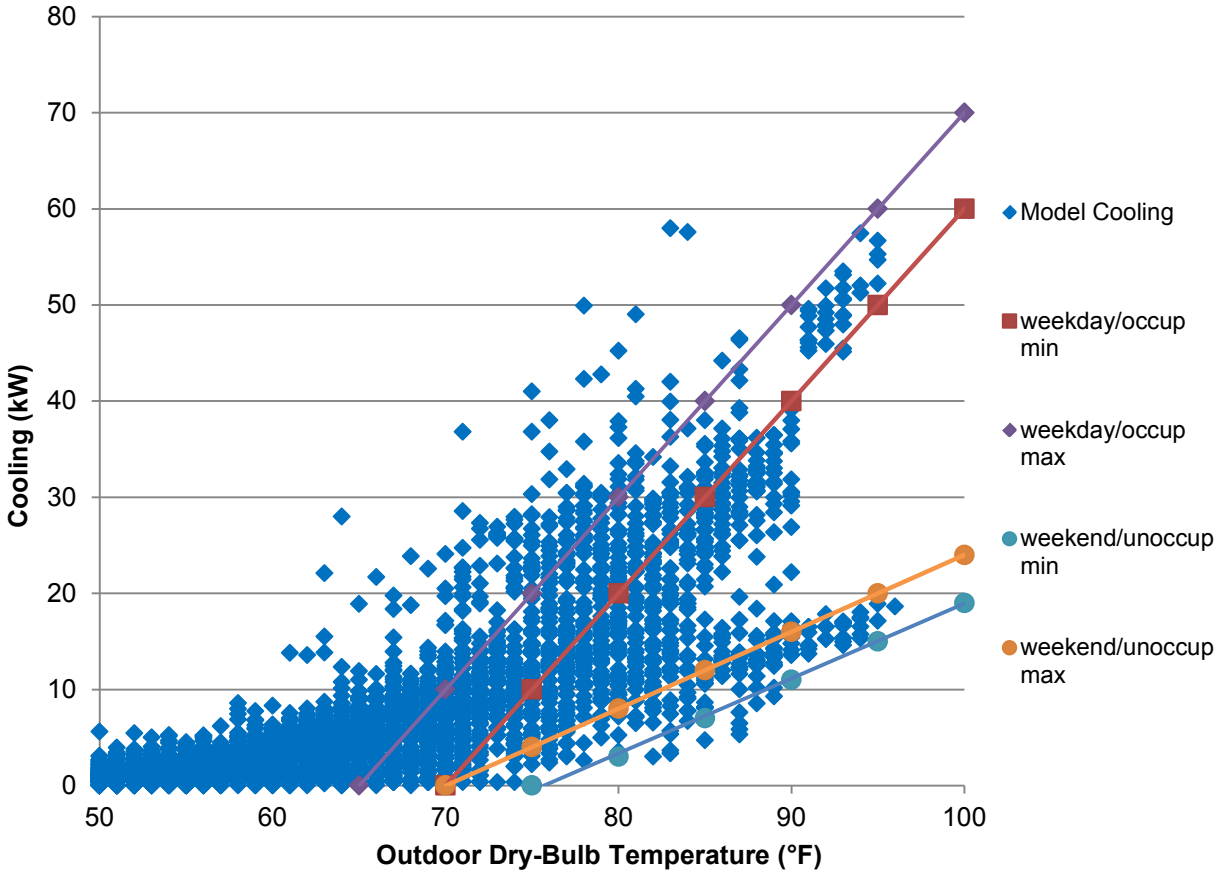

Figure 4–12 Modeled hourly cooling with weekday and weekend control limits

These univariate (depending on the dominant explanatory variable) curve fits for acceptable control limits were based on the as-built model prediction for cooling energy use, and successfully predicted actual cooling energy use (Figure 4–13. Setting end-use budgets, and tracking actual energy use in real time against these budgets, are effective methods for realizing net-zero design expectations for the RSF.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

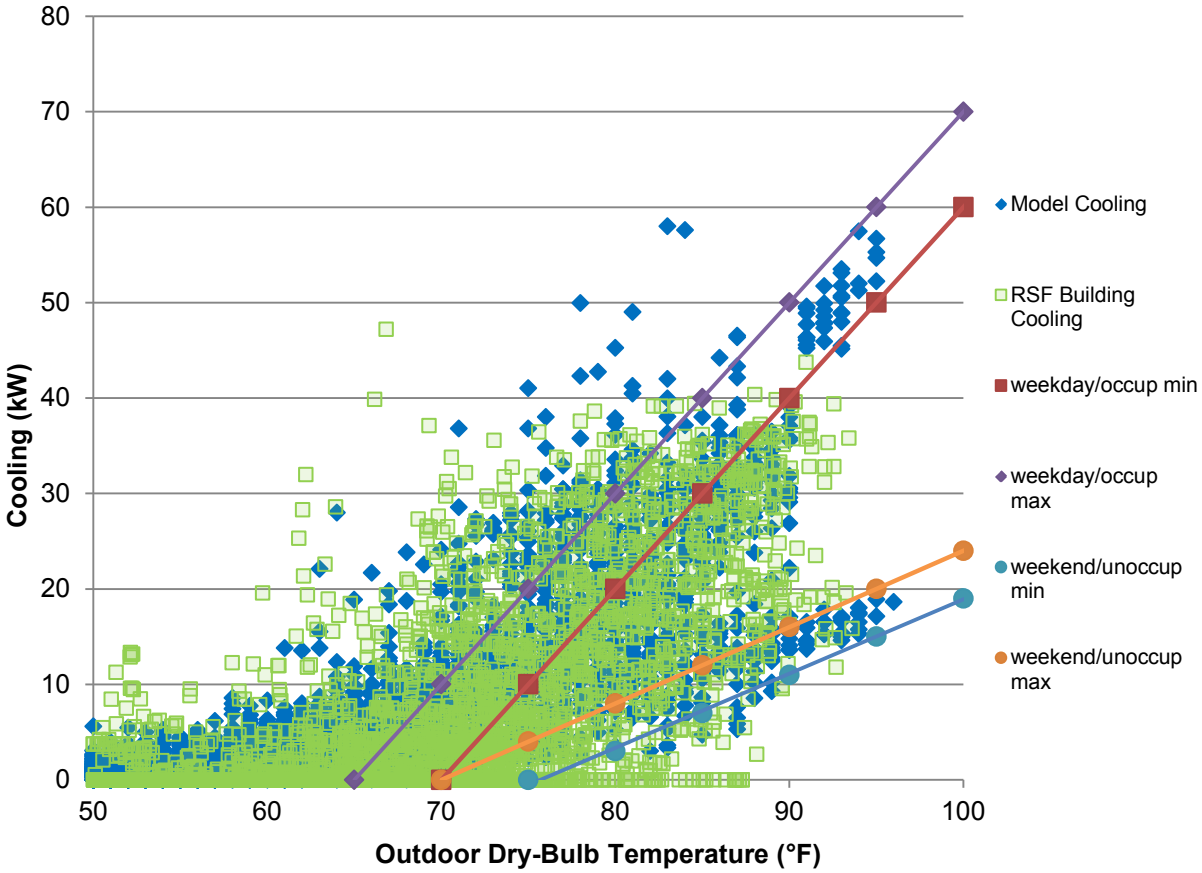

Figure 4–13 Measured and modeled hourly cooling

This manual curve fit process effectively provided first-order control limits; however, its limitations are evident. Cooling energy use is not always best predicted based solely on occupancy state and outdoor dry-bulb temperature; outdoor humidity and solar gains can also contribute significantly to cooling energy use. These limitations suggested the need to develop better methods for automating these control limit curve fits and including multiple dominant variables. Subsequent sections of this report discuss these improvements.

4.3 Discussion

The first process used to determine the predicted range models for the RSF, the first building represented by the Building Agent, is presented as an example of the energy assurance process emerging in the commercial buildings industry. The process consisted of design and metered data comparisons, the validation or development of an asset model for systems, and an operational model for control variables, where necessary. Then, a visually determined continuous polynomial curve fit equation or lookup table was applied to the asset and operational models and data, respectively.

For noncontrolled systems such as PV, visually bounding asset model predictions relative to known dominant variables to determine reduced-order equations is a simple and sufficient process. The information occupants need about these types of systems is whether the system is working and whether degradation is occurring.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

On the other end of the spectrum, complex controlled building systems such as fans and pumps need more rigorous approaches to determining predicted energy use ranges. The reduced-order lookup table approach used for the RSF Building Agent is not sufficient to capture the hourly and daily differences between potential and actual performance. This leads to missed opportunities for energy savings.

The first attempt to compare real-time energy end uses to model expectations using a simple method that captures first-order effects was simplistic and required engineering insights, but the Building Agent visualization still provided value to the RSF in meeting energy goals. The value in predicted range visualization has been realized at NREL through numerous anecdotes to date:

- The high plug load use at night was consistently high relative to the predicted ranges. This led NREL energy engineers, information technologists, and managers to work together to communicate options to employees for turning off monitors, computer, and other miscellaneous loads at night. Best practices were developed based on the lessons learned (Lobato et al. 2011).
- A high nighttime lighting load has led to training sessions that remind custodial crews to turn off ambient lights when they leave an area and to use the egress switches when possible, as these often provide sufficient light for the tasks being performed.
- A consistently high data center load relative to the model gave the data center manager evidence to request funding for improved hot aisle containment strategies.
- Last but not least, the whole-building energy display comparing the RSF energy use to the contractual EUI and net-zero energy goals have led to continued emphasis on meeting energy goals during building upgrade projects. For example, the addition of a 24-hour visualization room for security was not originally designed for dedicated heating, cooling, or ventilation. Dedicated systems were added once the building energy engineer was able to show that the building would tip past its annual operating goal if the building systems were to be used for this 24-hour occupancy program.

The visualizations have provided information and the ranges have added the needed justification for building energy projects.

The RSF dashboard development case study reveals the importance of defining a complete architecture, such as Building Agent, for a building-occupant interface. The architecture must be considered in early project planning to ensure hardware (such as submeters) is in the purview of the design and construction team. Also, owners must account for the human and computing resources necessary to create as-built asset models and ongoing operational models. A final construction, or asset model, is sufficient for representing some end uses, but an operational model is needed for the high energy-saving potential systems that are controlled either by occupants or automatically, or both.

The Building Agent dashboard moved the RSF operational practice from presenting noncontextualized building performance to a first-order data analysis with boundary conditions (i.e., presenting end uses with predicted ranges of acceptable energy use).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

5.0 Univariate Control Limits based on Quantile Analysis

5.1 Overview

This second method improves on the previous method by developing an empirical quantile analysis process for calculating the control limits, rather than the manual curve-fit described in the previous chapter. For the sake of brevity, electricity consumption for cooling provided by chilled water is considered the energy end use of choice to illustrate the second and third methods. Here, chilled water load is converted to an electricity equivalent using a constant coefficient of performance (COP) of 7.8.

In this section, measured RSF I performance data for 2011 (consisting of wings B and C) are analyzed and compared against the RSF I calibrated building energy model. Wing A, also known as RSF II, was opened in 2012. Because no separate energy model has been developed for RSF II, modeled energy data for the combined RSF I and II (simply RSF) scale the energy use data of RSF I on the basis of building floor area. The third method (based on quantile regression) and discussed in Chapter 6, will investigate 2012 data for RSF I and RSF II.

Initially, annual average daily quantile plots (box plots) are generated for the cooling energy end use, followed by a distinction between weekday and weekend. Next, a winter/summer seasonal distinction is defined and the end-use consumption portrayed for winter and summer periods. At its lowest level of resolution, each end use is distinguished by season and day of the week, revealing differences between weekdays and weekend days and holidays, but also between weekdays. Next, scatter matrix plots are presented to reveal pairwise correlations among measured variables and the most important independent variables for each end use are identified. The chapter continues with annual and seasonal carpet plots for the cooling end use, allowing visual detection of seasonal patterns and unique events, and identifies the dominant independent variables.

Finally, the chapter concludes with the development of the second control limit method based on the frequency of occurrences and the implementation in the Building Agent energy dashboard.

5.2 Measured Performance Data

5.2.1 Measured Data Annual Box Plots

5.2.1.1 Average Daily Behavior

The RSF receives chilled water from a central chilled water plant serving a district cooling system on the NREL campus. The RSF I (wings B and C) cooling energy consumption profile shows peak annual values of around 40 kW electricity demand based on a simplified model, assuming a constant COP of 7.8. This high COP value is attained—and has been verified through measurement—as a result of a best-in-class, fully variable, water-cooled chilled water plant operated in a very dry climate with very low wet-bulb temperatures. Considering that RSF I has a conditioned floor area of 220,000 ft^2, the resultant cooling demand of less than 0.2 W/ft^2 is extremely low. Clearly, the RSF is a high-performance building, and the mixed-mode design—which involves natural ventilation, ground-coupled ventilation air precooling, good passive shading, high mass and high thermal resistance envelope, very low internal gains, and radiant cooling—is very effective.

The box plot profiles in Figure 5–1 show the median values (thick black horizontal lines) biased toward the first quartiles, confirming that during most hours of the year the cooling demand is

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

low; seasonal summer cooling perspective is thus offered below. The convention for declaring outliers is to use values that are either greater than 1.5 times the interquartile range above the third quartile or 1.5 times the interquartile range below the first quartile.

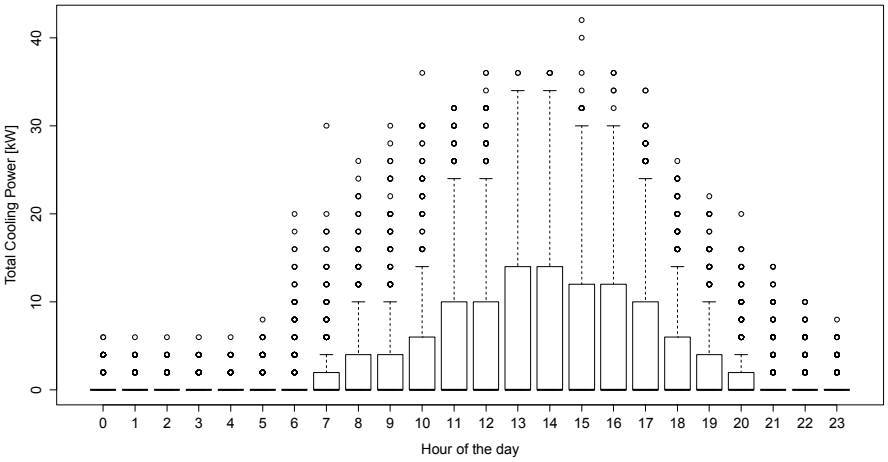

Figure 5–1 Measured annual average daily cooling end use power as a function of time of day

5.2.1.2 Separate Weekday and Weekend Behavior

In the next step, the data were disaggregated into weekday and weekend patterns to eliminate the variance in the box plots (see Figure 5–2) based simply on the difference in building use between weekdays and weekends. Cooling-related electricity demand for any hour of the day is a long-tailed distribution, naturally bounded by zero at the low end. Much variance in the data remains unexplained with an annual average time-of-day perspective, and motivates a seasonal distinction.

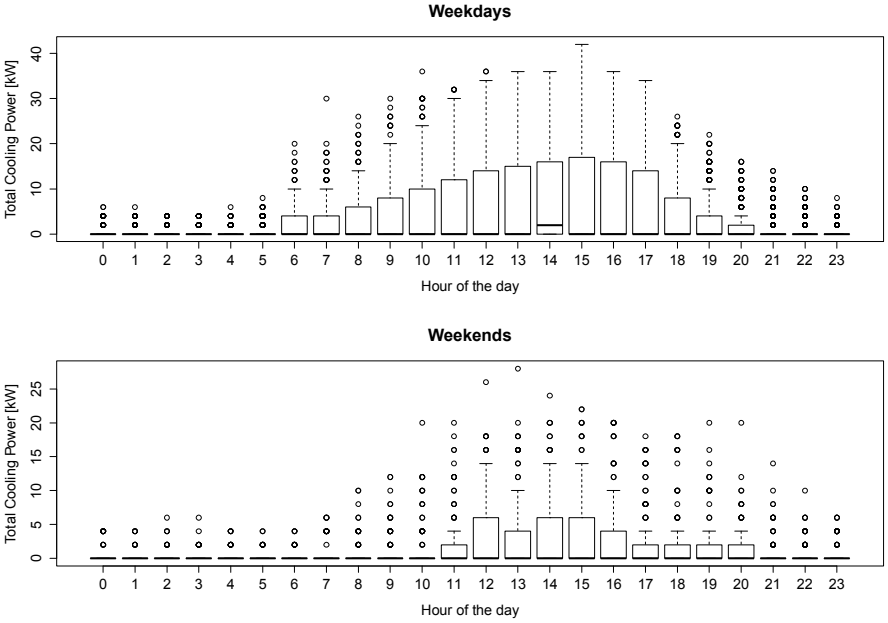

Figure 5–2 Measured annual average weekday and weekend cooling energy end use as a function of time of day

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

5.2.2 Measured Data Seasonal Box Plots

5.2.2.1 Average Daily Behavior

This section shows average daily profiles separated by winter and summer periods. Summer is considered June through October (5 months); winter is considered January through May and November through December (7 months). The summer box plot in Figure 5–3 shows a distinct profile with median values rising from 06:00 to 13:00 and peaking during the early afternoon period of 14:00 to 16:00, thereafter falling again, with significant cooling remaining after occupancy from 18:00 to 21:00.

The winter profile shows no discernible patterns, with most consumption being very low except for a large number of outliers as high as 20 kW; these outliers are expected to be introduced by the artificially abrupt summer/winter distinction. Winter observations will not be further studied.

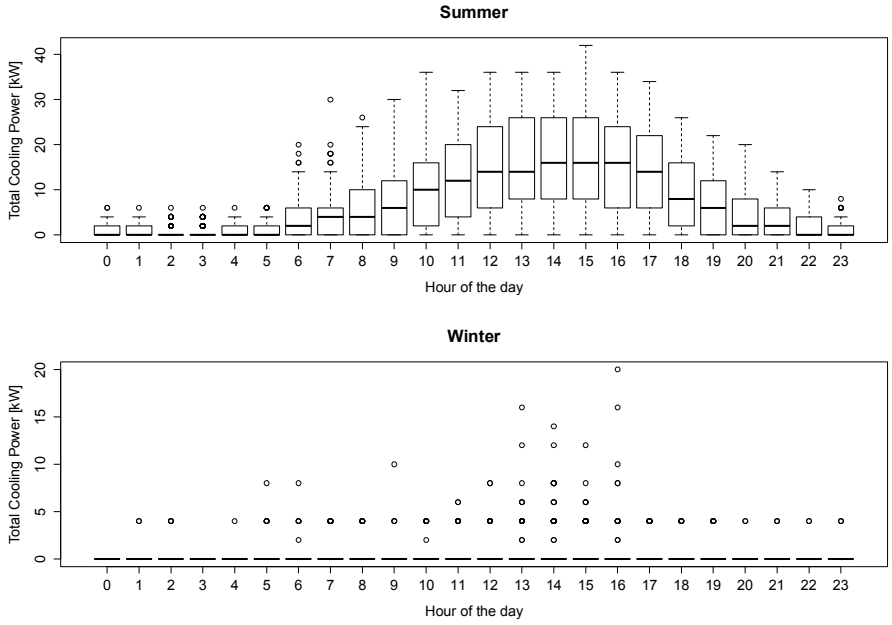

Figure 5–3 Measured annual average summer and winter cooling energy end use as a function of time of day

5.2.2.2 Daily Behavior by Weekday During the Summer

Because the average summer day has a distinct diurnal profile, each day of the week is presented during the summer. The weekend days (Saturday and Sunday) exhibit much lower cooling energy consumption than do the weekdays, suggesting either day-typing into weekday and weekend days, or alternatively, distinguishing between occupied and unoccupied periods (see Figure 5–4). Mondays reveal the highest median cooling demand, followed by Tuesdays. Wednesdays and Thursdays are next lower; Fridays have the lowest weekday cooling demand. The high Monday values are likely due to the extra load resulting from cooling a building that was experiencing a weekend temperature setup.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

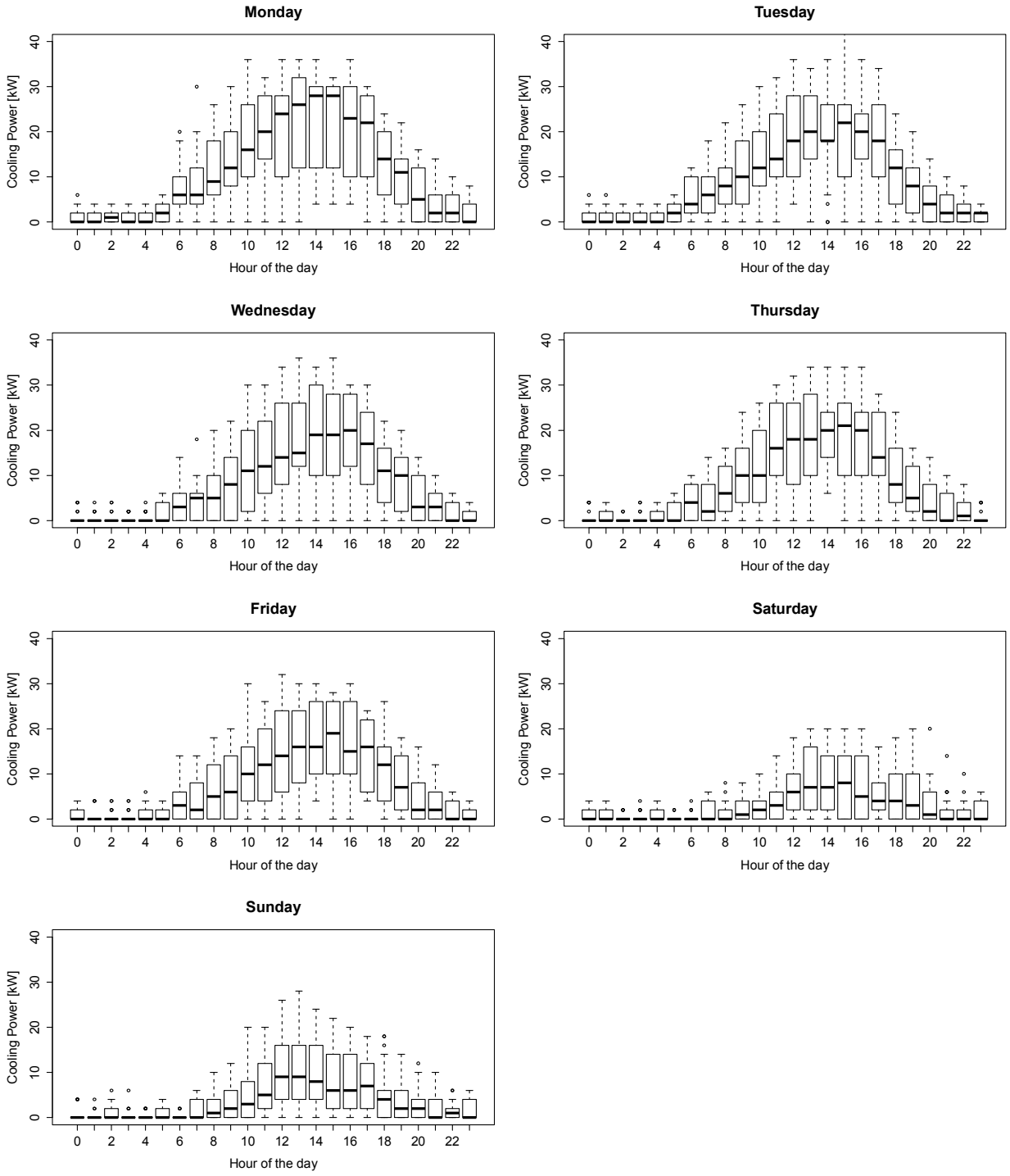

Figure 5–4 Measured average summer cooling energy end use by day of week as a function of time of day

5.2.3 Measured Data Scatter Matrix Plots

The analysis thus far investigated temporal dependencies along season, day type, and hour of day; this section analyzes dependencies of variables relative to each other. After the data have been split into separate variables that contain either summer or winter data, scatter matrix plots

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

are created for summer and winter periods separately to find pairwise correlations. The pairwise Pearson's correlation coefficients are in the upper panels and the scatter plots are in the lower panels. In statistics, the Pearson product-moment correlation coefficient is a measure of the linear correlation between two variables, giving a value between +100% and −100%, inclusive. It is widely used as a measure of the strength of linear dependence between two variables.

During the summer, a relatively strong correlation of 71% between dry-bulb ambient temperature (Temp.Sum.Mod) and measured cooling power can be measured, and a weaker one (51%) between global horizontal solar and cooling power (see Figure 5–5). Ambient temperature and global horizontal insolation were the only available exogenous drivers considered for cooling. Hence, if only one main effect variable can be chosen to model cooling energy demand during the summer, it should be ambient dry-bulb temperature.

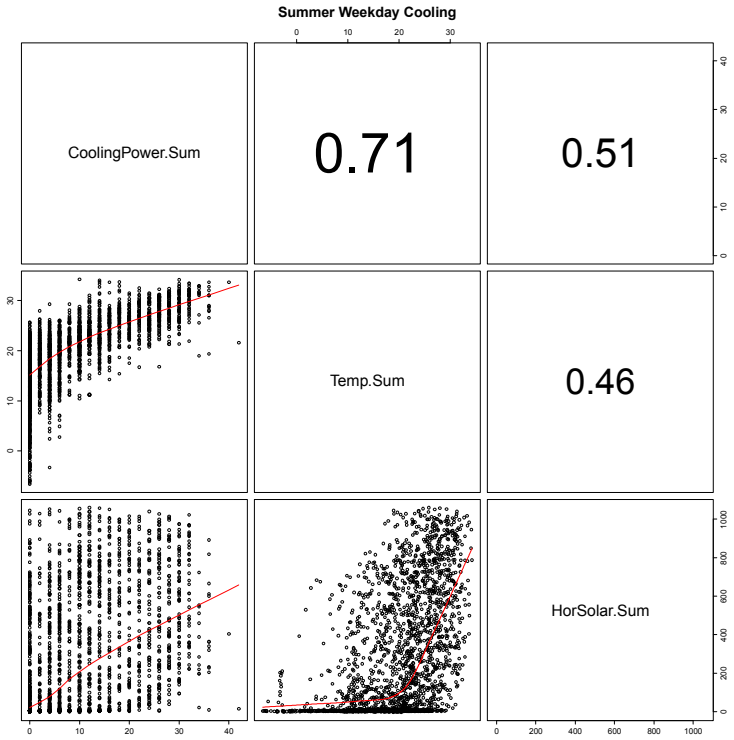

Figure 5–5 Pairwise correlation for measured summer cooling energy end use for ambient dry-bulb temperature and global horizontal insolation

5.2.4 Carpet Plots

In this section, carpet plots (also known as heat maps) are prepared for various time frames, allowing one to identify patterns easily and simply with visual inspection.

The annual carpet plot in Figure 5–6 reveals the peak cooling demand to fall into the time period of late June to late August and from 11:00 to 18:00. Compared to the measured building performance, the cooling demand drops sharply at the end of occupied hours (18:00). Vertical lines show weekends and holidays. The summer carpet plot in Figure 5–7 shows high cooling demand focused on several hot weeks during late July and late August. Moreover, 2 hot weeks with high cooling energy demand occurred during late September and early October 2011.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

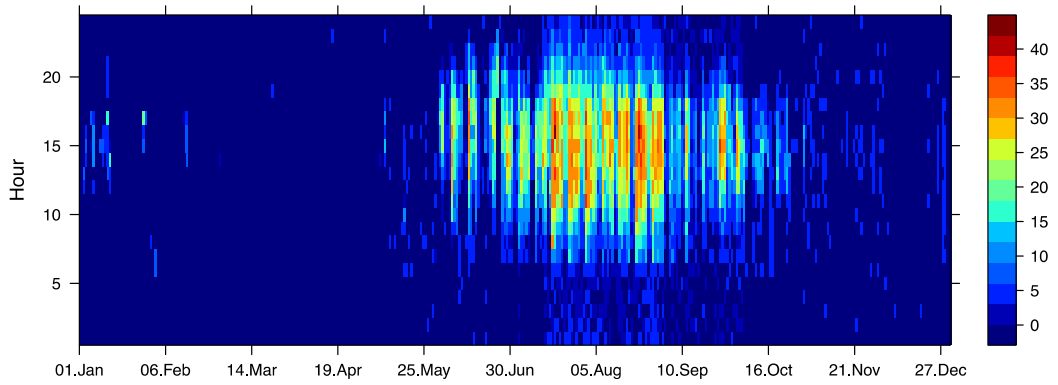

Figure 5–6 Annual carpet plot for measured cooling energy end use (hour of the day versus day of the year)

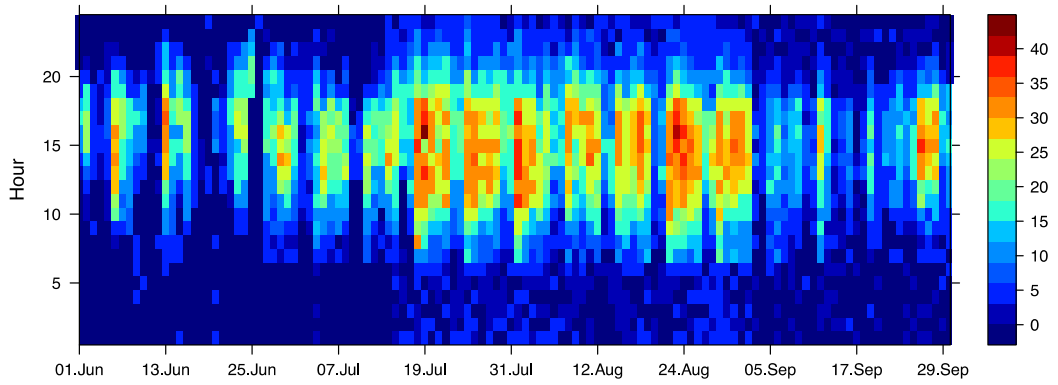

Figure 5–7 Summer carpet plot for measured cooling energy end use (hour of the day versus day of the year)

5.3 Building Energy Simulation Data

An analysis that is similar to the one prepared for the measured performance data is now presented for the building energy simulation data. The basis of the simulated data is a detailed eQuest building energy simulation of RSF I that uses Typical Meteorological Year 3 (TMY3) weather data for Boulder, Colorado.

5.3.1 Modeled Data Annual Box Plots

5.3.1.1 Average Daily Behavior

The modeled cooling energy consumption profile in Figure 5–8 shows peak annual values of around 60 kW electricity demand based on a simplified model assuming a constant COP of 7.8 and TMY3 weather data, which is roughly 20 kW higher than the measured 2011 data.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

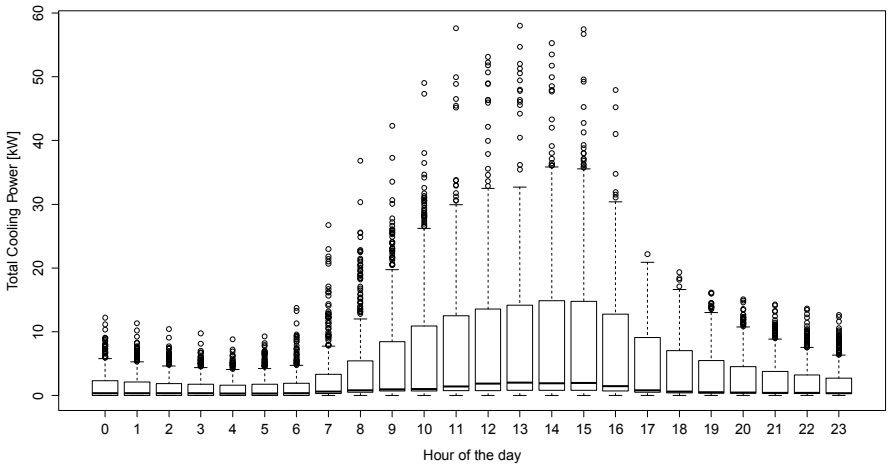

Figure 5–8 Modeled annual average daily cooling energy end use as a function of time of day

5.3.1.2 Separate Weekday and Weekend Behavior

Next, the data were disaggregated into weekday and weekend patterns to eliminate the variance in the box plots based on the difference in building use between weekdays and weekends (see Figure 5–9). Compared to the measured performance data, the diurnal profile for weekdays appears similar; the most striking difference is that the peak third quartile cooling hours occur earlier in the day. The third quartile values of the weekend cooling data, conversely, exhibit a more regular, approximately sinusoidal pattern.

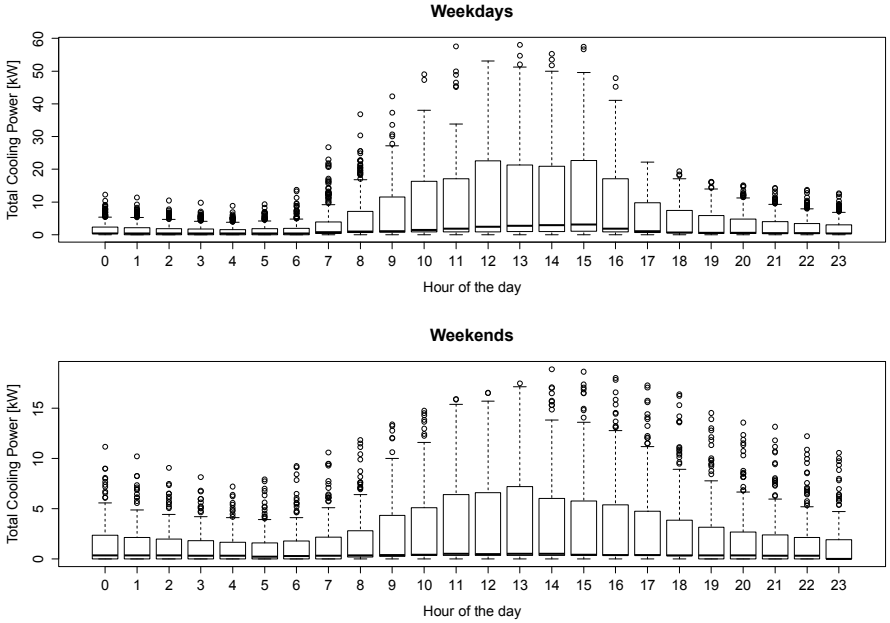

Figure 5–9 Modeled annual average weekday and weekend cooling energy end use as a function of time of day

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

5.3.2 Modeled Data Seasonal Box Plots

5.3.2.1 Average Daily Behavior

This section shows average daily profiles separated by winter and summer periods. Recall, summer is considered June through October (5 months); winter is considered January through May and November through December (7 months). The summer box plot shows a distinct profile, with median values rising from 7:00 to 12:00 and peaking during the early afternoon period of 13:00 to 15:00, thereafter falling with an exponential delay through the evening and night hours, with a minimum value observed at 4:00 (Figure 5–10).

The winter profile shows a pattern similar to the summer case; however, the values are significantly lower except for numerous outliers as high as 25 kW. Winter observations will not be further studied.

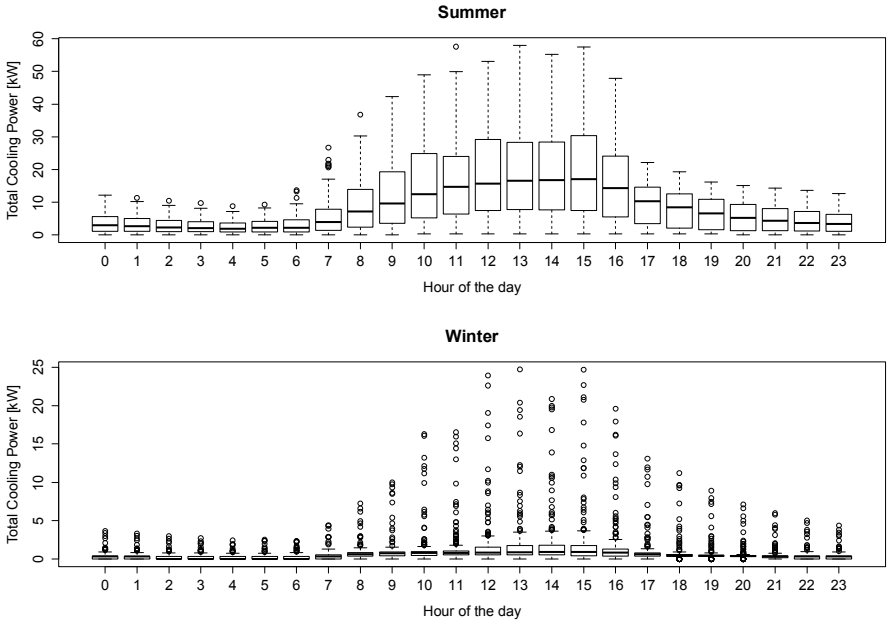

Figure 5–10 Modeled annual average summer and winter cooling energy end use as a function of time of day

5.3.2.2 Daily Behavior by Weekday During the Summer

Here, each day of the week during the summer is presented (Figure 5–11). Weekend days (Saturday and Sunday) again exhibit much lower cooling consumption than the weekdays, suggesting either day-typing into weekday and weekend days, or alternatively, distinguishing between occupied and unoccupied periods. Unlike the case of the measured performance data, no clear differences between the 5 weekdays can be observed in the simulated data.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

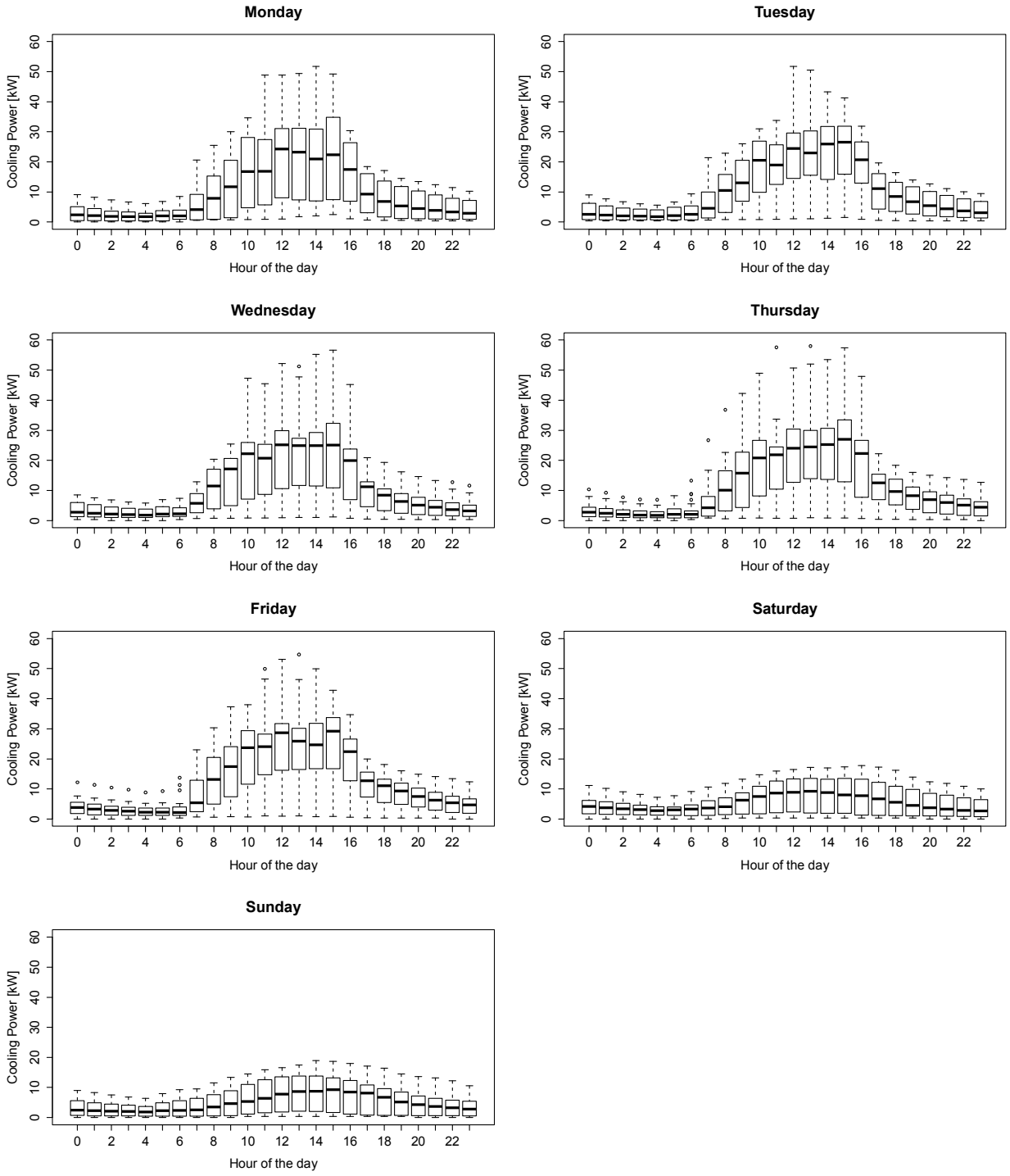

Figure 5–11 Measured average summer cooling energy end use by day of week as a function of time of day

5.3.3 Modeled Data Scatter Matrix Plots

The analysis thus far investigated temporal dependencies along season, day type, and hour of day. Here, dependencies of variables relative to each other are analyzed. After the data have been split into separate variables that contain either summer or winter data, scatter matrix plots

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

are created for summer and winter periods separately to find pairwise correlations (Figure 5–12). The correlation coefficients can be found in the upper panels and the scatter plots can be found in the lower panels.

During the summer, a strong correlation of 81% between dry-bulb ambient temperature (Temp.Sum) and modeled cooling power can be measured, and a weaker one (63%) between global horizontal solar and cooling power. Apparently, the stronger dependence on ambient dry-bulb temperature relative to the case of measured cooling data reveals the more deterministic nature of the simulation scenario relative to the noisier measured case.

As before, if only one independent variable can be chosen to model cooling energy demand during the summer, it should be ambient dry-bulb temperature.

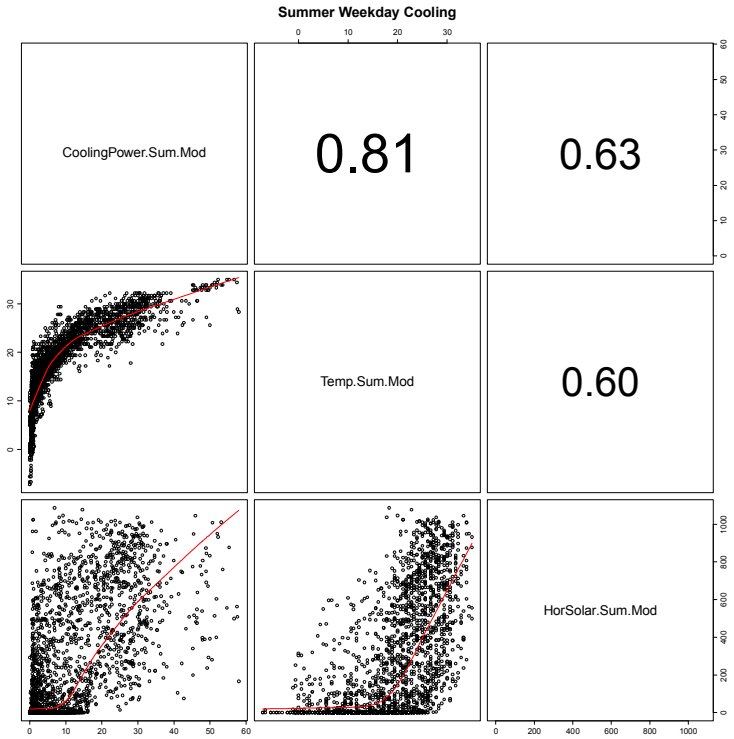

Figure 5–12 Pairwise correlation for modeled summer cooling energy end use for ambient dry-bulb temperature and global horizontal insolation

5.3.4 Carpet Plots

In this section, carpet plots are prepared for various time frames, allowing one to identify patterns easily and simply with visual inspection.

The annual carpet plot reveals the peak cooling demand to fall into the time period of mid-July to late August and from 10:00 to 17:00. Vertical lines show weekends and holidays. The summer carpet plot shows high cooling demand focused on a number of hot weeks during early and late July. Moreover, 2 hot weeks with high cooling energy demand occurred during late June and mid-September in the Boulder, Colorado TMY3 dataset. Naturally, because the simulation is based on TMY data rather than on an actual weather year, these differences in the occurrence of peak cooling periods are expected (Figure 5–13 and Figure 5–14).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

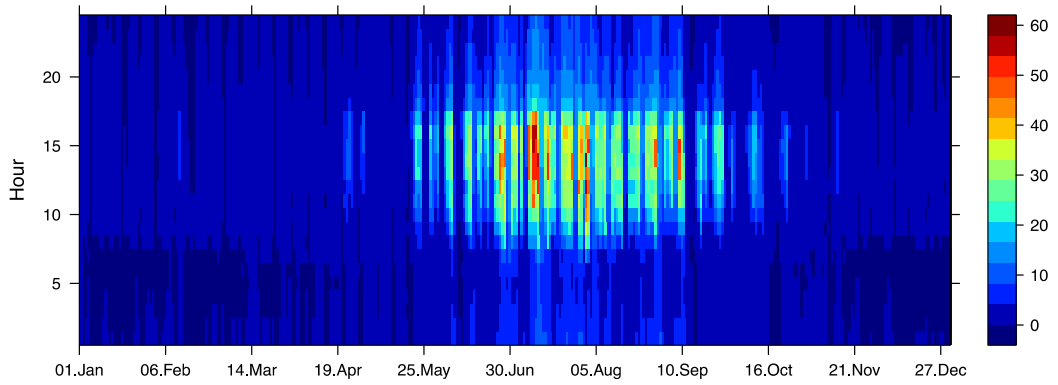

Figure 5–13 Annual carpet plot for modeled cooling energy end use (hour of the day versus day of the year)

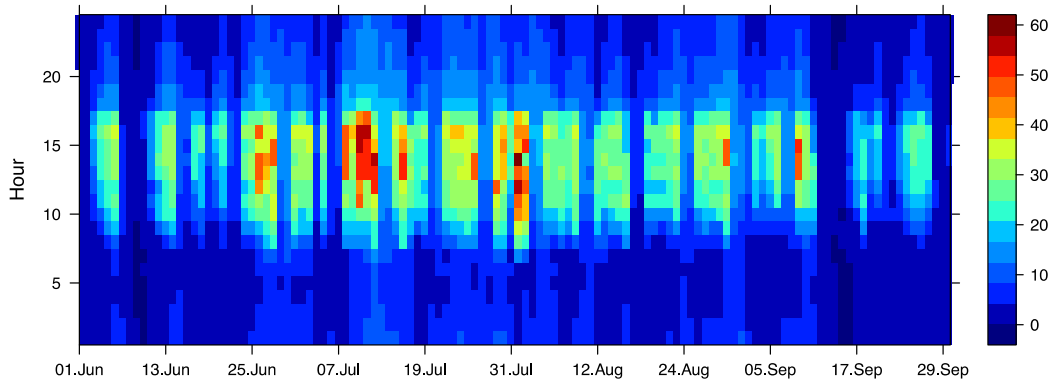

Figure 5–14 Summer carpet plot for modeled cooling energy end use (hour of the day versus day of the year)

5.4 Development of Frequency-Based Control Limits

In this section, the dominant trends discovered in the data exploration are revisited and improved models are suggested. As mentioned before, the process is illustrated by focusing on the cooling energy end use modeling. In contrast to the preceding discussion of trends in the observed data, the improved models were developed using the measured data for 2012 for the completed RSF consisting of RSF I (wings B and C) and RSF II (wing A). Using the more recent 2012 data may be preferable over using 2011 data.

5.4.1 Approach To Control Limit Development

A simple approach is suggested for the improved models: For each end use, ranges of expected values are generated for both empirical (measured) and modeled data as a function of the dominant independent (main effect) variable. The empirical ranges represent the historical performance of the building; modeled ranges represent the ideal performance of the building, based on a calibrated building model.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Measured building performance data are collected over a meaningful time horizon, such as the last year or last calendar year, here 2012. The modeled data are from a calibrated building energy model (eQuest); a PV system simulation conducted in Solar Advisor Model 2013 using the simplified PVWatts calculation engine was adopted to predict PV system electric power output. Only cooling energy end use is discussed in detail in this report, so the PV simulation does not come into play. Three potential factors would cause a divergence between the measured and modeled data:

- The energy model is for the B and C wings only.
- The measured data are for 2012 weather and occupancy patterns, but the model is using TMY3 weather and some other occupancy pattern.
- The eQuest model cannot model all the dynamic behavior of a thermodynamically complex building such as the RSF.

The second concern is partially addressed by presenting only static relationships between cooling energy consumption and weather variables; although the weather patterns were different, the static mapping is expected to be similar. A model is never perfect, but it still provides predictive and diagnostic capabilities.

For each end use and each data source (measured versus modeled), quantile tables are produced for relevant seasons and occupied periods using the main effect variables that have displayed the highest correlation with the particular end use: season, occupancy state, time of day, ambient dry-bulb temperature, global insolation on a vertical south-facing surface, and global horizontal insolation. Although the measured data include global insolation on a vertical surface facing south, these values had to be calculated for the modeled data using fundamental solar geometry relationships.

The generated quantile tables are then saved as comma separated values (CSV) files for future use by the Building Agent energy information system. We decided to use the lower quartile (25^{th} percentile) as the lower bound and the upper quartile (75^{th} percentile) as the upper bound, which leaves the range displayed as the central 50% of the expected values.

The following list of quantile tables for empirical data ("Obs" for observed or measured) and modeled ("Mod") data is saved for future analysis:

- Cooling by occupancy status as a function of ambient dry-bulb temperature: Cooling.Occ.Mod.csv, Cooling.Occ.Obs.csv, Cooling.Unocc.Mod.csv, Cooling.Unocc.Obs.csv
- Heating by occupancy status as a function of ambient dry-bulb temperature: Heating.Occ.Mod.csv, Heating.Occ.Obs.csv, Heating.Unocc.Mod.csv, Heating.Unocc.Obs.csv
- Lighting by season and occupancy status as a function of global insolation on a south-facing vertical surface: Lighting.Sum.Occ.Mod.csv, Lighting.Sum.Occ.Obs.csv, Lighting.Sum.Unocc.Mod.csv, Lighting.Sum.Unocc.Obs.csv, Lighting.Win.Occ.Mod.csv, Lighting.Win.Occ.Obs.csv, Lighting.Win.Unocc.Mod.csv, Lighting.Win.Unocc.Obs.csv

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

- Mechanical by day type (WD being weekday and WE being weekend) as a function of time of day: Mechanical.WD.Mod.csv, Mechanical.WD.Obs.csv, Mechanical.WE.Mod.csv, Mechanical.WE.Obs.csv
- Plug loads by day type (WD being weekday and WE being weekend) as a function of time of day: PlugLoads.WD.Mod.csv, PlugLoads.WD.Obs.csv, PlugLoads.WE.Mod.csv, PlugLoads.WE.Obs.csv
- PV System Output as a function of global horizontal insolation: PV.Mod.csv, PV.Obs.csv.

In essence, for every end use at any point in time, two control ranges are available. The question arises whether to use the measured or the modeled ranges. If only the measured (or observed) ranges are used, there is little to no encouragement to engage occupants to improve the energy consumption. Hence, a calibrated building model may provide more ambitious yet realistic performance goals. Of course, savings goals can also be articulated through a percentage reduction of the measured values, say 20%. However, as with the comparison for plug load and lighting energy end use (not shown in this report), the model is at times far off, either vastly underestimating or overestimating the actual building performance. If the model were predicting unreasonably high end-use values (higher than were attained in the past), there would be no incentive to improve. Conversely, if the model is predicting unreasonably low values, which cannot be attained without negatively impacting the building's primary purpose, the model suggestions are invalid.

We adopted the following approach to resolve this dilemma: The model suggested range is considered invalid whenever the lower observed limit (here lower quartile LQO) is greater than upper model limit (here upper quartile UQM) or when the upper empirical limit UQO is smaller than the lower model limit LQM. In case the model is invalid, the observed range is applied.

The model is considered valid in all other cases. In case the model is valid, the lower limit (LL) is the lesser of either the lower model or the lower observed limit and the upper limit (UL) is the smaller of either the upper model or the upper observed limit. In other words: If UQO > UQM & UQO < LQM, then LL = LQO and UL = UQO; else LL = min(LQO,LQM) and UL = min(UQO,UQM).

Figure 5–15 shows a hypothetical development of empirical and model ranges over time. Eight distinct time periods are indicated, from 1 to 8. Beginning at 1, the requirement for a valid model is fulfilled for periods 1 to 3. During phases 4 and 8, the model is invalid. The labels next to the arrows indicate the result of applying the above minimization statements.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

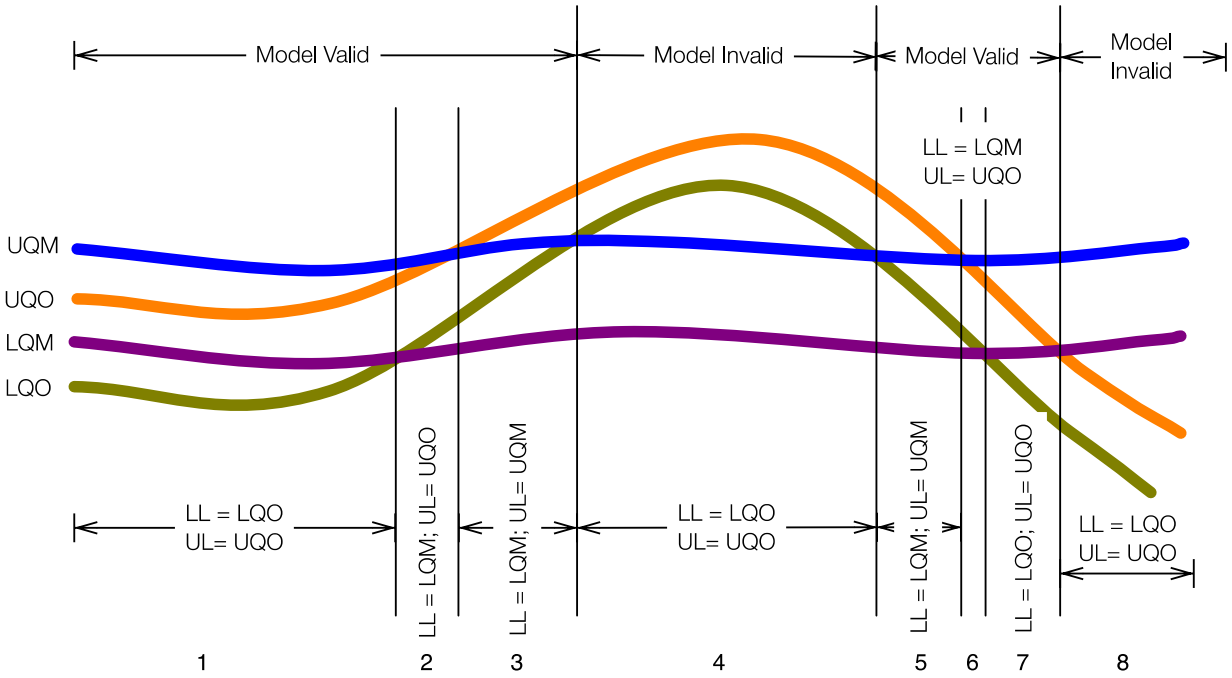

Figure 5–15 Illustration of selection of applicable LCL and UCL

When these rules are applied, the following list of results tables for measured data ("Obs" for observed) and modeled ("Mod") data are saved for future use by the Building Agent energy information system:

- Cooling during summer months and by occupancy status as a function of ambient dry-bulb temperature: Cooling.Sum.Occ.csv and Cooling.Sum.Unocc.csv as well as cooling during summer months without occupancy distinction: Cooling.Sum.csv. In addition, cooling for all months distinguished only by occupancy status as a function of ambient dry-bulb temperature: Cooling.Occ.csv and Cooling.Unocc.csv.
- Heating during winter months and by occupancy status as a function of ambient dry-bulb temperature: Heating.Win.Occ.csv and Heating.Win.Unocc.csv.
- Lighting by season and occupancy status as a function of global insolation on a south-facing vertical surface: Lighting.Sum.Occ.csv and Lighting.Sum.Unocc.csv as well as Lighting.Win.Occ.csv and Lighting.Win.Unocc.csv.
- Mechanical by day type (WD being weekday and WE being weekend) as a function of time of day: Mechanical.WD.csv and Mechanical.WE.csv.
- Plug loads by day type (WD being weekday and WE being weekend) as a function of time of day: PlugLoads.WD.csv and PlugLoads.WE.csv.
- PV System Output as a function of global horizontal insolation: PV.csv.

5.4.2 Cooling Energy Consumption Versus Ambient Temperature

The strongest correlation for cooling energy consumption was, not surprisingly, found for ambient air temperature. Restricting the analysis to summer months, the quantiles for all hours are presented first and then for occupied hours and for unoccupied hours during weekday nights

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

and weekends. In Figure 5–16 and Figure 5–17, the width of each box indicates the number of data available in the corresponding bin.

5.4.2.1 All Summer Hours

Developing so-called energy signatures for cooling energy end use; i.e., a depiction of the cooling demand as a function of ambient dry-bulb temperature, enables the analyst to detect weather-dependent and weather-independent end use components. An inspection of the measured cooling energy end use versus ambient temperature for all summer hours revealed somewhat irregular consumption patterns for ambient temperatures below 20.8°C, albeit at a very low level. A nearly linear increase of total cooling power for the RSF with ambient temperature can be observed for temperatures above 20.8°C.

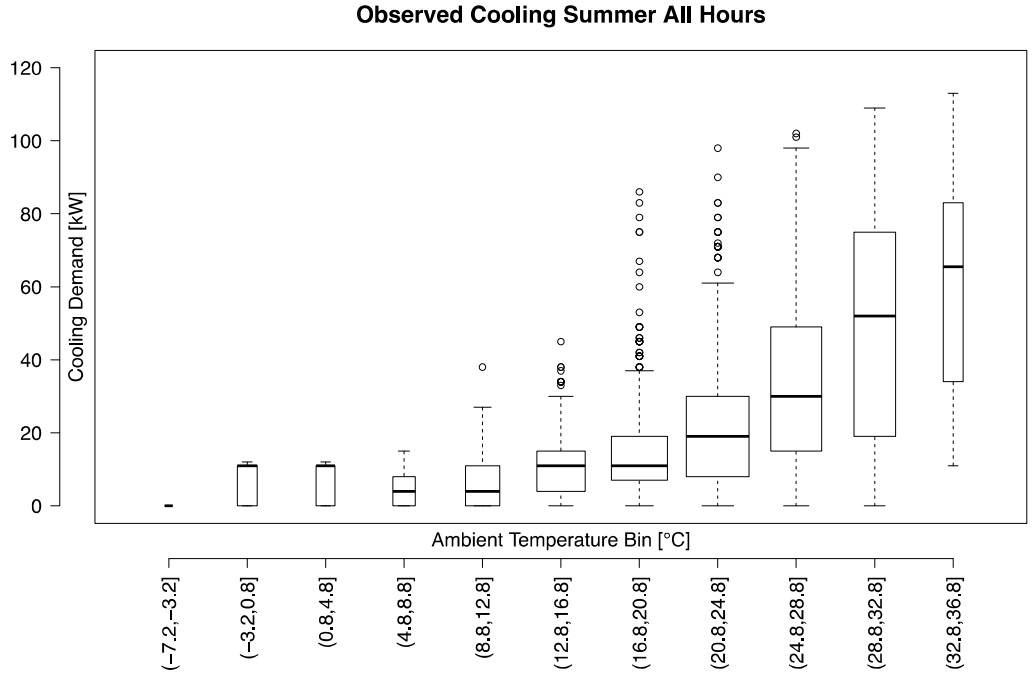

Figure 5–16 Measured cooling energy end use versus ambient temperature for all summer hours

On the other hand, inspecting the modeled cooling energy end use versus ambient temperature for all summer hours, one can detect a much smoother development of the median total cooling power. In both the measured and modeled cases, a joint treatment of all summer hours leaves a large number of outliers, which will be remedied below by distinguishing between occupied and unoccupied hours.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

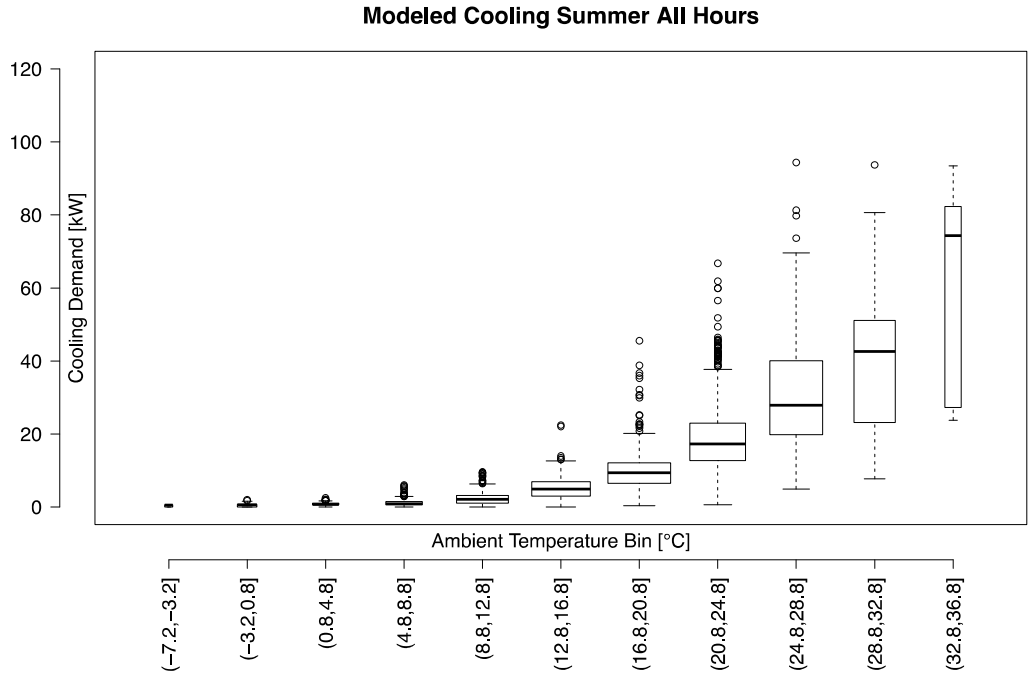

Figure 5–17 Modeled cooling energy end use versus ambient temperature for all summer hours

Applying the energy-conservative control logic introduced above, the following table of control limits can be generated for all summer hours. The LCL corresponds to the first quartile and the UCL to the third quartile. Table 5–1 has been implemented in the Building Agent energy information system by linearly interpolating the control limits for the currently observed ambient temperature between the midpoints of two adjacent ambient temperature bins.

Table 5–1 Cooling Energy End-Use UCLs and LCLs (kW) Versus Ambient Temperature (°C) for All Summer Hours

Bin	(–7.2, –3.2)	(-3.2, 0.8)	(0.8, 4.8)	(4.8, 8.8)	(8.8, 12.8)	(12.8, 16.8)	(16.8, 20.8)	(20.8, 24.8)	(24.8, 28.8)	(28.8, 32.8)	(32.8, 36.8)
1st Quartile	0	0	0	0	0	3	6	8	15	19	27
3rd Quartile	0	1	1	2	3	7	12	23	40	51	82

5.4.2.2 Occupied Summer Hours

Prompted by the large number of outliers for the measured and modeled cooling energy signatures, we separated the datasets into occupied and unoccupied summertime periods (Figure 5–18 and Figure 5–19). For higher ambient temperatures, the occupied summer hours are associated with higher cooling energy consumption than the corresponding data for all summer hours.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

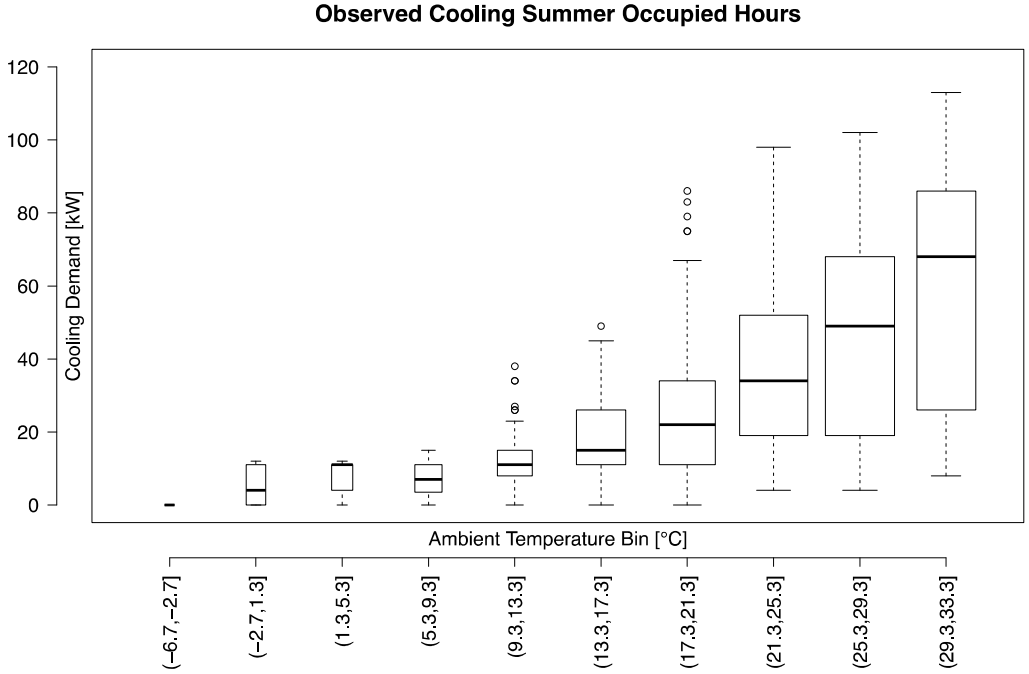

Figure 5–18 Measured cooling energy end use versus ambient temperature for occupied summer hours

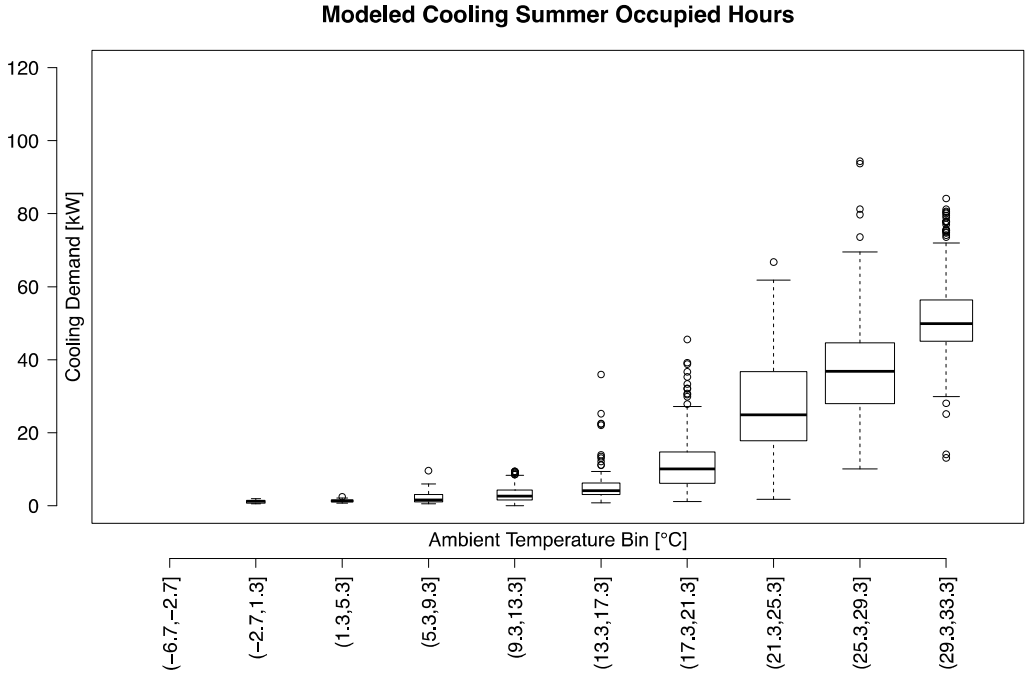

Figure 5–19 Modeled cooling energy end use versus ambient temperature for occupied summer hours

The energy-conservative control logic can be applied to generate the following table of control limits for occupied summer hours. Table 5–2, Figure 5–20, and Figure 5–21 require the occupancy status of the building and the ambient temperature.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Table 5–2 Cooling Energy End-Use UCLs and LCLs (kW) Versus Ambient Temperature (°C) for Occupied Summer Hours

Bin	(–6.7, –2.7)	(-2.7, 1.3)	(1.3, 5.3)	(5.3, 9.3)	(9.3, 13.3)	(13.3, 17.3)	(17.3, 21.3)	(21.3, 25.3)	(25.3, 29.3)	(29.3, 33.3)
1st Quartile	0	0	4	4	8	11	6	18	19	26
3rd Quartile	0	1	11	11	15	26	15	37	45	56

5.4.2.3 Unoccupied Summer Hours

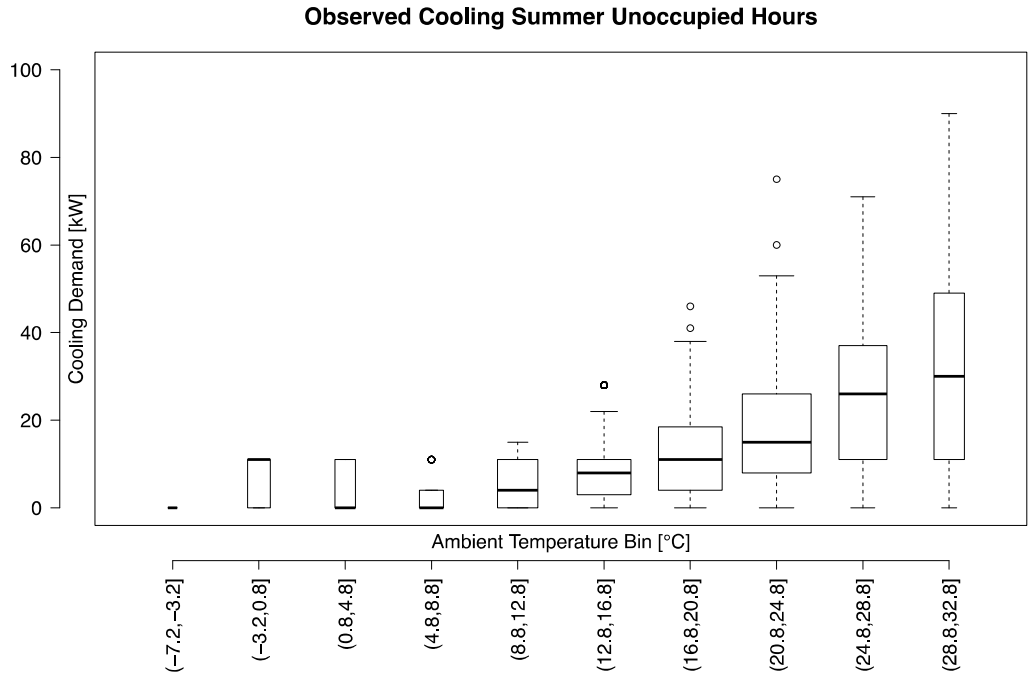

Figure 5–20 Measured cooling energy end use versus ambient temperature for unoccupied summer hours

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

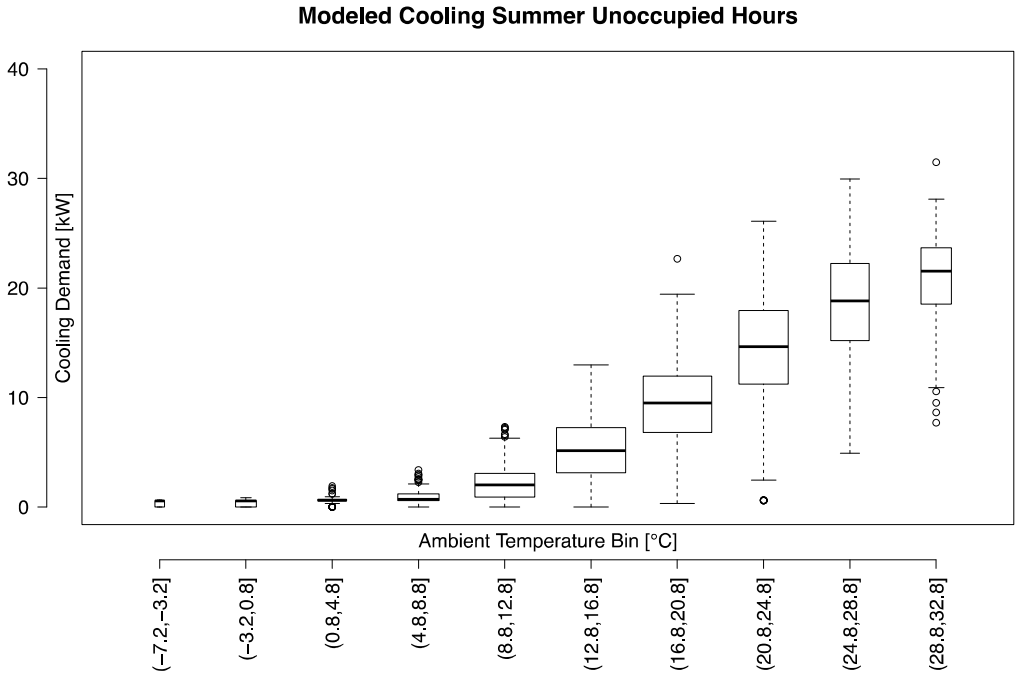

Figure 5–21 Modeled cooling energy end use versus ambient temperature for unoccupied summer hours

The energy-conservative control logic can be applied to generate the following table of control limits for unoccupied summer hours. Again, Table 5–3 requires the occupancy status of the building and the ambient temperature.

Table 5–3 Cooling Energy End Use UCLs and LCLs (kW) Versus Ambient Temperature (°C) for Unoccupied Summer Hours

Bin	(–7.2, –3.2)	(–3.2, 0.8)	(0.8, 4.8)	(4.8, 8.8)	(8.8, 12.8)	(12.8, 16.8)	(16.8, 20.8)	(20.8, 24.8)	(24.8, 28.8)	(28.8, 32.8)
1st Quartile	0	0	0	0	0	3	4	8	11	11
3rd Quartile	0	1	1	1	3	7	12	18	22	24

5.4.2.4 All Occupied Hours

After this first approach in the Building Agent environment was implemented, it became apparent that the seasonal distinction did not offer performance improvements beyond the selection of ambient dry-bulb temperature as the independent variable. Thus, separate energy signatures for total RSF cooling power for all occupied and unoccupied hours for observed and modeled data were prepared.

The occupied period data show that the observed energy signature rises steeply for lower ambient temperatures than the model predicts: the positively sloped portion of the energy signature begins with the (11,15) °C bin, whereas in the case of the modeled data the rise begins not before the (15,19) °C bin, indicating a lower balance point temperature for the actual RSF than the model (Figure 5–22 through Figure 5–25). The lower balance point temperature is likely due to higher cumulative internal and equipment heat gains.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

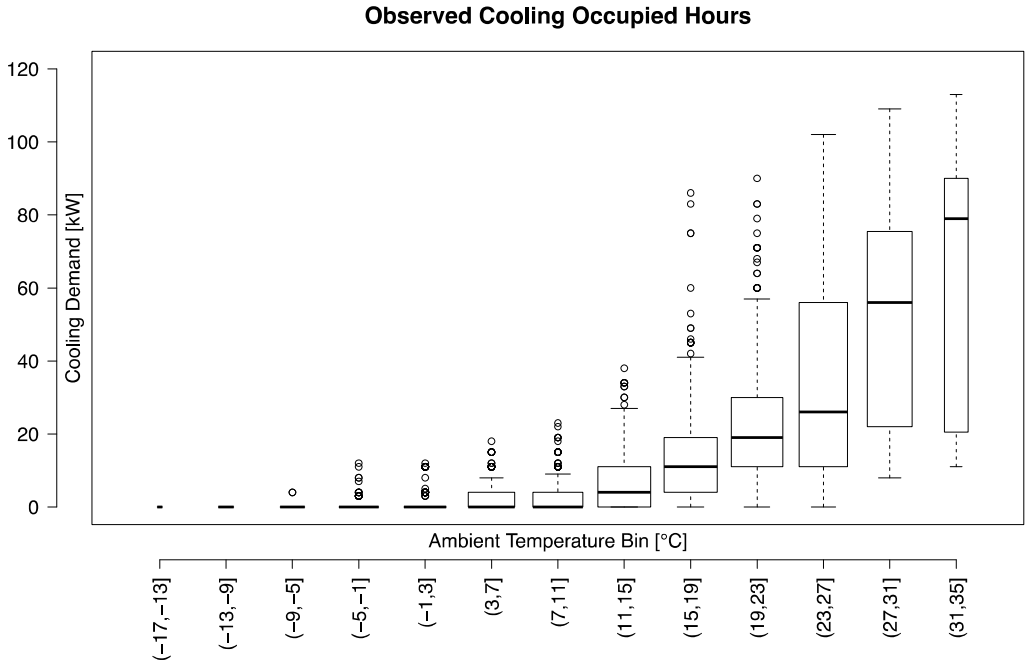

Figure 5–22 Measured cooling energy end use versus ambient temperature for occupied hours

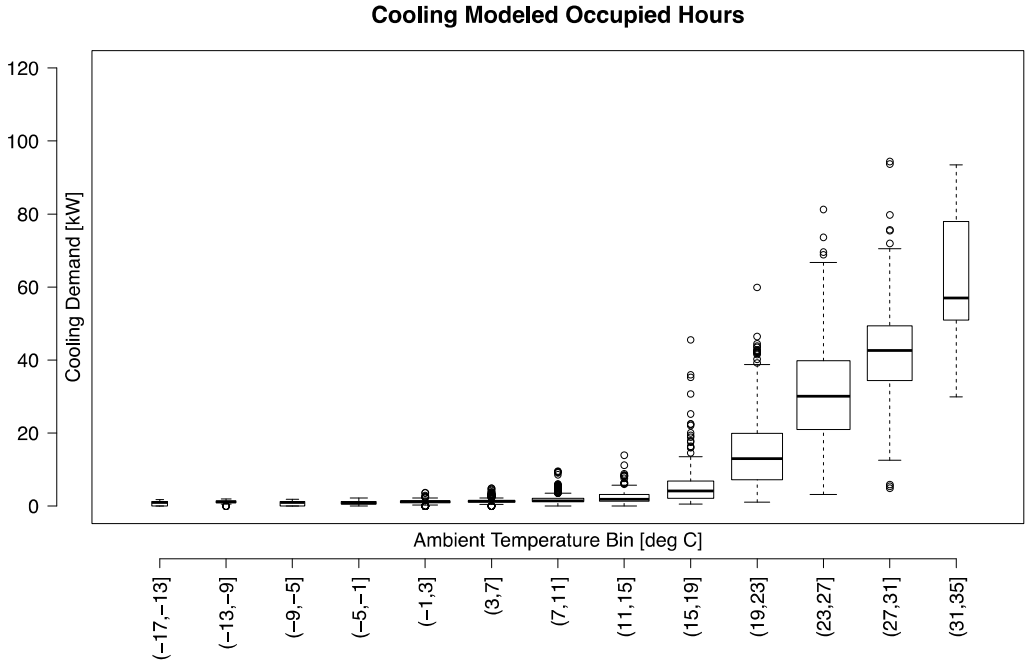

Figure 5–23 Modeled cooling energy end use versus ambient temperature for occupied hours

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Table 5–4 Cooling Energy End Use UCLs and LCLs (kW) Versus Ambient Temperature (°C) for Occupied Hours

Bin	(–17, –13)	(–13, –9)	(–9, –5)	(–5, –1)	(–1, 3)	(3, 7)	(7, 11)	(11, 15)	(15, 19)	(19, 23)	(23, 27)	(27, 31)	(31, 35)
1st Quartile	0	0	0	0	0	0	0	0	2	7	11	22	20
3rd Quartile	0	0	0	0	0	2	2	3	7	20	40	49	78

5.4.2.5 All Unoccupied Hours

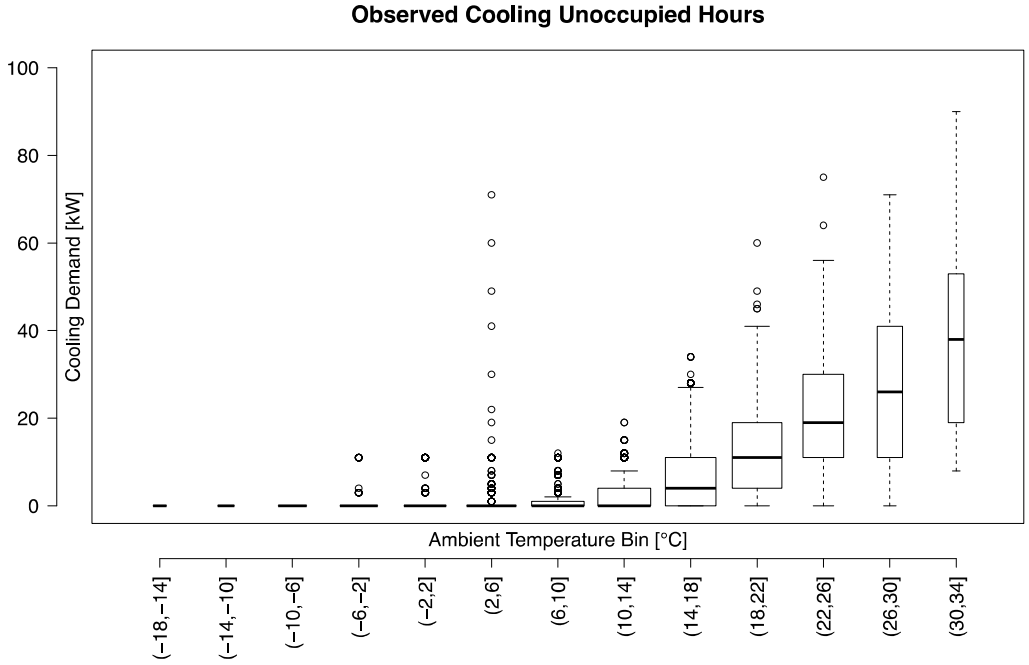

Figure 5–24 Measured cooling energy end use versus ambient temperature for unoccupied hours

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

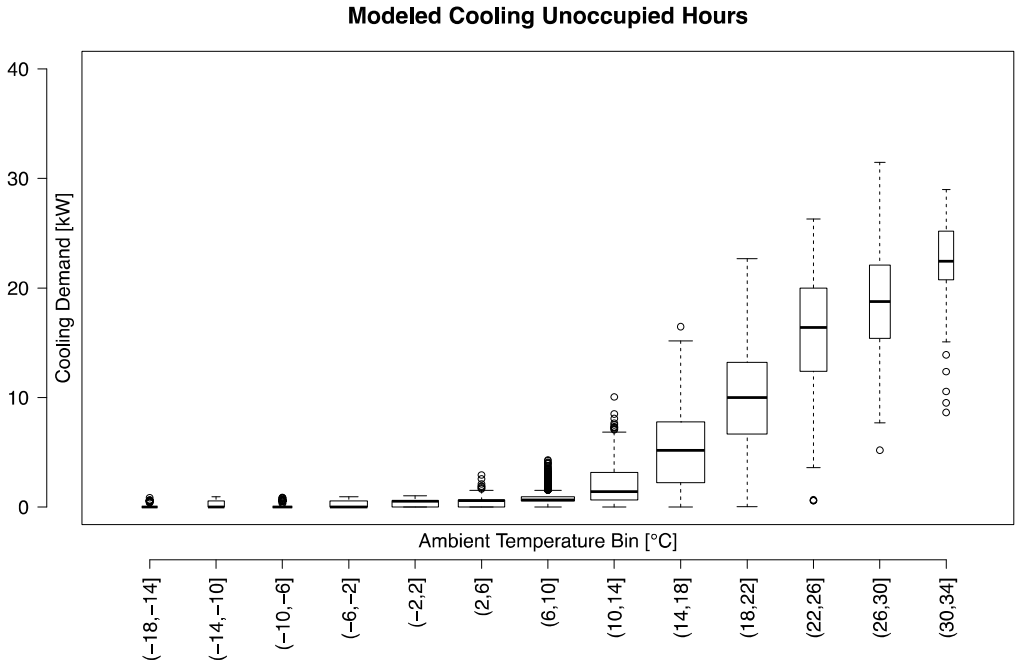

Figure 5–25 Modeled cooling energy end use versus ambient temperature for unoccupied hours

Table 5–5 Cooling Energy End-Use UCLs and LCLs (kW) Versus Ambient Temperature (°C) for Unoccupied Hours

Bin	(–18, –14)	(–14, –10)	(–10, –6)	(–6, –2)	(–2, 2)	(2, 6)	(6, 10)	(10, 14)	(14, 18)	(18, 22)	(22, 26)	(26, 30)	(30, 34)
1st Quartile	0	0	0	0	0	0	0	0	0	4	11	11	19
3rd Quartile	0	0	0	0	0	0	1	3	8	13	20	22	25

5.5 Energy Dashboard Implementation

The control limits shown in Table 5–4 and Table 5–5 have been implemented in the Building Agent system. Figure 5–26 shows the instantaneous cooling energy end use and Figure 5–27 the historical development. The gray band in Figure 5–26 and Figure 5–27 is defined by the LCLs and UCLs developed in this section. Evidently, for the time period shown, the building exhibited cooling energy consumption above the UCL.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Figure 5–26 Dashboard view of instantaneous total cooling power (kW) in the RSF along with LCLs and UCLs shown as a gray band above

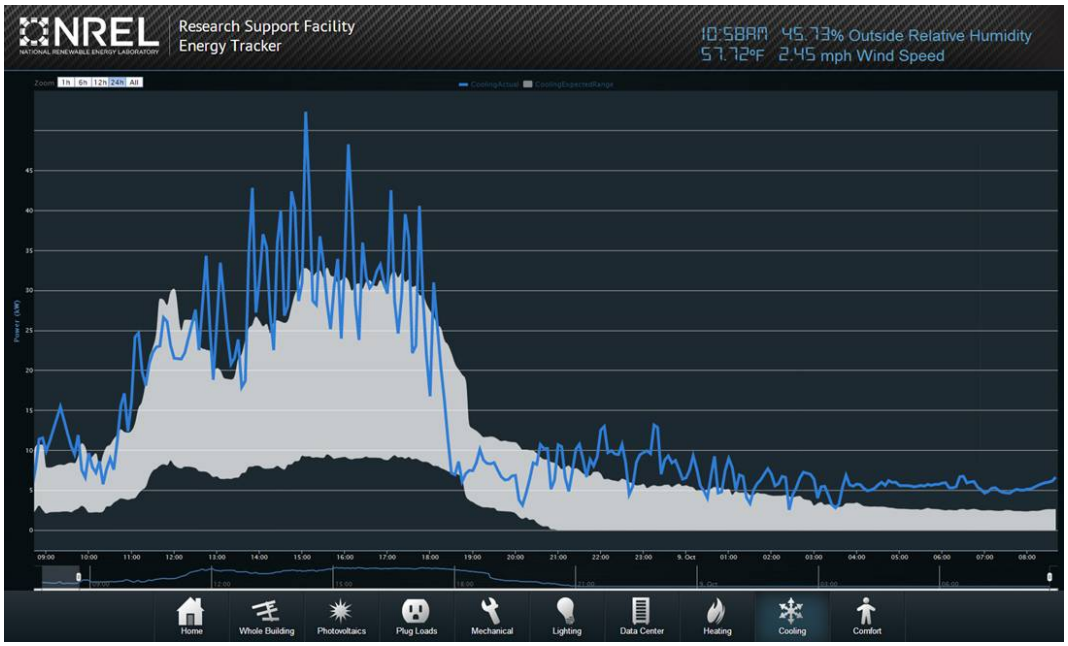

Figure 5–27 Dashboard view of historical total cooling power (kW) (blue) development in the RSF along with control limits

5.5.1 Comparison With Engineering Judgment Method

A comparison between the first engineering judgment-based method and the second quantile analysis shown in Figure 5–28 reveals that (1) significant subjective judgment represented by a simplistic linear representation in first approach did not capture mild weather hours cooling performance; and (2) the second quantile analysis method proved to be as effective as the first, and was often a more precise method with tighter acceptable control limits, offering better insight into operations (see Figure 5–29).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

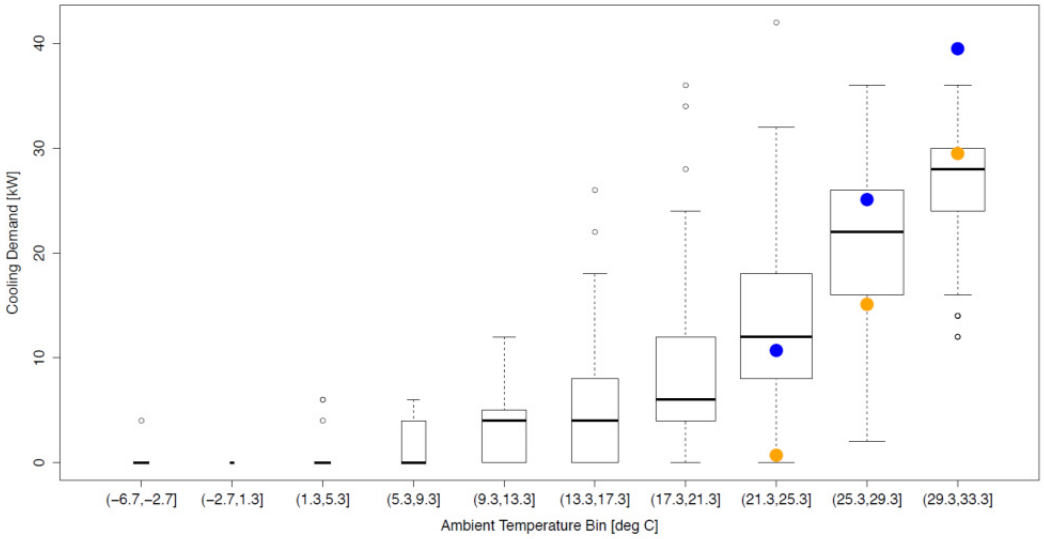

Figure 5–28 Modeled cooling summer occupied hours from quantile analysis compared with UCLs (blue dot) and LCLs (orange dot) from engineering judgment method

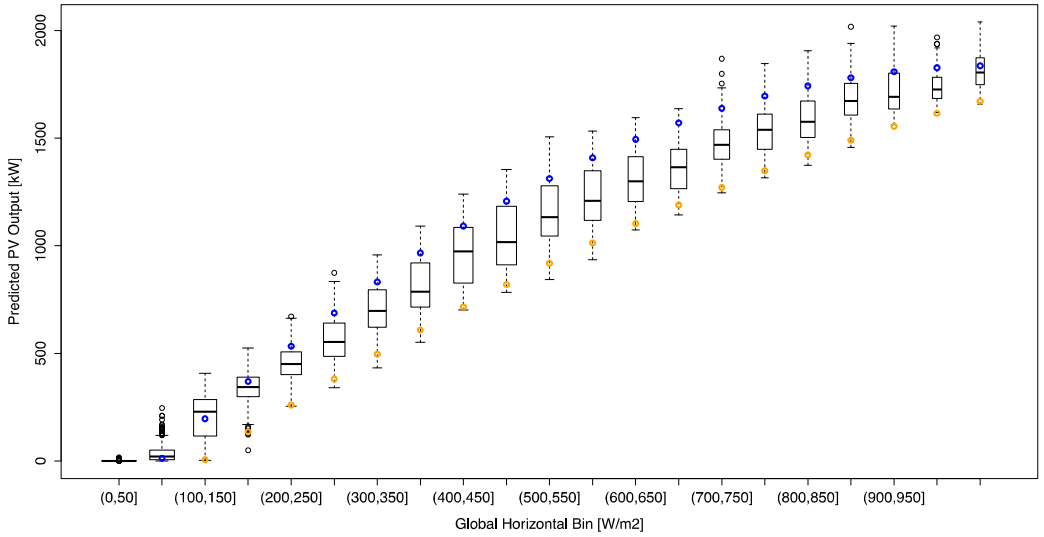

Figure 5–29 Modeled PV quantile analysis compared with UCLs (blue dot) and LCLs (orange dot) from engineering judgment method

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

6.0 Multivariate Control Limits Based on Quantile Regression

6.1 Motivation

In the second method presented, control limits were defined as lower and upper quartiles of empirical distributions of energy end use versus the dominant independent variable; e.g., ambient dry-bulb temperature or time of day. The resulting nonlinear functions were saved as tables for implementation in the energy dashboard. Forcing these control limits onto univariate relationships implies, however, that other available explanatory variables or covariates are ignored and requires that the dominant independent variable be discovered manually in advance.

A universally applicable methodology is desired that harnesses *all* the available covariates of an energy end use to predict UCLs and LCLs as functions of all the relevant independent variables, not just the dominant one. In essence, a methodology is sought that can be applied to any situations where control limits on energy end use are needed, whether a sophisticated monitoring system makes dozens of variables available or whether only a few points are monitored.

6.2 Background on Quantile Regression

Quantile regression analysis is used in statistics and econometrics. Whereas the method of ordinary least squares (OLS) results in estimates that approximate the conditional mean of the response variable given certain values of the predictor variables, quantile regression aims to estimate either the conditional median or other quantiles of the response variable. In this project, the control limits on the energy end use are defined as the quantiles of interest.

One advantage of quantile regression, relative to OLS regression, is that the quantile regression estimates are more robust against outliers in the response measurements. Yet, the main attraction of quantile regression is that different measures of central tendency and statistical dispersion are available to conduct a more comprehensive analysis of the relationship between variables.

Although statistical models arising from the method of least squares can be solved using numerical linear algebra, quantile regression models lead to linear programming problems solved using either the simple method in case of moderate sized problems or using the interior point method for large data problems. As implemented in the statistical computing environment R used for this work, the quantile regression function computes an estimate on the n-th conditional quantile function of the response, given the covariates, as specified by a formula argument. As in the case of linear regression, the quantile regression function assumes a linear specification for the quantile regression model; i.e., that the formula defines a model that is linear in its parameters, not necessarily in its model structure.

6.3 Methodology

6.3.1 Data Analysis Procedure

As shown in Figure 6–1, the process adopted begins with the acquisition of end use and weather data for the actual building and the detailed building simulation model. The datasets are disaggregated by building operational mode (day type, occupancy status, and season) and any multicollinearity is removed from the datasets by means of principal component analysis (PCA) explained below. The significant PCs are retained for construction of linear models, while reducing the complexity of the model using stepwise regression that minimizes the Bayesian Information Criterion (BIC) as the performance/complexity metric. Next, based on the

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

parsimonious linear models attained through stepwise regression, models for preselected conditional quantiles are generated for the monitored and the building simulation data.

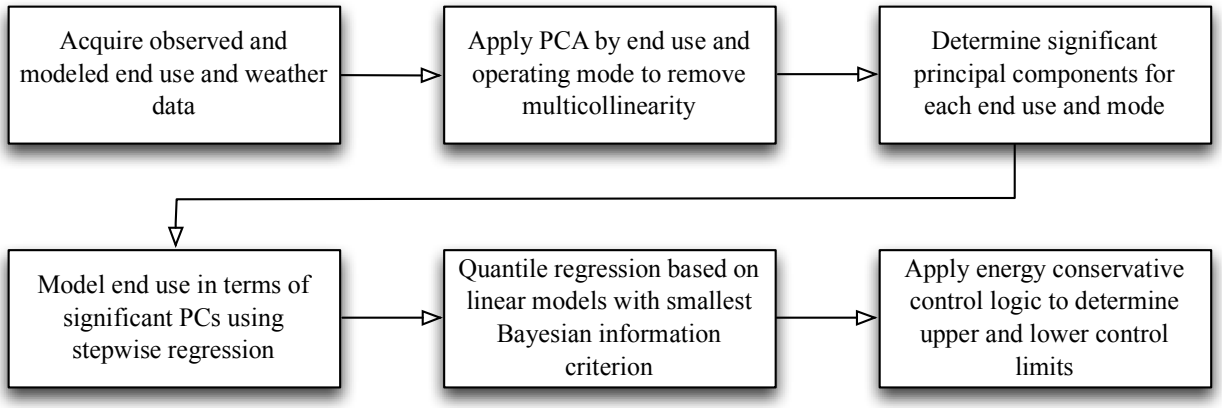

Figure 6–1 Flowchart of analysis for developing individual end use control limits based on quantile regression

The last step of the data analysis procedure involves applying the same energy-conservative data logic between the measured and the modeled data of the building, as in the case of frequency-based quantile analysis. The following sections apply this analysis and control limit estimation method to cooling energy use.

6.3.2 End-Use Mode Definition

Within a given end use, further classifying data of similar types into exhaustive, mutually beneficial, but fundamentally different groups or "modes" can be useful. The analysis in this report uses three such modes:

- Day type (weekday or weekend)
- Occupancy state (occupied or unoccupied)
- Season (winter or summer).

Day type is determined by day of the week, with Saturday and Sunday considered "weekend" days, and all others "weekdays." Occupancy state is determined by hour of the day; occupied hours are taken as 6:00 to 18:00; all other hours are taken as unoccupied. Lastly, season is defined by month, with summer taken as June–October, and all other months taken as winter. Although the seasonal distinction has been available, seasonal models for summer cooling and winter cooling have not been developed; rather, one model was developed for occupied periods and one for unoccupied periods, and for measured and modeled data.

6.3.3 Model Validation

Because the time series of the data and subsequent models developed are complex (although no true time series analysis is performed), this model is partially validated in two ways:. (1) the standard errors of quantile regression fits are computed using Markov Chain Marginal Bootstrapping, which provides a measure of confidence for the UCLs and LCLs; and (2) overall model quality is assessed using a combination of the R^2 values for linear models of individual end-use modes, model relevance as indicated by the F test, and finally visual assessment when compared to measured data.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

6.4 Measured Cooling Data

6.4.1 Occupied Periods

Independent variables used in multivariate regression are often not independent; rather, they are collinear. That is, various regressors share similar behavior. Thus, models built may have a good fit but may be suspect as predictive models because model coefficients and predictions show large standard errors and uncertainty bands, which lead to unstable predictions. Also, in models with collinear regressors, the regression coefficients are no longer proper indicators of relative physical importance. PCA is the most popular method for removing such effects of collinearity and summarizing the variance in the regressor set. This reduces the dimensionality of the multivariate datasest and increases model robustness.

PCA is a technique for singular decomposition of covariance matrices that effectively ranks the dominant modes of variation, or PCs, in a set of data. One first constructs a covariance matrix of potential predictor variables (variables are first normalized to a mean of 0 and a standard deviation of 1). The scaled variables are contained in a matrix, X, with variables separated by column. The covariance matrix, S, of these data is then given by

$$S = \frac{1}{N-1} X^T X$$

in which $N-1$ represents the degrees of freedom for the variance measure, N being the total number of observations in the training dataset. The resulting square covariance matrix is then decomposed by solving the eigenvalue problem

$$SE = \lambda E$$

The resulting diagonal eigenvalue matrix, λ, contains the variance present in each PC of the data. The eigenvectors for each PC are the columns of the square matrix E, and provide a set of weights on the variables in the original dataset that can be used to easily identify key predictor variables. The original dataset can further be transformed into the "PC space" by projecting the original dataset onto the new set of coordinate axes in the eigenvector matrix

$$P = XE$$

The matrix P contains the PCs of X and is the same size as the original dataset. Because of the eigen decomposition and coordinate transformation, the PCs have the unique and desirable properties of being orthogonal and uncorrelated. As a result, all multicollinearity between the PCs is eliminated, and they can be used as predictors in regression models with less ambiguity as to which variable combinations form the best set.

Then, the significant leading PCs are selected as those that contribute more than the average share of variance description (Figure 6–2).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

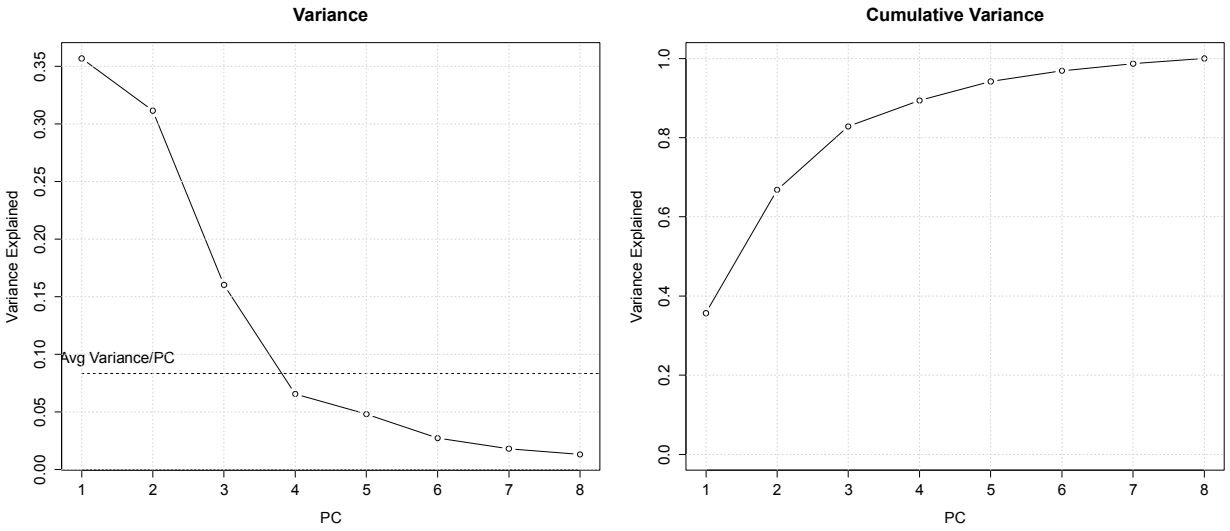

Figure 6–2 Individual and cumulative variance explained by PCs for measured cooling energy end use during occupied periods

Thus, the dimensionality of measured occupied cooling variables is reduced from 8 to 3. It is also useful to examine the loading factor for the first two significant PCs to assess the validity of the orthogonalization in terms of known influences on this particular end use. Figure 6–3 shows the loading factors for the most descriptive PCs.

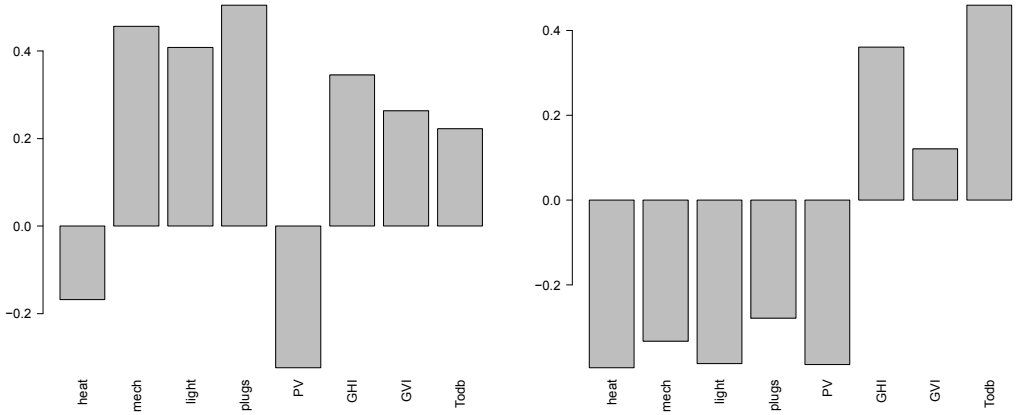

Figure 6–3 Loading factors for the two dominant PCs for measured cooling energy end use during occupied periods

In the case of the measured RSF data, only limited weather data are available. The dominant loadings of the first PC are associated with the mechanical, light, and plug load end uses, followed by horizontal and vertical insolation and ambient temperature. The loadings of the second PC emphasize ambient temperature and global horizontal radiation, followed by the other end uses with the opposite sign.

Next, these significant PCs are used to determine a parsimonious model for measured occupied cooling energy use using stepwise regression. Explanatory variables used in the regression are

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

the dominant PCs identified in the PCA and a set of exogenous cycle variables for hour of the day h and day of the year d.

$$C_{h,1} = \sin\frac{2\pi h}{24};\ C_{h,2} = \cos\frac{2\pi h}{24}; C_{d,1} = \sin\frac{2\pi d}{365};\ C_{h,2} = \cos\frac{2\pi d}{365}$$

These cycle variables are designed to capture the diurnal and annual periodicities in the energy end use development and significantly improve model performance. They are considered exogenous, because they are process independent and independently available.

The linear model could contain any of the variables and their first- and second-order interaction terms for the significant PCs, and the stepwise regression is conducted using the BIC and not Akaike's Information Criterion. BIC is a cost-complexity metric used to find a parsimonious model that balances model prediction performance with model complexity,

$$BIC = -2\ln L + k\ln N$$

where L is the maximized value of the likelihood function for the estimated model, k the number of free parameters to be estimate, and N the number of observations. Assuming that the model errors are independent and identically distributed according to a normal distribution and the boundary condition that the derivative of the log likelihood with respect to the true variance is zero, this becomes (up to an additive constant, which depends only on N and not on the model),

$$BIC = N\ln\sigma_\varepsilon^2 + k\ln N$$

where σ_ε^2 is the error variance

$$\sigma_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x_i})^2$$

and $\widehat{x_i}$ is the predicted and x_i the measured data. The stepwise regression attempts to minimize BIC. In general, BIC penalizes free parameters more strongly than does Akaike's Information Criterion.

Stepwise regression is a powerful model building technique that combines backward elimination and forward selection processes. Initially, the correlation coefficient between the response variable and each regressor is computed. The most highly correlated regressor enters the regression, OLS minimization is applied, and the goodness of fit (here BIC) is computed. Then, the second most highly correlated regressor is added to the first one and OLS is conducted again. For each regressor, the Student-*t* statistic is computed to test for the significance of each coefficient. Insignificant regressors are removed. The process ends when no more regressors enter or leave the OLS process.

Figure 6–4 shows a summary of the linear model created by the process. The model contains as regressors the individual PCs PC1, PC2, and PC3, three of the cycle variables, all first-order interaction terms between the PCs and the cycle variables, as well as all first-order interaction terms between the three cycle variables. Each of the 18 model variables is significantly different from zero, as confirmed by high Student-*t* scores and associated low Type I errors. Moreover, the overall model exhibits a very high *F*-statistic of 915, which indicates that it is significantly different from a random model and the R^2 statistic of 79% reveals that 79% of the variance in the data can be explained by the linear model.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

```
## lm(formula = meas.cool.occ$cool ~ PC1 + PC2 + PC3 + sinH + sinD +
##     cosD + PC1:sinH + PC1:sinD + PC1:cosD + PC2:sinH + PC2:sinD +
##     PC2:cosD + PC3:sinH + PC3:sinD + PC3:cosD + sinH:sinD + sinH:cosD +
##     sinD:cosD, data = meas.cool.occ.rot.exp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.141  -2.030   0.029   1.696  21.686
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0090     0.1035   19.40  < 2e-16 ***
## PC1           1.5324     0.0448   34.22  < 2e-16 ***
## PC2           1.0176     0.0723   14.08  < 2e-16 ***
## PC3           2.4601     0.1073   22.93  < 2e-16 ***
## sinH         -1.6226     0.1246  -13.02  < 2e-16 ***
## sinD         -3.0610     0.1014  -30.20  < 2e-16 ***
## cosD         -1.7520     0.1784   -9.82  < 2e-16 ***
## PC1:sinH     -0.7632     0.0609  -12.54  < 2e-16 ***
## PC1:sinD     -1.2995     0.0602  -21.60  < 2e-16 ***
## PC1:cosD     -1.5173     0.0530  -28.65  < 2e-16 ***
## PC2:sinH     -0.3537     0.0792   -4.46  8.2e-06 ***
## PC2:sinD     -0.4521     0.0845   -5.35  9.1e-08 ***
## PC2:cosD     -0.5519     0.0674   -8.19  3.3e-16 ***
## PC3:sinH     -0.5968     0.1126   -5.30  1.2e-07 ***
## PC3:sinD     -1.3785     0.1189  -11.59  < 2e-16 ***
## PC3:cosD     -2.5909     0.1096  -23.63  < 2e-16 ***
## sinH:sinD     1.3590     0.1481    9.18  < 2e-16 ***
## sinH:cosD     0.7609     0.1875    4.06  5.0e-05 ***
## sinD:cosD     3.6440     0.2652   13.74  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.79 on 4361 degrees of freedom
## Multiple R-squared: 0.791,Adjusted R-squared: 0.79
## F-statistic:  915 on 18 and 4361 DF,  p-value: <2e-16
```

Figure 6–4 Linear model structure summary report for measured cooling energy end use during occupied periods

The model residuals are analyzed using a quantile-quantile or Q-Q plot (see Figure 6–5), which is a plot of the quantiles of two distributions against each other. The pattern of points in the plot is used to compare the two distributions. Linear regression theory dictates that the residuals should be normally distributed with zero mean. Deviations from this assumption are revealed in a Q-Q plot by the deviations of the actual residual quantile distribution from the assumed normal distribution. The deviations of the model residuals from the assumed normal deviations are fairly significant, indicating that the residuals exhibit nonuniformity. Causes for nonuniform residuals include missing regressor variables, outliers, heteroscedasticity (nonconstant variance in the residuals), inappropriate model structure, and serial (auto) correlation.

The Durbin–Watson (DW) test is applied to test for first-order autocorrelation of the residuals. If no serial or autocorrelation is present, the expected value of DW is 2. If the model underfits, DW would be less than 2; it would be greater than 2 for an overfitted model, the limiting range being 0–4. If the DW statistic is substantially less than 2, there is evidence of positive serial correlation. As a rough rule of thumb, if DW is less than 1.0, there may be cause for alarm. Small DW values indicate that successive error terms are, on average, close in value to one another, or positively correlated. The model reveals a score of 0.59 and is therefore underfitting with positively correlated residuals.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

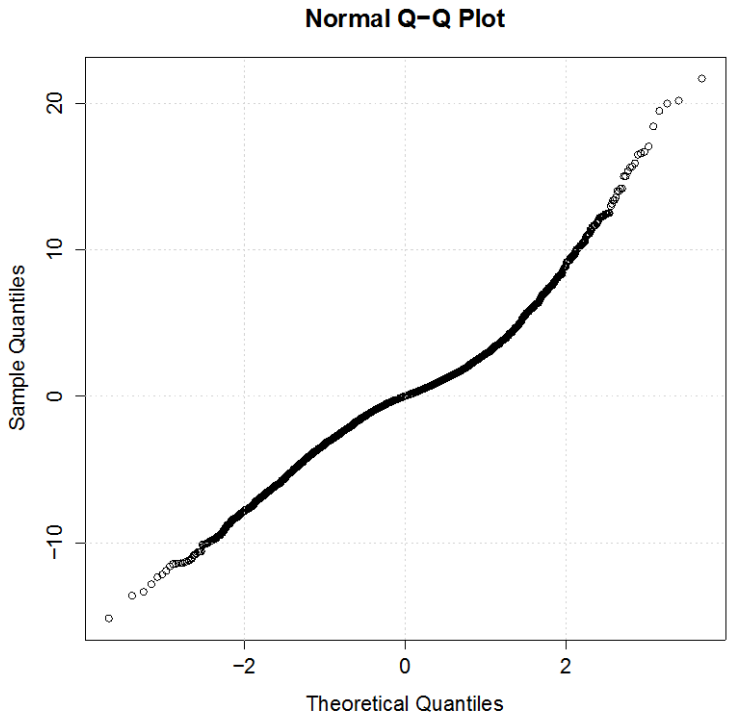

Figure 6–5 Q-Q normal plot for residuals of linear model for measured cooling energy end use during occupied periods

The available remedies for the nonuniform residuals in this case include (1) the addition of additional regressor variables; (2) transformation of the dependent variable; e.g., by means of logarithmic transformation; and (3) the use of weighted least squares instead of OLS. The best option may be to measure additional variables that are causally related to cooling energy end use, as long as they can be easily measured. However, these additional measurements are currently not available.

Next, the control limits for this mode and end use are modeled using quantile regression for the 40^{th}, 50^{th} (median), and 60^{th} quantiles and the computed 40^{th} and 60^{th} quantiles. The mean predictions are plotted against measured cooling energy consumption in Figure 6–6. Predicted control limits and the mean prediction increase roughly linearly with the measured cooling. Also, a meter resolution of 2 kW is evident in the measured data.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

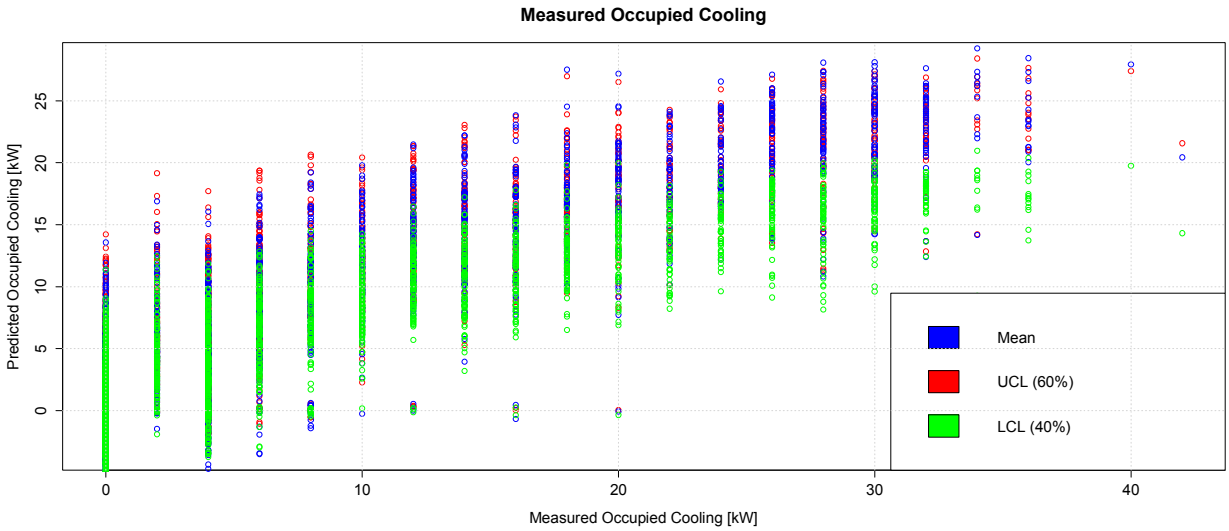

Figure 6–6 Scatter plot of the quantile (40th and 60th) and linear (mean) regression predictions for measured cooling energy end use during occupied periods

Next, for the occupied periods, a time series plot of the quantile predictions and the measured occupied data is provided in Figure 6–7. In general, control limits increase during the summer and decrease during the shoulder periods and winter; the measured cooling consumption follows that trend.

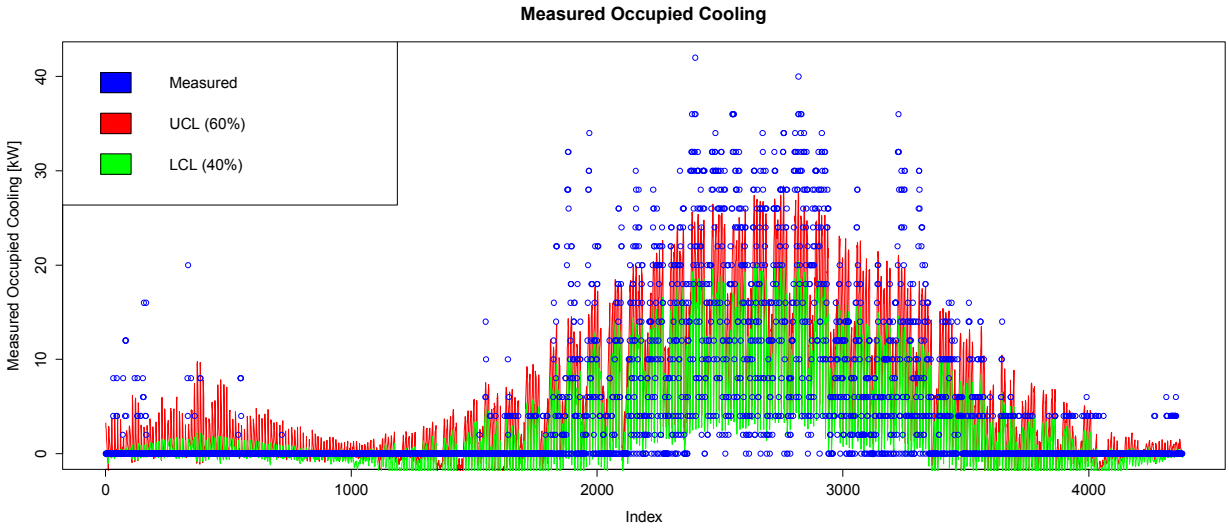

Figure 6–7 Time series plot of the quantiles (40th and 60th) and measured cooling energy end use during occupied periods

Finally, the median quantile prediction is compared against the mean prediction, along with the mean prediction against the measured occupied cooling data in Figure 6–8. The slope of the predicted median is slightly below the predicted mean, confirming that the mean is biased positively because of high-valued outliers of cooling energy end use. Moreover, nonphysical negative values of cooling energy can be observed. The mean prediction is below the measured values above approximately 20 kW.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

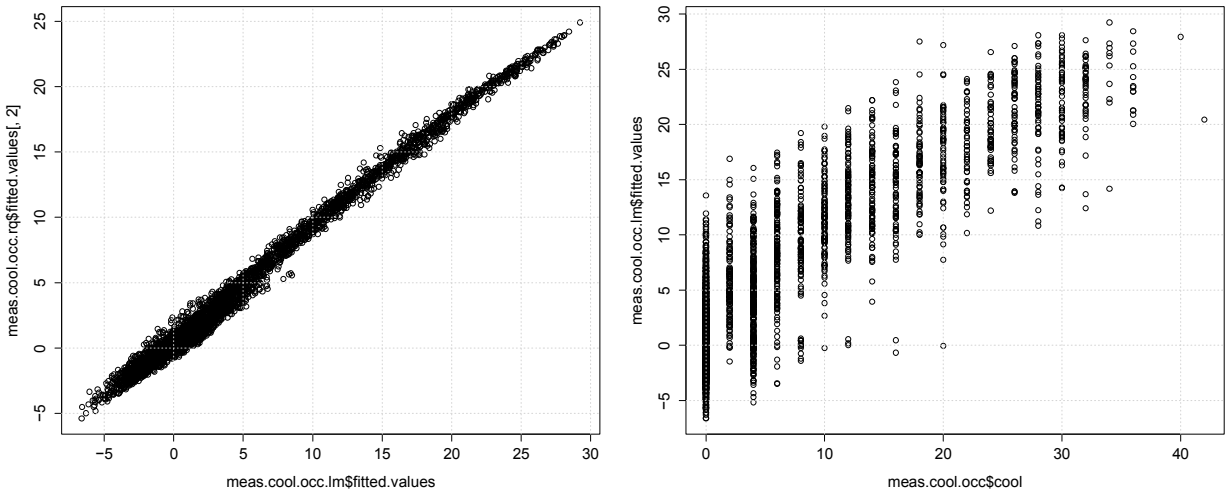

Figure 6–8 Scatter plot of the median versus mean regression predictions (left) and of the mean regression prediction versus the measured cooling data (right) for measured cooling energy end use during occupied periods

6.4.2 Unoccupied Periods

The analysis for the measured unoccupied cooling energy use is nearly identical to that outlined in Section 6.4.1, but considers only those measured data points deemed "unoccupied" (hours 00:00 to 5:00 and 18:00 to 23:00). First, PCA is applied to the data and the four most significant PCs are selected (see Figure 6–9 and Figure 6–10.

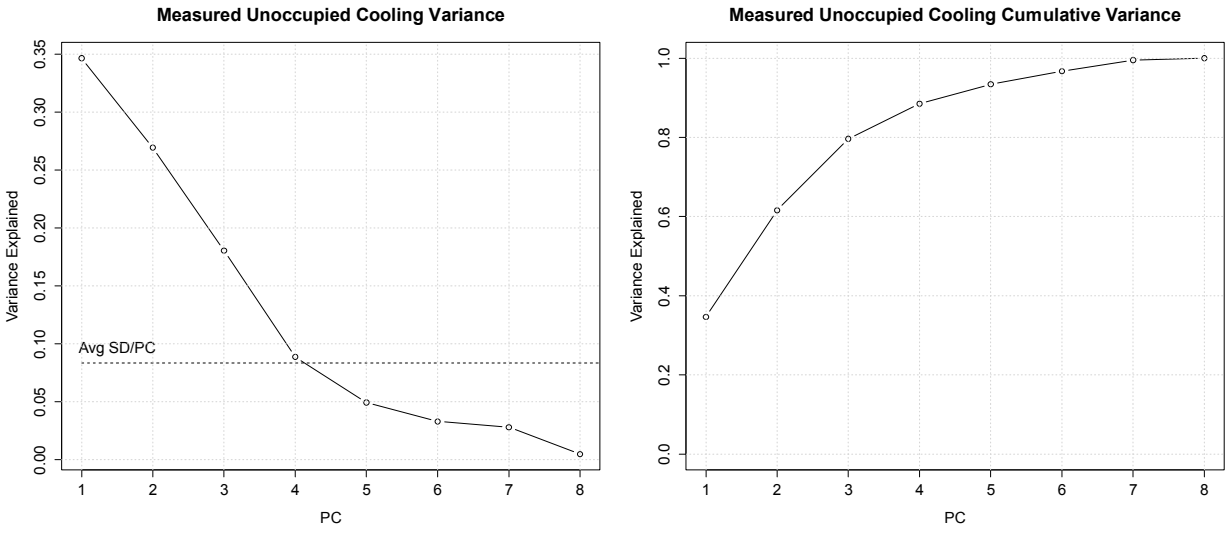

Figure 6–9 Individual and cumulative variance explained by PCs for measured cooling energy end use during unoccupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

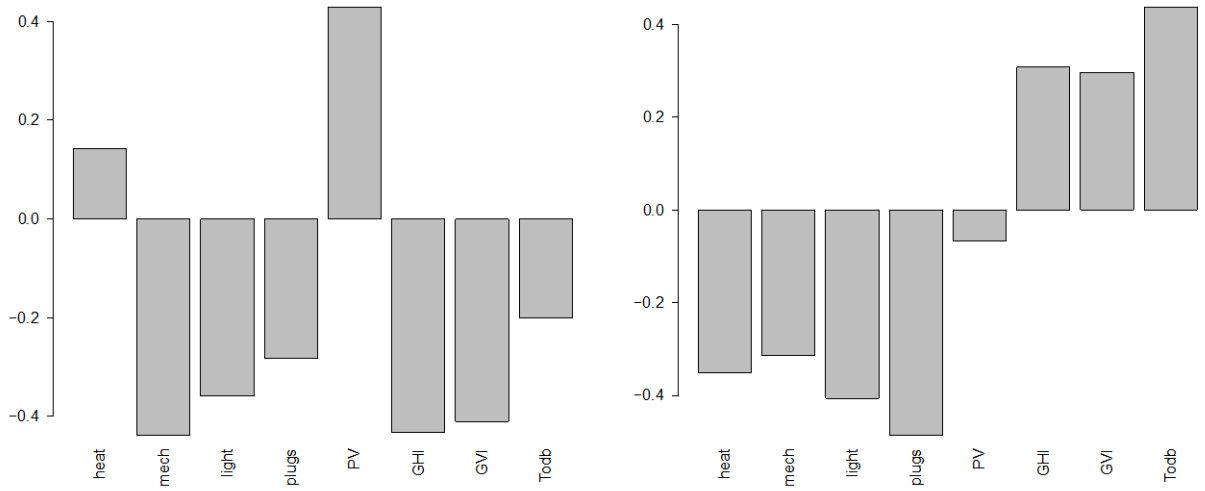

Figure 6–10 Loading factors for the two dominant PCs for measured cooling energy end use during unoccupied periods

Again, the dominant PC is driven most directly by the mechanical end use, horizontal and vertical insolation, followed by light and plug loads, and then ambient temperature. The significant four PCs are used to determine a parsimonious model for measured unoccupied cooling energy use.

The model contains as regressors all four cycle variables, most of the first-order interaction terms between PC1, PC3, and PC4 and the cycle variables, as well as all first-order interaction terms between the four cycle variables. Each of the 22 model variables is significantly different from zero, as confirmed by high Student-*t* scores and associated low Type I errors. Moreover, the overall model exhibits a high *F*-statistic of 429, which indicates that it is significantly different from a random model, and the R^2 statistic of 72% reveals that 72% of the variance in the data can be explained by the linear model. In summary, the linear model for the unoccupied measured cooling periods is slightly inferior to the linear model for the occupied measured cooling periods.

The Q-Q plot in Figure 6–11 shows that the deviations of the model residuals from the assumed normal deviations are again fairly significant, indicating that the residuals exhibit nonuniformity. The DW test revealed a score of 0.93, suggesting that the model is underfitting with positively correlated residuals.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

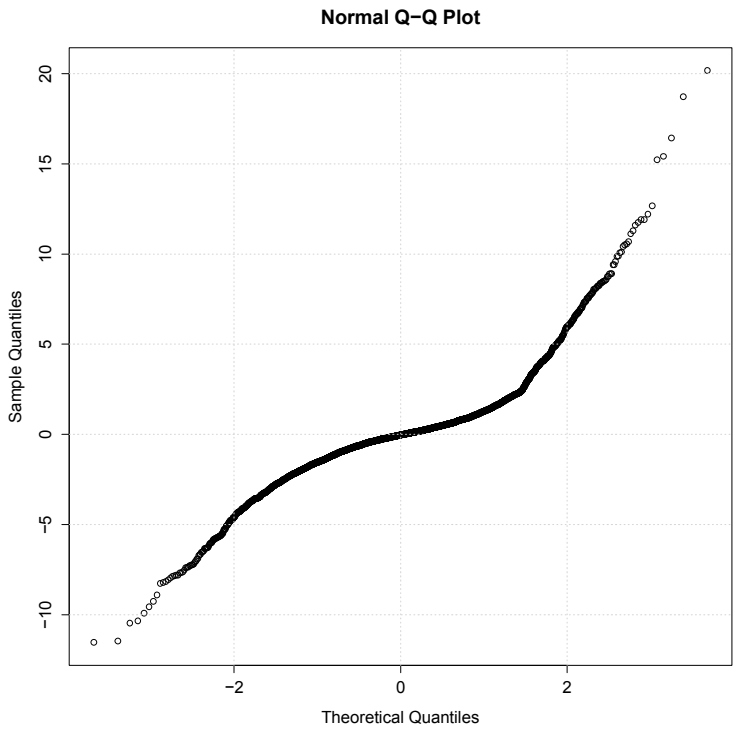

Figure 6–11 Q-Q normal plot for residuals of linear model for measured cooling energy end use during unoccupied periods

Next, the control limits for this mode and end use are modeled using quantile regression for the 40th, 50th (median), and 60th quantiles and the computed 40th and 60th quantiles along with the mean predictions are plotted against measured cooling energy consumption (Figure 6–12).

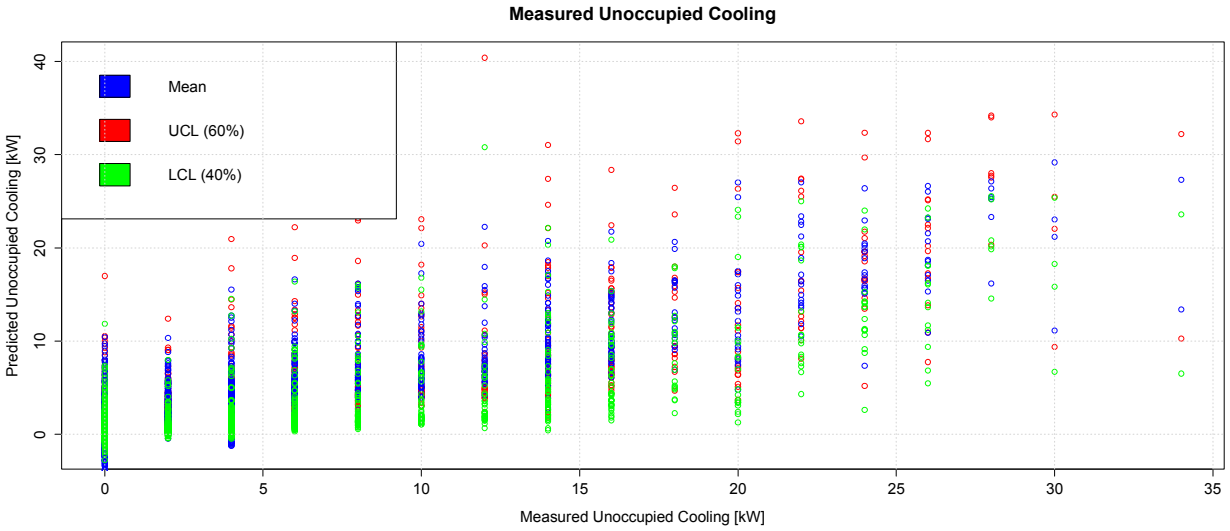

Figure 6–12 Scatter plot of the quantile (40th and 60th) and linear (mean) regression predictions for measured cooling energy end use during unoccupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Next, for the unoccupied periods, a time series plot of the quantile predictions and the measured occupied data is provided. In general, control limits increase during the summer and decrease during the shoulder periods and winter, with the measured cooling consumption following that trend (Figure 6–13).

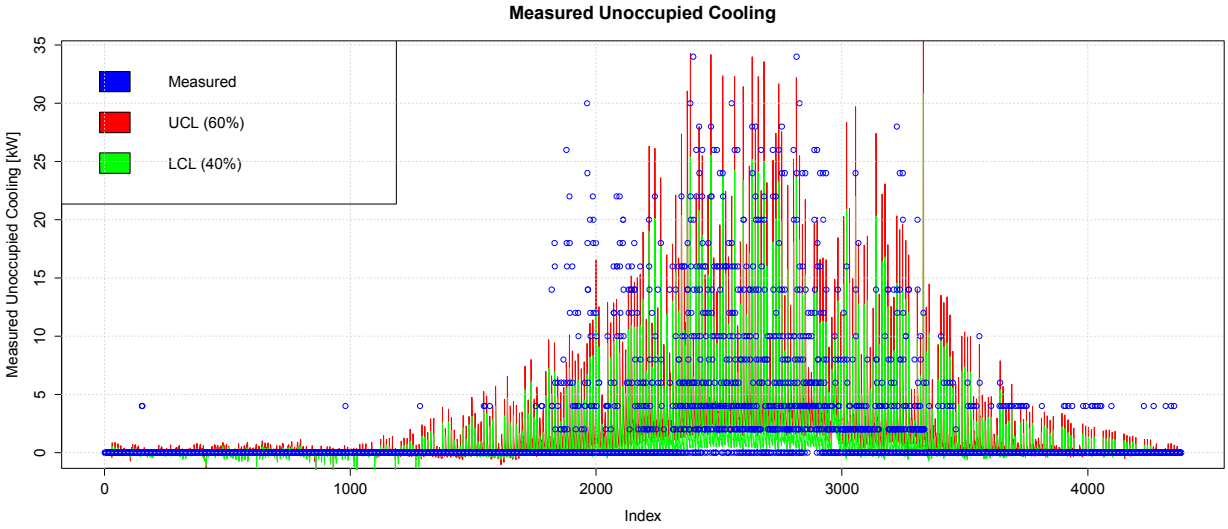

Figure 6–13 Time series plot of the quantiles (40th and 60th) and measured cooling energy end use during unoccupied periods

Finally, the median quantile prediction is compared against the mean prediction, along with the mean prediction against the measured unoccupied cooling data in Figure 6–14. The median versus mean reveals an interesting curvature, almost quadratic behavior. The median falls below the mean up to about 20 kW, and recovers with a steeper slope from 20 kW to the maximum observed values. Moreover, the mean prediction is uniformly below the measured values for the entire range of observed data.

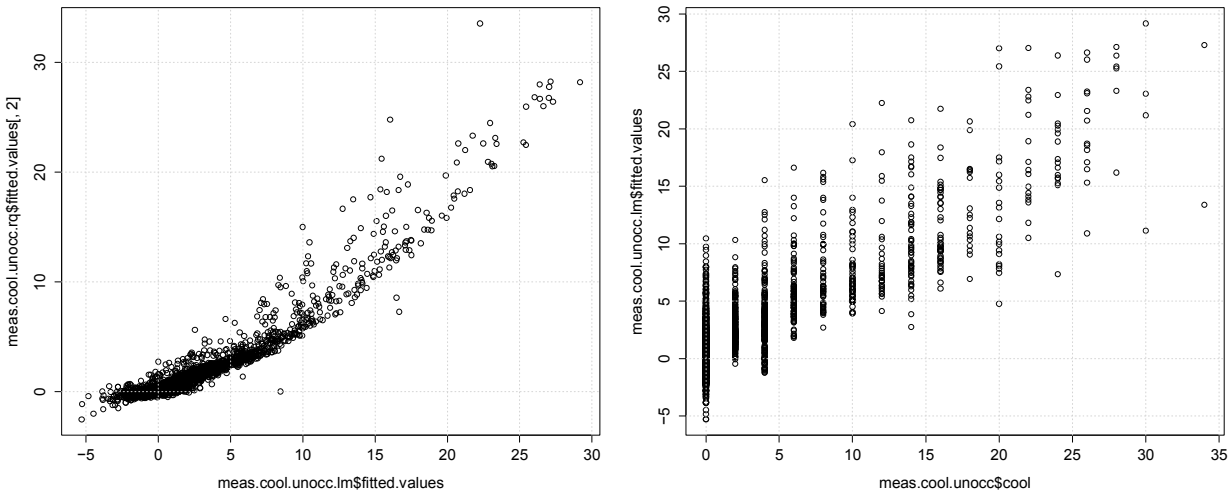

Figure 6–14 Scatter plot of the median versus mean regression predictions (left) and of the mean regression prediction versus the measured cooling data (right) for measured cooling energy end use during unoccupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Now that occupied and unoccupied periods have been separately analyzed and control limits developed, the LCLs and UCLs can be plotted against the measured cooling energy consumption for the entire year (see Figure 6–15). With 8,760 hours available, the 40th and 60th quantiles follow the seasons as intuition would dictate, rising from low values in the spring to peak values around mid-August.

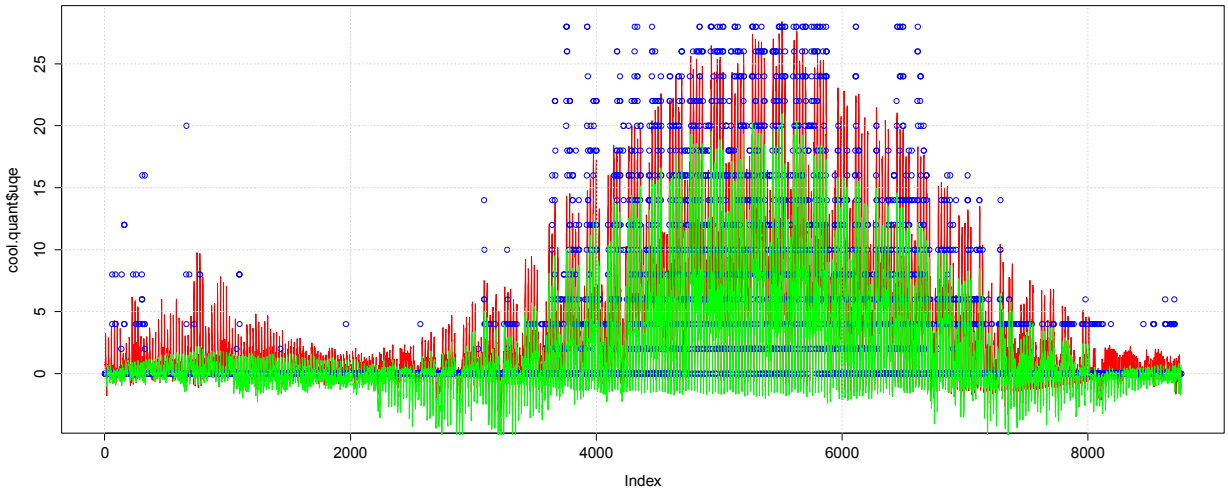

Figure 6–15 Annual development of LCLs (green) and UCLs (red) based on the measured cooling energy consumption data, along with the actual measured values (blue)

6.5 Modeled Cooling Data

6.5.1 Occupied Periods

The analysis procedure between modeled and measured data is identical; the primary differences are the data source and the number of variables included. First the multicollinearity is removed from the data through PCA on the scaled modeled data. Then, significant PCs are selected as those that contribute more than the average share of variance description (see Figure 6–16).

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

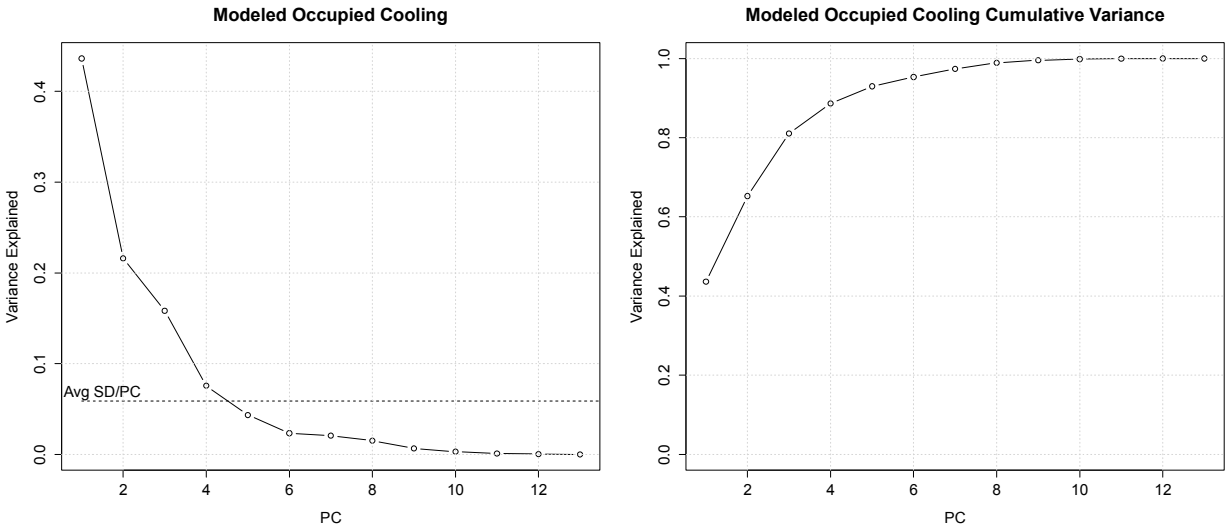

Figure 6–16 Individual and cumulative variance explained by PCs for modeled cooling energy end use during occupied periods

Significantly more information is available in the case of the modeled data because a complete weather archive is used, in particular relative humidity, ambient wet-bulb temperature, additional solar variables such as diffuse and beam horizontal, as well as direct normal components. In Figure 6–17, ambient dry-bulb temperature Todb is in (°F) and Todb.1 is in (°C); the PCA finds identical loadings for both. In the dominant PC, high loadings are associated with dry-bulb and wet-bulb temperatures (or relative humidity with the opposite sign), then with global horizontal insolation. As expected, the second dominant PC exhibits loadings that emphasize all other variables not previously emphasized in the first PC, in particular lights, plug loads, and mechanical systems.

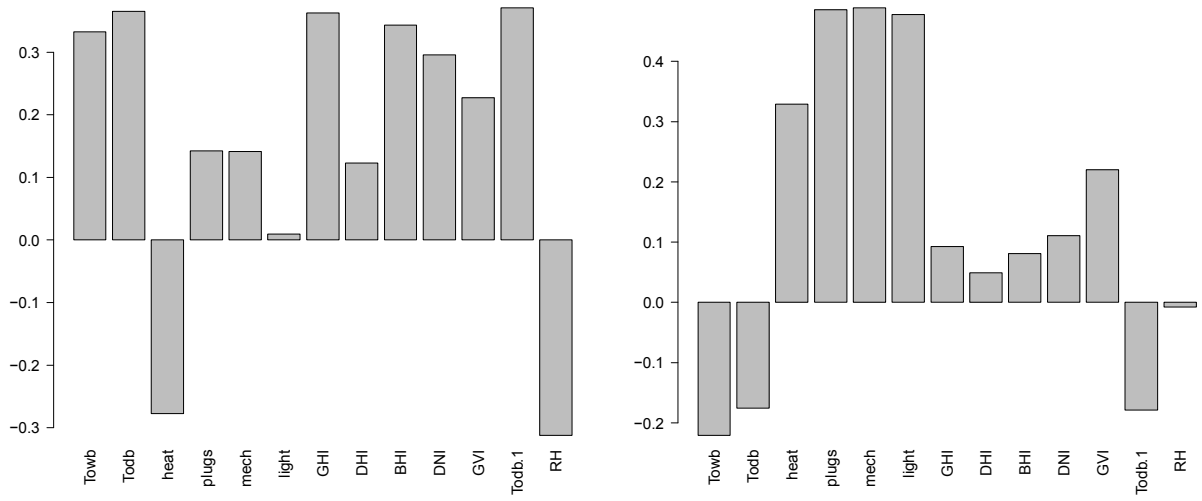

Figure 6–17 Loading factors for the two dominant PCs for modeled cooling energy end use during occupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Figure 6–18 shows a summary of the linear model created by the process. The model contains as regressors the individual PCs PC1–4, all four cycle variables, five first-order interaction terms between the four cycle variables, as well as numerous first-order interaction terms between the PCs and the cycle variables. Except for $C_{h,1}$, each of the 27 model variables is significantly different from zero, as confirmed by high Student-*t* scores and associated low Type I errors. Moreover, the overall model exhibits a very high *F*-statistic of 2612, which indicates that it is significantly different from a random model and the excellent R^2 statistic of 94% reveals that 94% of the variance in the data can be explained by the linear model. Evidently, the linear model built on the simulated data is superior to the model built on the measured data.

```
## lm(formula = mod.cool.occ$cool ~ PC1 + PC2 + PC3 + PC4 + sinH +
##     cosH + sinD + cosD + PC1:PC2 + PC1:PC3 + PC1:PC4 + PC1:sinH +
##     PC1:cosH + PC1:sinD + PC1:cosD + PC2:PC3 + PC2:PC4 + PC2:cosD +
##     PC3:PC4 + PC3:cosH + PC3:sinD + PC3:cosD + sinH:sinD + sinH:cosD +
##     cosH:sinD + cosH:cosD + sinD:cosD, data = mod.cool.occ.rot.exp)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.847 -1.347  0.156  1.319 17.915
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.6574     0.2324   -7.13  1.2e-12 ***
## PC1           0.2993     0.0895    3.35  0.00083 ***
## PC2          -0.5587     0.0406  -13.76  < 2e-16 ***
## PC3          -3.1186     0.1389  -22.45  < 2e-16 ***
## PC4           1.3462     0.0700   19.22  < 2e-16 ***
## sinH         -0.1986     0.0817   -2.43  0.01512 *
## cosH         -5.3617     0.3453  -15.53  < 2e-16 ***
## sinD         -1.8512     0.1574  -11.76  < 2e-16 ***
## cosD         -3.1749     0.2396  -13.25  < 2e-16 ***
## PC1:PC2      -0.1319     0.0220   -5.99  2.2e-09 ***
## PC1:PC3      -0.6057     0.0303  -20.00  < 2e-16 ***
## PC1:PC4       0.4884     0.0251   19.48  < 2e-16 ***
## PC1:sinH     -0.4815     0.0409  -11.78  < 2e-16 ***
## PC1:cosH     -2.5951     0.1063  -24.41  < 2e-16 ***
## PC1:sinD     -0.9577     0.0485  -19.75  < 2e-16 ***
## PC1:cosD     -1.8728     0.0472  -39.68  < 2e-16 ***
## PC2:PC3      -0.2778     0.0291   -9.55  < 2e-16 ***
## PC2:PC4      -0.2069     0.0277   -7.47  9.9e-14 ***
## PC2:cosD      0.6261     0.0618   10.14  < 2e-16 ***
## PC3:PC4      -0.2193     0.0399   -5.50  4.0e-08 ***
## PC3:cosH     -1.0357     0.1920   -5.39  7.3e-08 ***
## PC3:sinD      0.8260     0.0482   17.13  < 2e-16 ***
## PC3:cosD      2.1600     0.0670   32.23  < 2e-16 ***
## sinH:sinD    -0.6706     0.0962   -6.97  3.6e-12 ***
## sinH:cosD    -1.3185     0.1170  -11.27  < 2e-16 ***
## cosH:sinD    -1.2535     0.2313   -5.42  6.3e-08 ***
## cosH:cosD    -2.3073     0.3193   -7.23  5.8e-13 ***
## sinD:cosD     0.6197     0.1517    4.08  4.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.38 on 4352 degrees of freedom
## Multiple R-squared: 0.942,Adjusted R-squared: 0.942
## F-statistic: 2.61e+03 on 27 and 4352 DF,  p-value: <2e-16
```

Figure 6–18 Linear model structure summary report for modeled cooling energy end use during occupied periods

Residual analysis is conducted using a Q-Q plot, which is a plot of the quantiles of two distributions against each other. The pattern of points in the plot is used to compare the two distributions. Linear regression theory dictates that the residuals should be normally distributed with zero mean. Deviations from this assumption are revealed in a Q-Q plot by the deviations of the actual residual quantile distribution from the assumed normal distribution.

The Q-Q plot shows significant deviations of the modeled residuals for high values as shown in Figure 6–19, indicating that the residuals exhibit nonuniformity. The DW tested for over- or underfitting and revealed a score of 0.55, which is lower than the critical score of 2.0. The

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

model is therefore underfitting with positively correlated residuals. The addition of other available regressor variables would likely improve the model quality, similar to the case of the measured cooling energy consumption.

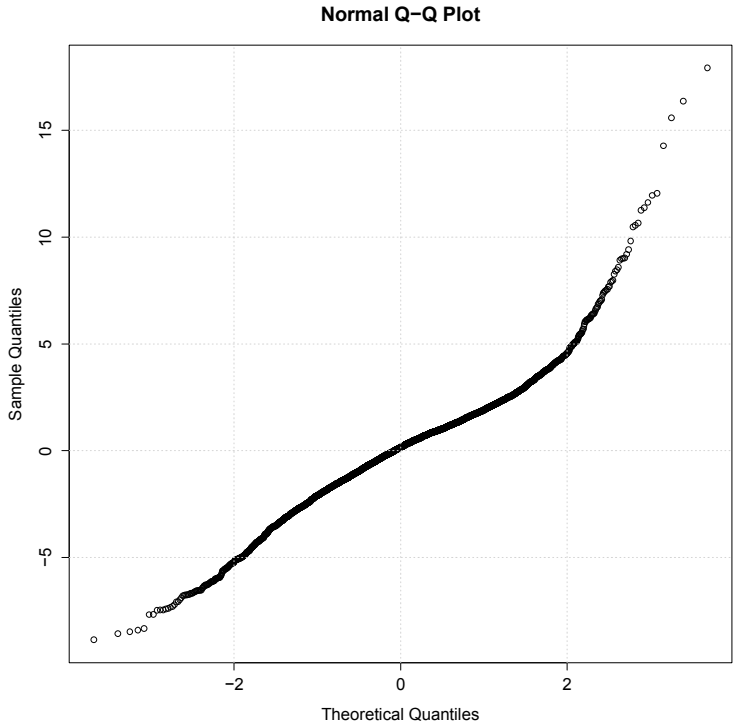

Figure 6–19 Q-Q normal plot for residuals of linear model for modeled cooling energy end use during occupied periods

Next, the control limits for this mode and end use are modeled using quantile regression for the 40^{th}, 50^{th} (median), and 60^{th} quantiles and the computed 40^{th} and 60^{th} quantiles along with the mean predictions are plotted against modeled cooling energy consumption in Figure 6–20. Predicted control limits and the mean prediction increase almost linearly with the measured cooling. Compared to the measured cooling data case, much less dispersion is observed in the simulation data.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

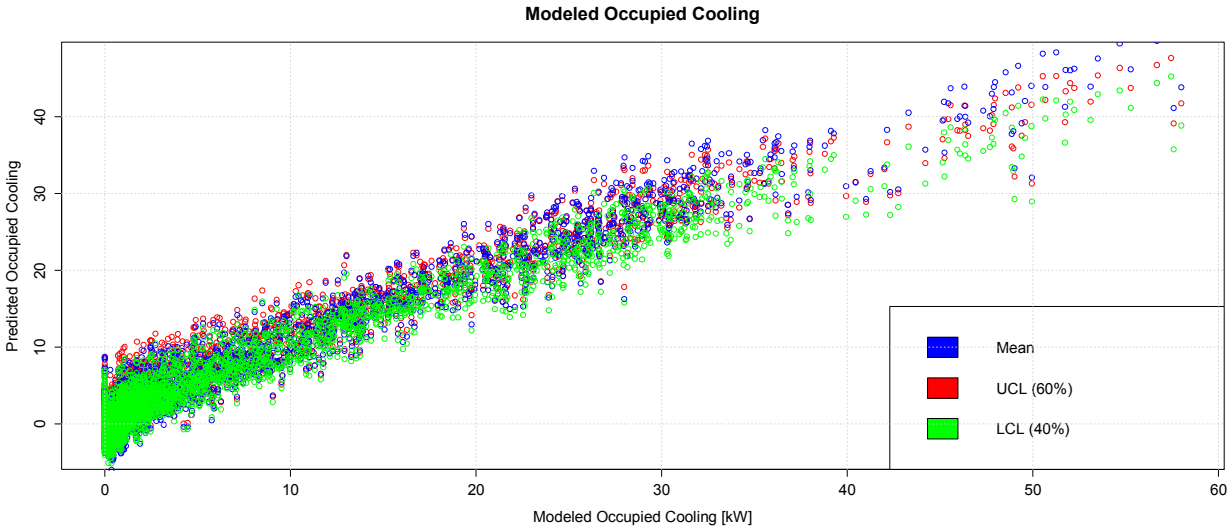

Figure 6–20 Scatter plot of the quantile (40th and 60th) and linear (mean) regression predictions for modeled cooling energy end use during occupied periods

Next, for the occupied periods, a time series plot of the quantile predictions and the measured occupied data is provided in Figure 6–21. In general, control limits increase during the summer and decrease during the shoulder periods and winter; the measured cooling consumption follows that trend. Significantly less dispersion is evident in the time series representation of the simulated performance.

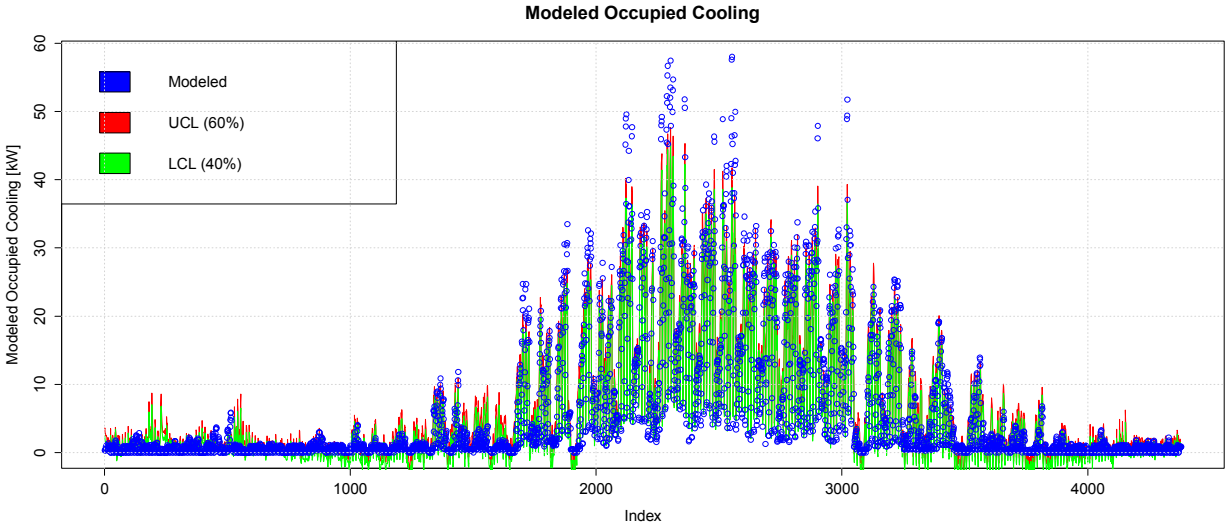

Figure 6–21 Time series plot of the quantiles (40th and 60th) and modeled cooling energy end use during occupied periods

Finally, the median quantile prediction is compared against the mean prediction, along with the mean prediction against the measured occupied cooling data in Figure 6–22. The slope of the predicted median is slightly below the predicted mean, confirming that the mean is biased positively because of high-valued outliers of cooling energy end use. As before, nonphysical negative values of cooling energy can be observed. Finally, the mean prediction is only slightly lower than the measured values.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

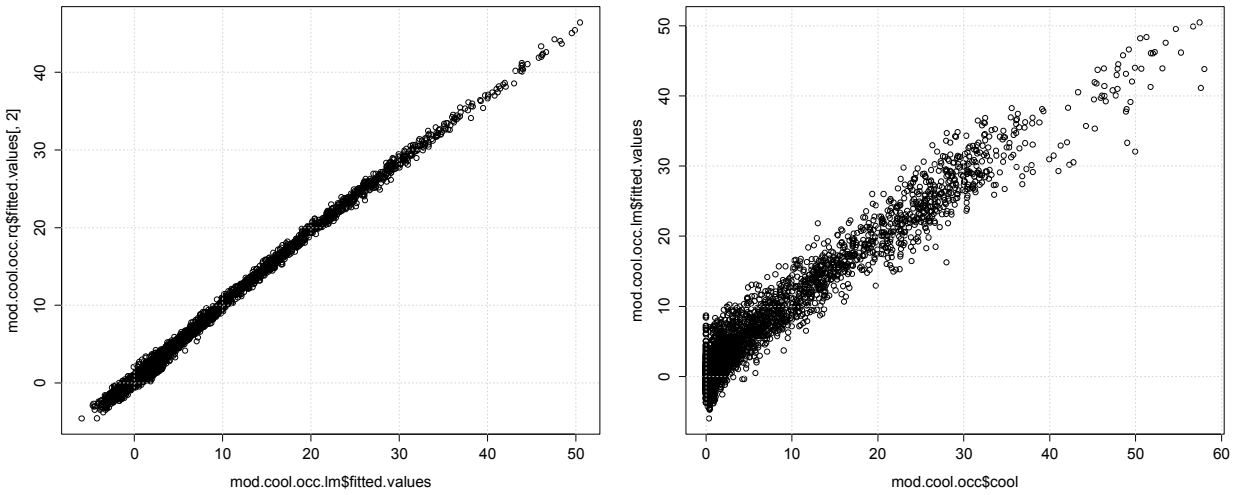

Figure 6–22 Scatter plot of the median versus mean regression predictions (left) and of the mean regression prediction versus the modeled cooling data (right) for modeled cooling energy end use during occupied periods

6.5.2 Unoccupied Periods

Analysis for the modeled unoccupied cooling use is nearly identical to that outlined in Section 6.5.1, but considers only those modeled data points deemed unoccupied (hours 00:00 to 5:00 and 18:00 to 23:00 and weekends). First, PCA is applied to the data. Significant PCs are selected, as shown in Figure 6–23.

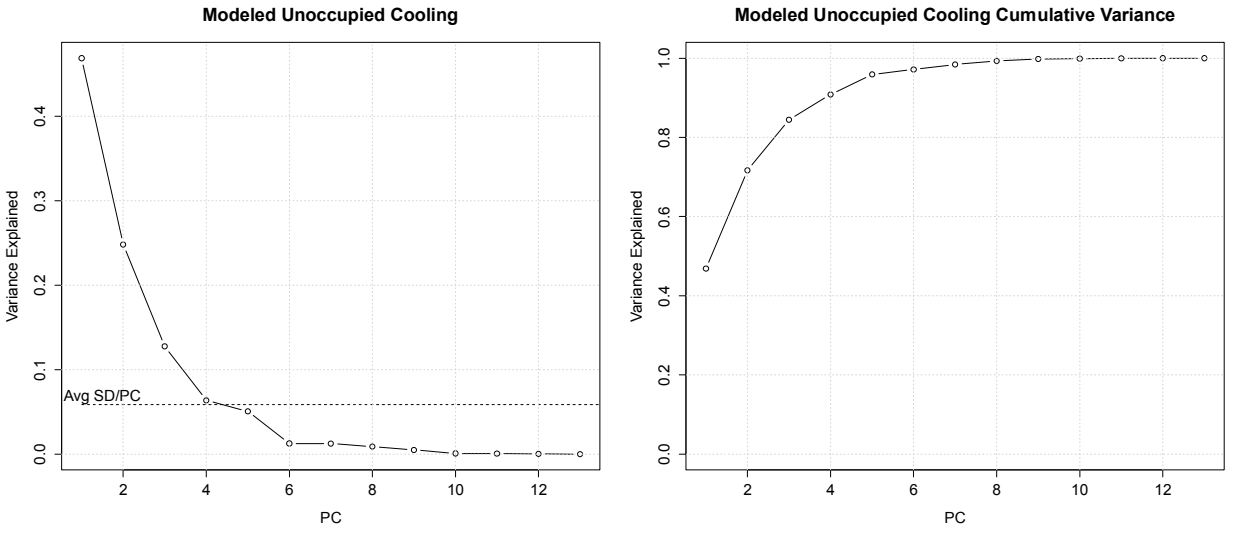

Figure 6–23 Individual and cumulative variances explained by PCs for modeled cooling energy end use during unoccupied periods

In Figure 6–24, ambient dry-bulb temperature Todb is in (°F) and Todb.1 in (°C); the PCA finds identical loadings for both. In the dominant PC, high loadings are associated with dry-bulb and wet-bulb temperatures (or relative humidity with the opposite sign), then global horizontal

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

insolation, followed by beam horizontal insolation. As before, the second dominant PC exhibits loadings that emphasize all other variables not previously emphasized in the first PC, in particular lights, plug loads, and mechanical systems.

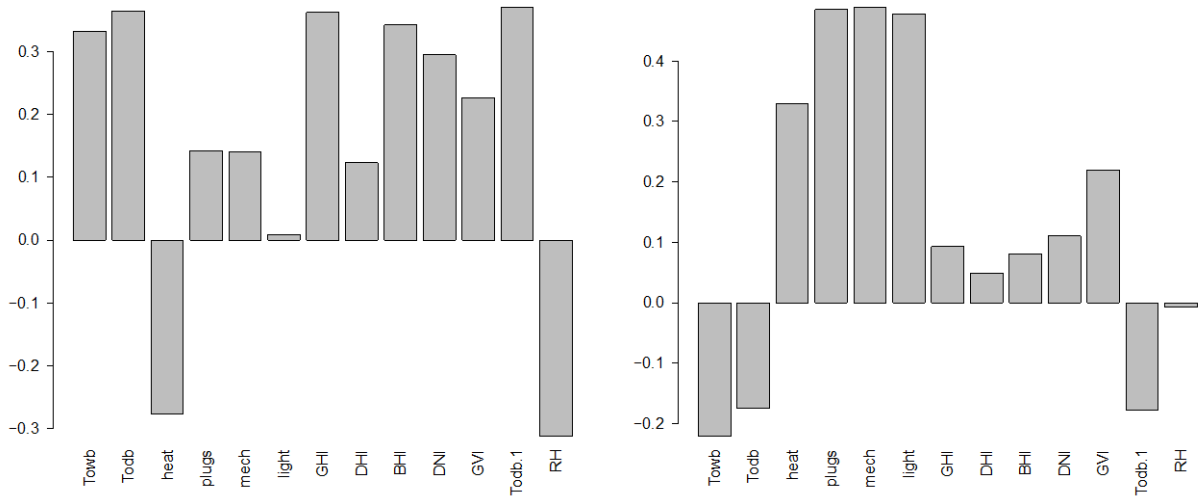

Figure 6–24 Loading factors for the two dominant PCs for modeled cooling energy end use during unoccupied periods

The model contains as regressors the individual PCs PC1–4, all four cycle variables, three first-order interaction terms between the four cycle variables, as well as a large number of first-order interaction terms between the PCs and the cycle variables. Except for $C_{d,2}$, each of the 27 model variables is significantly different from zero, as confirmed by high Student-*t* scores and associated low Type I errors. Moreover, the overall model exhibits a very high *F*-statistic of 3020, which indicates that it is significantly different from a random model and the R^2 statistic of 95% reveals that 95% of the variances in the data can be explained by the linear model. In summary, the linear model for the unoccupied model cooling periods is slightly superior to the linear model for the occupied modeled cooling periods.

The Q-Q plot in Figure 6–25 shows similar deviations from the theoretical normal distribution as the modeled occupied case and the DW score of 0.42 confirms that the model is underfitting with positively correlated residuals.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

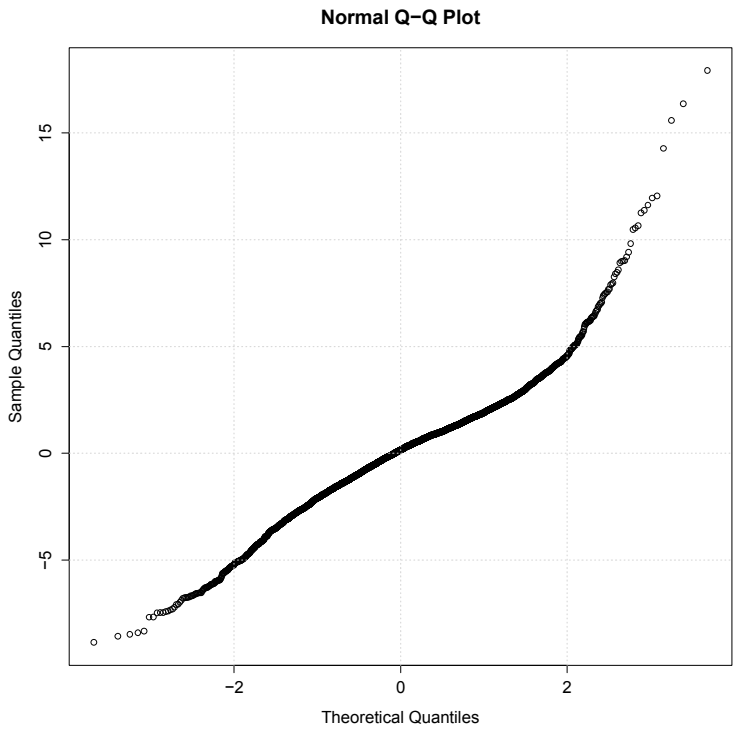

Figure 6–25 Q-Q normal plot for residuals of linear model for modeled cooling energy end use during unoccupied periods

Next, the control limits for this mode and end use are modeled using quantile regression for the 40th, 50th (median), and 60th quantiles and the computed 40th and 60th quantiles; the mean predictions are plotted against modeled cooling energy consumption in Figure 6–26. Predicted control limits and the mean prediction increase almost linearly with the measured cooling. Compared to the measured cooling data case, much less dispersion occurs in the simulation data.

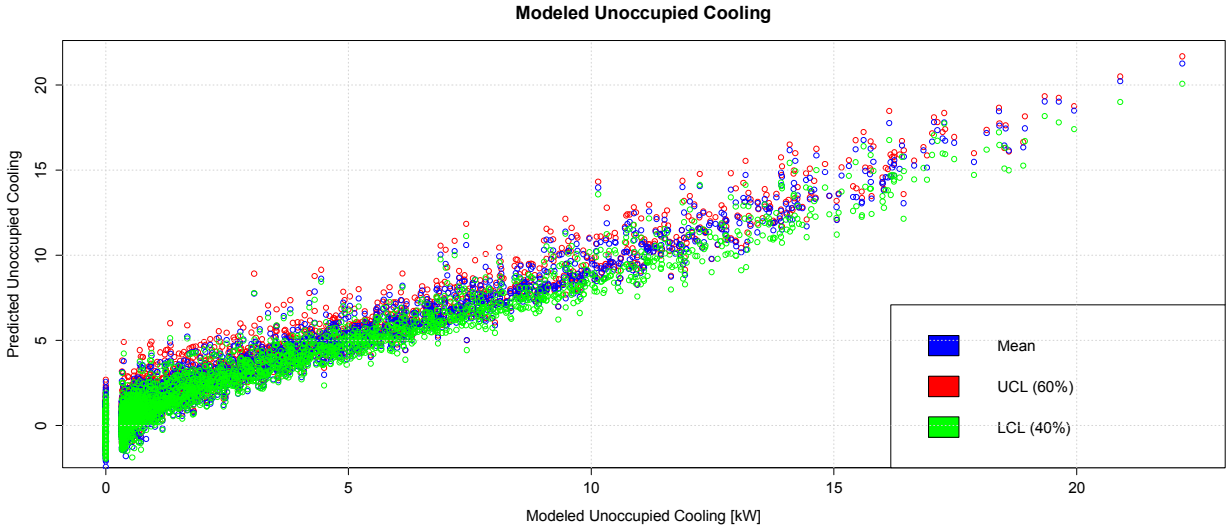

Figure 6–26 Scatter plot of the quantile (40th and 60th) and linear (mean) regression predictions for modeled cooling energy end use during unoccupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

For the unoccupied periods, a time series plot of the quantile predictions and the measured occupied data is prepared (Figure 6–27). In general, control limits increase during the summer and decrease during the shoulder periods and winter, with the measured cooling consumption following that trend. Significantly less dispersion is evident in the time series representation of the simulated performance.

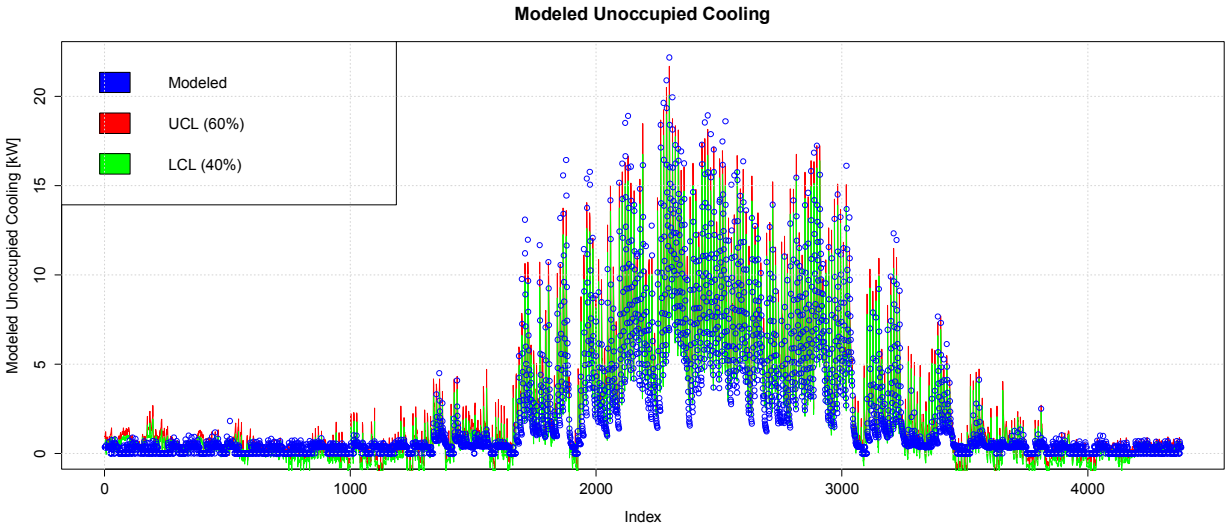

Figure 6–27 Time series plot of the quantiles (40th and 60th) and modeled cooling energy end use during unoccupied periods

Finally, the median quantile prediction is compared against the mean prediction, along with the mean prediction against the measured occupied cooling data in Figure 6–28. The predicted median follows the predicted mean proportionately, revealing that the mean is not biased for the modeled unoccupied data. Similarly, the mean prediction is proportional to the measured values. As before, nonphysical negative values of cooling energy can be observed.

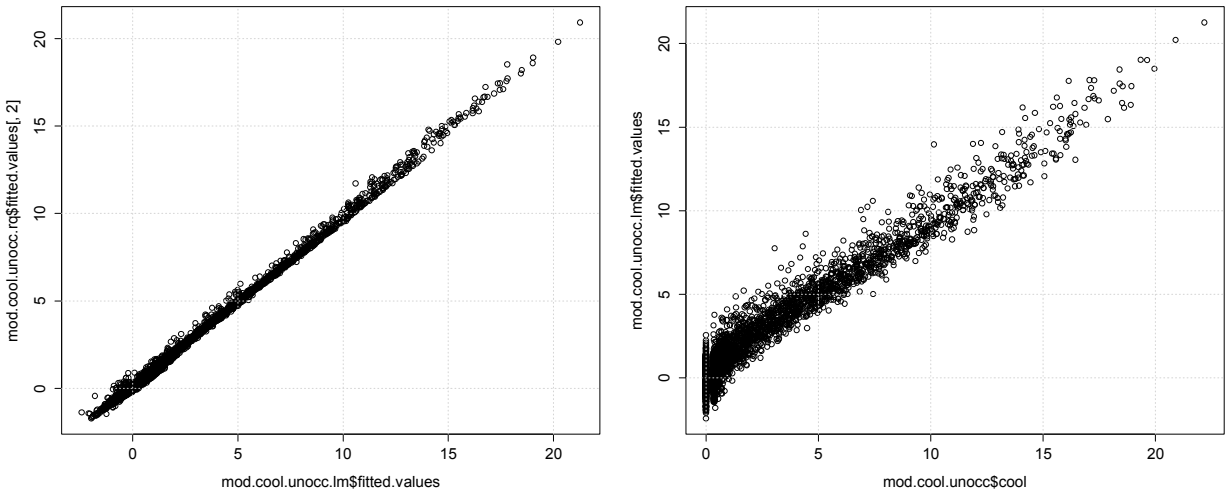

Figure 6–28 Scatter plot of the median versus mean regression predictions (left) and of the mean regression prediction versus the modeled cooling data (right) for modeled cooling energy end use during unoccupied periods

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Now that occupied and unoccupied periods have been separately analyzed and control limits developed, the LCLs and UCLs can be plotted against the measured cooling energy consumption (Figure 6–29). With 8,760 hours available, the 40th and 60th quantiles follow the seasons as intuition would dictate, rising from low values in the spring to peak values around early August.

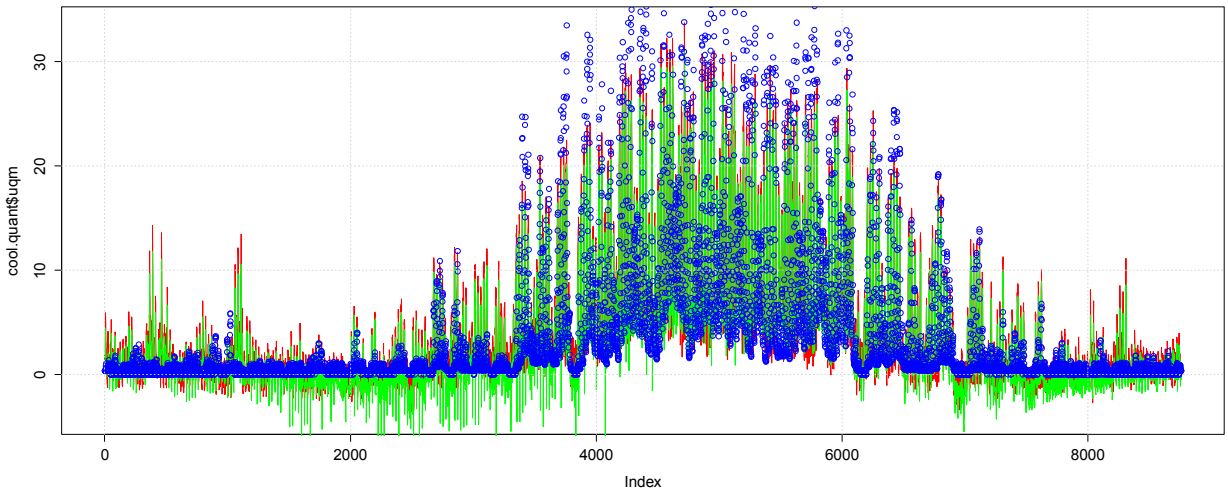

Figure 6–29 Annual development of LCLs (green) and UCLs (red) based on the modeled cooling energy consumption data, along with the actual measured values (blue)

6.6 Data Logic Application

With measured and modeled data analyzed to produce UCLs and LCLs throughout the year, the last step in determining the overall control limits suggested by this analysis is to apply the energy-conservative data logic described in Section 5.4.1. Note that all nonpositive fits are removed, as these values are physically meaningless and will not affect the integrity of the models when compared to the actual cooling data.

Now that the energy-conservative data logic has been applied to the control limits based on measured and modeled data, the final LCLs and UCLs can be plotted against the measured cooling energy consumption (Figure 6–30.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

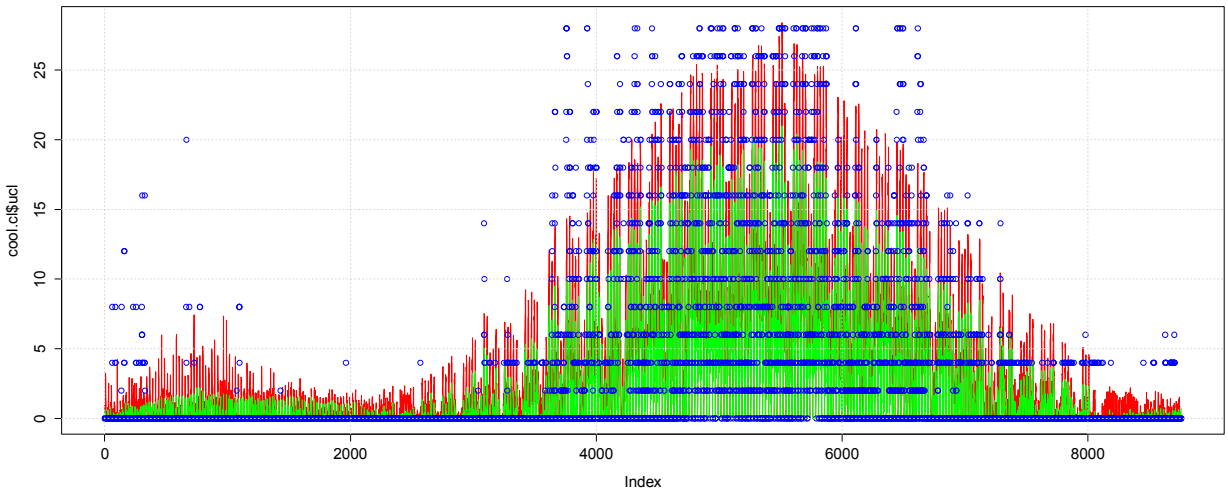

Figure 6–30 Annual development of LCLs (green) and UCLs (red) based on energy-conservative data logic along with the measured values (blue)

Time series plots are provided for the month of August in Figure 6–31 and for the second week of August in Figure 6–32. The 60th percentile is consistently predicted above the 40th percentile and the measured cooling data are fairly close to the relatively tight control limits.

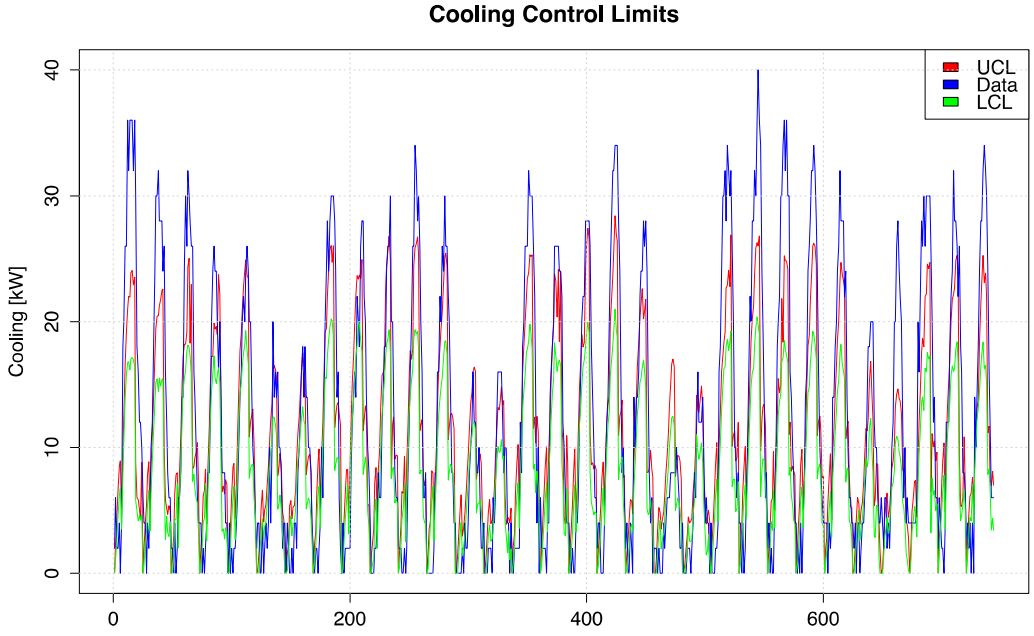

Figure 6–31 Summer month cooling UCLs (red) and LCLs (green) along with measured cooling end use

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

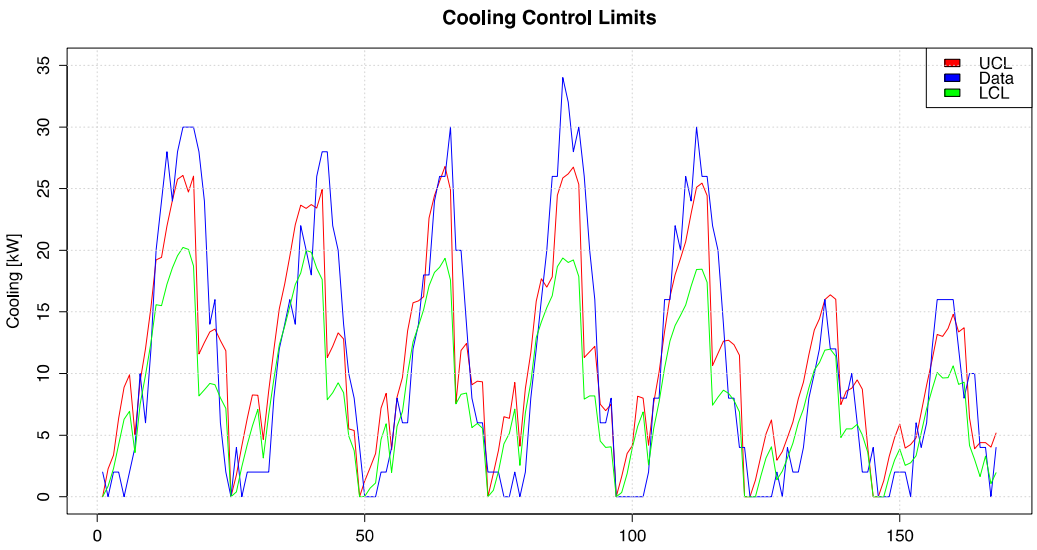

Figure 6–32 Summer week cooling UCLs (red) and LCLs (green) along with measured cooling end use

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

7.0 Summary and Next Steps

7.1 Summary

The report presented three different but related approaches to determine performance bounds on building energy performance for display on an energy information system: (1) an approach based on engineering judgment-informed curve fitting; (2) an approach based on the frequency of occurrences; and (3) a quantile regression-based approach. The process was illustrated using electricity consumption for cooling provided by chilled water. Yet, the same approach can be used to analyze other energy end uses such as heating, lighting, plug loads, mechanical systems (fans and pumps), and even PV power generation.

In the first approach, using engineering judgment and manual curve fitting, end-use control limits were derived using static lookup-based performance targets, combined with polynomial curve-fit models. These control limits were developed based on a combination of historical end-use data and as-built energy modeling, with single main effect variables selected for each end use, such as solar irradiance for lighting performance.

In the second frequency-based approach, measured building end-use energy data were used to develop empirical frequency distributions for each end use as a function of the dominant independent variable determined from visual inspection and a pairwise correlation analysis. Appropriate control limits were defined as ranges in which a certain percentage of end use consumption has been observed historically (here the central 50% of observations between the 25^{th} and 75^{th} percentiles), as a function of the dominant independent variable. The same process was then applied to simulated energy end use data sourced from a calibrated building energy model.

Effectively, two sets of upper and lower control limits were generated from measured and simulation-based data. An energy-conservative control logic was then applied to select one set of control limits for each situation. This set of LCLs and UCLs for each energy end use was adopted in the Building Agent energy dashboard system for presentation to building operators and occupants.

A more universal methodology that harnesses *all* the available covariates of an energy end use was desired to predict UCLs and LCLs as a function of all the relevant independent variables, not just the dominant one. In essence, a methodology was needed that can be applied to any situations where control limits on energy end use are needed, whether a sophisticated monitoring systems makes dozens of variables available or whether only a few points are monitored.

Therefore, in the third quantile regression-based approach, the building energy end use was modeled linearly in terms of other end uses, exogenous variables, and weather data, all of which were transformed to orthogonal PCs, and the linear models were used to generate control limits using quantile regression for the measured and simulated data. In this report, quantile regression was used to predict the 40^{th} and 60^{th} quantiles. The same energy-conservative control logic was then applied to the two sets of quantile regression-based control limits.

By offering the viewer of the energy dashboard real-time expected range of energy end-use consumption in addition to the current instantaneous value, conclusions can be drawn about the relative energy intensity. If the current value falls within the expected range, the building energy end use can be considered acceptable, not requiring any action from facility operators or

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

occupants. Alternatively, if the energy use exceeds the UCL, the dashboard indicates opportunities for energy savings. The greater the departure, the more significant the savings potential will likely be.

The first approach shows engineering judgment-based control limits; the second and third approaches present to the viewer the more energy-conservative control limits based on the comparison between modeled and observed data. Each method encourages occupants and operators to maintain the building's energy performance in a more energy conservative state.

7.2 Next Steps

Four steps lie ahead:

1. The quantile regression presented in this report will need to be deployed in the RSF dashboard system.
2. Once completed, the dashboard system should be expanded to the ESIF, where metering infrastructure has recently been commissioned.
3. As mentioned throughout the report, adopting control limits representing the 25^{th} and 75^{th} percentile may be too loose, and adopting control limits representing the 40^{th} and 60^{th} percentile may be too stringent for building occupants to respond actively. Finding control limit values that encourage proactive, energy-conscious behavior would involve adopting an initial set of control limits, deploying a survey to the occupants, and then adjusting the control limit definition as the outcome of the survey suggests.
4. Future work is needed to determine the relative performance of physical models deployed in real time rather than the statistical models presented here. For example, beginning with a steady-state system model, the measured PV output exceeded the expected range for many hours, suggesting the temperature during those hours was lower and thus the PV output higher than for the average temperature used to establish the control limits. This observation revealed the limitation associated with selecting only one dominant variable and suggested more advanced approaches that account for all explanatory variables.

The quantile regression approach presented is one valid pathway to account for all explanatory variables. Alternatively, a physical system model may be preferable. For example, a Monte Carlo-based execution of the PVWatts simulation model using available measured data for either a historical analysis or predicted variables and probability distributions for unobserved and uncertain input variables for a predictive analysis could be pursued. A predictive application of a physical model is discussed further.

Once the Monte Carlo sampling terminates upon convergence, the numerous possible daily time series for each energy end use (or PV production) is analyzed by developing empirical cumulative distribution functions for each hour of the day. Relevant quantiles (e.g., 10%, 25%, 40%, 60%, 75%, and 90%) are turned into colored lines (see Figure 7–1). At any time of the next day, the actual energy end use will be plotted as a time series on top of these quantile envelopes. The blue line represents the observed end-use energy consumption, plotted on top of the three confidence envelopes. For the example case illustrated, in the last couple of hours the actual end use has begun to increase above the 75% expected quantile.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

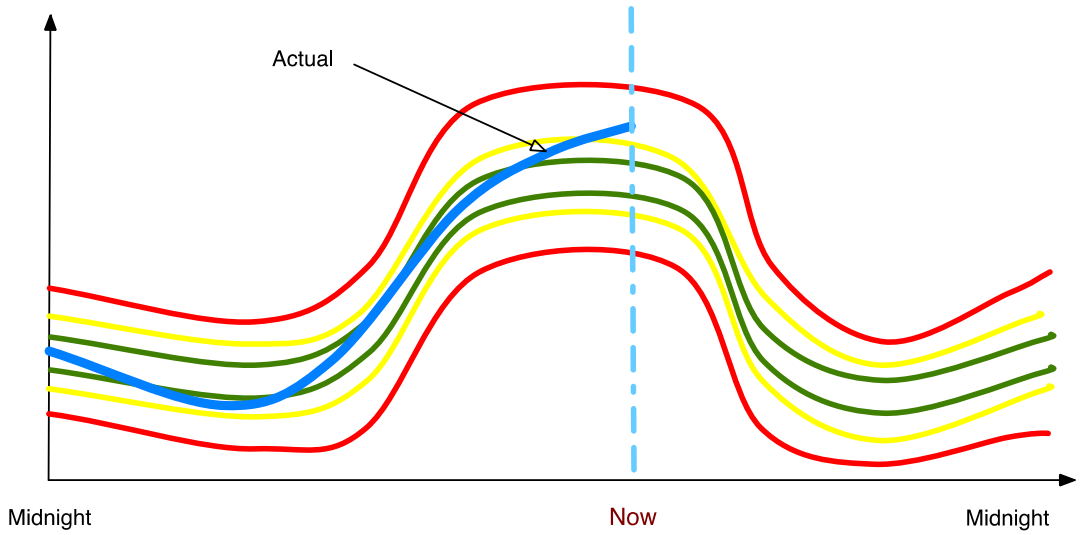

Figure 7–1 Predicted quantile time series derived from numerous physical model runs executed in the context of a Monte Carlo analysis

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

8.0 References

DOE. (2010). 2010 Buildings Energy Data Book. Washington, DC: U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy

Leach, M.; Lobato, C.; Hirsch, A.; Pless, S.; Torcellini, P. (2010). *Technical Support Document: Strategies for 50% Energy Savings in Large Office Buildings*. Golden, CO: National Renewable Energy Laboratory. NRELTP-550-49213. www.nrel.gov/docs/fy10osti/49213.pdf

Lobato, C.; Pless, S.; Sheppy, M.; Torcellini, P. (2011). "Reducing Plug and Process Loads for a Large Scale, Low Energy Office Building: NREL's Research Support Facility." Presented at ASHRAE Winter Conference, 2011. Preprint. Golden, CO: National Renewable Energy Laboratory. NRELCP-5500-49002. www.nrel.gov/docs/fy11osti/49002.pdf

Long, N.; Scheib, J.; Pless, S. (2013). "Using Models to Provide Predicted Ranges for Building-Human Interfaces." In Proceedings of the 13th International Conference of the International Building Performance Simulation Association, 2013. Chambery, France.

Pless, S.; Torcellini, P.A.; Shelton, D. (2011). "Using an Energy Performance Based Design-Build Process to Procure a Large Scale Low-Energy Building." Presented at ASHRAE Winter Conference, 2011. Preprint. Golden, CO: National Renewable Energy Laboratory. NRELCP-5500-51323. www.nrel.gov/docs/fy11osti/51323.pdf

Schott, M.; Long, N.; Scheib, J. (2012). "Progress on Enabling an Interactive Conversation between Commercial Building Occupants and Their Building to Improve Comfort and Energy Efficiency." In Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings. Monterey, CA. www.aceee.org/files/proceedings/2012/data/papers/0193-000234.pdf

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.