Energy and Buildings 87 (2015) 95–104

Contents lists available at ScienceDirect

Energy and Buildings

journal homepage: www.elsevier.com/locate/enbuild

Algorithm for automating the selection of a temperature dependent change point model

Mitchell T. Paulus*, David E. Claridge, Charles Culp

Texas A&M University, United States

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form 7 November 2014
Accepted 8 November 2014
Available online 18 November 2014

Keywords:
Energy modeling
Regression
EnergyPlus
Measurement and verification

ABSTRACT

An algorithm was developed to automate the process of selecting a temperature dependent change point model. Regression models based solely on outdoor air temperature for monitoring and verification purposes are common. The correct change point model shape is determined through a series of three tests. The first test checks whether the coefficients of the model are the correct sign for the shape. The second test checks if the coefficients for the model are significant. The final test checks whether enough data points are present in each temperature region of the model. The algorithm was tested with synthetic EnergyPlus electricity and natural gas data for an outpatient hospital, medium office building, large office building, large hotel, secondary school, and warehouse, with weather data from Chicago, Miami, Seattle, and Fairbanks. The algorithm was able to select the most appropriate temperature dependent change point model for all 48 cases tested. The algorithm can be used in an automated energy modeling routine for monitoring and verification or for checking human decision-making in the energy modeling process.

1. Introduction

Regression modeling is often used to determine savings from a building energy retrofit. The regression model can predict the baseline, or pre-retrofit, energy use based on influential parameters such as outdoor air temperature. The energy consumption for the building during the post-retrofit period can be predicted from the baseline regression model. This baseline regression model also allows an analyst to determine normalized savings under different building operating conditions.

Two widely used guidelines for the measurement and verification (M&V) process are ASHRAE Guideline 14: Measurement of Energy and Demand Savings, and the International Performance Measurement and Verification Protocol (IPMVP) [1,2]. Both of these guidelines detail three major savings determination approaches in addition to deemed savings. The approaches include retrofit isolation, whole building, and calibrated simulation. The IPMVP breaks up the retrofit isolation approach into two options (A & B), and ASHRAE Guideline 14 breaks up the whole building approach into a performance and prescriptive path. The algorithm presented in this paper would be part of the whole-building performance or prescriptive path under ASHRAE Guideline 14 and part of Option C for the IPMVP.

ASHRAE Guideline 14 and the IPMVP are designed to provide a common set of terms and methods helping people involved in energy efficiency projects such as facility energy managers, energy service companies (ESCOs), and consultants. In particular, both documents provide details for inverse models used to determine energy savings. ASHRAE Guideline 14 presents more specific details regarding inverse modeling for the whole building approach than does the IPMVP.

The IPMVP has detailed the basic approach to savings determination with the following steps [2]. The algorithm presented in this paper relates to the baseline model creation, which is part of step 5 and is necessary for computing and reporting savings in step 8 of the IPMVP process.

1. Select the IPMVP Option consistent with the scope of the project (similar to selecting a path from ASHRAE Guideline 14).
2. Gather relevant energy and operating data from the baseline period.
3. Determine the energy savings program.
4. Prepare the measurement plan, and verification plan if necessary.
5. Design, install, and test any special equipment under the M&V plan.

* Corresponding author. Tel.: +1 262 483 9068.
E-mail address: paulusm14@gmail.com (M.T. Paulus).

http://dx.doi.org/10.1016/j.enbuild.2014.11.033
0378-7788/

6. After the energy savings measures are implemented, follow up with a commissioning process.
7. Gather energy and operating data from the post-retrofit period.
8. Compute and report savings in accordance with the M&V plan.

Change point models, also known as piecewise linear regression models, are often used in predicting heating and cooling energy consumption in residential and commercial buildings, and are specifically discussed in the IPMVP and in ASHRAE Guideline 14. The physical basis of the linear change point methods are well known, and with interpretation, certain parameters such as the balance point temperature can be estimated [3–6]. Degree-day approaches, often used for estimating the heating load for residential buildings, could be considered the first use of change point models for estimating energy use [7]. Fels generalized the degree-day approach and set the stage for using change point models measuring energy savings for other energy components and for commercial buildings [8]. However, other authors have warned against the dangers of assuming degree-day models and the simple physical basis behind them can be accurately applied to all buildings [9,10].

Many additional parameters and methods for data-driven modeling have been suggested to improve upon the single-variate linear approaches. Katipamula et al. [11] suggested adding other variables relating to the dew point at the cooling coil, solar load, and a breakup of internal and external loads. Multiple linear regression models have shown great functionality in terms of evaluating initial designs [12–14], the prediction of energy consumption and demand in several different building sectors [15–17], and predicting indoor temperature and relative humidity [18]. Measurement and verification can use any advanced mathematical technique for building energy prediction including Fourier series [19,20], support vector machines [21–28], neural networks [29–32], among others.

Heo and Zavala [33] and Burkhart et al. [34] have argued for Gaussian process modeling in M&V to capture the complexity and non-linearity of building energy consumption and increase the accuracy of the savings uncertainty estimates. This work focused on linear change point models at the monthly time scale because of the innate simplicity and the relationship to ASHRAE Guideline 14. The physical basis for outdoor air temperature based linear change point models is described in [35].

Two important ASHRAE research projects investigated inverse modeling, ASHRAE RP-1050 and RP-1404. ASHRAE RP-1050 produced the Inverse Modeling Toolkit (IMT), which was a Fortran 90 application for developing regression models specifically for building energy use [36]. A user of the Inverse Modeling Toolkit needs to select the type of regression model from experience. ASHRAE RP-1404 presented modeling techniques using a year of monthly utility bill data combined with a shorter span of sub-metered hourly or daily data necessary for generating change point M&V models [37–39]. Reducing the necessary time span of baseline data is more important for daily and hourly modeling where sub-metering is often required. Monthly utility bills are more often available.

A goal of this work is to reduce the amount of necessary experience or to eliminate the human decision-making in selecting a proper change point baseline regression model. Automating the selection of a temperature dependent change point regression model is dependent on the selection of a proper model shape. Some information regarding selection procedures can be found in [40,41]. In this paper, the term model shape, or the model type, refers to the models as shown in Fig. 1. The model shapes considered for this algorithm include 2P, 3P-Heating, 3P-Cooling, 4P, and 5P. The "NP" nomenclature stands for the number of parameters determined by the regression, which are the B's in Fig. 1.

Although 7 models are shown in Fig. 1, computationally there are only 5 different models. The 5 different models are 2P, 3P-Heating, 3P-Cooling, 4P, and 5P. The differences between the models are in the calculation procedures, particularly in how the transformed temperature variables are calculated. In fact, after the temperature transformations, the 2P, 3P-C, and 3P-H models are all simple linear regressions, and the 5P and 4P are multiple linear regressions of the same form. Some consider the 4P model to have a "cooling" and "heating" shape similar to the 3P models, but the differences lie in the signs of the coefficients instead of the calculation procedure for the 3P-C and 3P-H models.

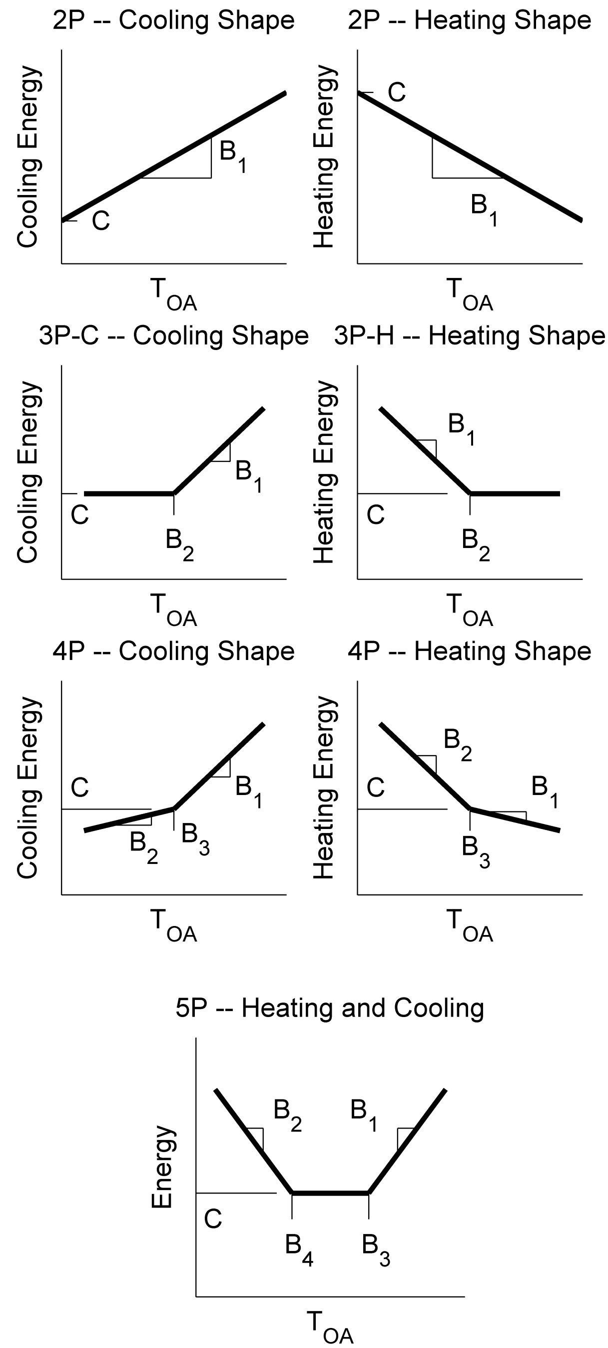

Fig. 1. IMT change point models. Top row: 2P cooling and heating models. Second row from top: 3P cooling and heating models. Third row from top: 4P models in the "cooling" and "heating" shape. Bottom row: 5P heating and cooling model.

Mathematically, the forms of the models are (using the nomenclature from ASHRAE Guideline 14)

$$2\text{P}: \quad E = C + B_1(T) \tag{1}$$

$$\text{3P-C}: \; E = C + B_1(T - B_2)^+ \tag{2}$$

$$\text{3P-H}: \; E = C + B_1(B_2 - T)^+ \tag{3}$$

$$\text{4P}: \; E = C + B_1(T - B_3)^+ + B_2(B_3 - T)^+ \tag{4}$$

$$\text{5P}: \; E = C + B_1(T - B_3)^+ + B_2(B_4 - T)^+ \tag{5}$$

where C is a constant, E is the energy use, T is the outdoor air dry-bulb temperature, and the B's are regression coefficients or temperature change points depending on the model type. The $()^+$ notation indicates when the parenthetic term evaluates to a negative number it is set to 0. As an example, the 5 parameters in the 5P model would be the regression coefficients C, B_1, and B_2, along with the two temperature change point values B_3 and B_4. These parameter values can be found in several ways, referenced in literature [42–44]. A straightforward method to calculate the change point is presented in Section 2.

The only independent variable considered at this time was the average monthly outdoor air temperature. As seen in ASHRAE Guideline 14, univariate models are an accepted baseline modeling approach at the monthly time scale.

2. Theory

2.1. Determining the temperature change point and model coefficients

The coefficients for the models are calculated using standard least squares regression techniques. A transformed temperature variable is used for the 3P, 4P, and 5P models. The transformed temperature variable is calculated by subtracting a change point temperature from the measured temperature and setting negatives values to 0, or doing the opposite, subtracting the measured temperature values from a change point temperature and then setting negative values to 0.

Literature describing algorithms for determining the temperature change points were presented in Section 1. However, with modern computing power, a simpler, "brute force" method is now plausible for personal implementation if desired. A simple method is detailed in this section.

First, a step size for the temperature change point is chosen. A step size of 0.15 °C [0.25 °F] has worked well in tests. The minimum search limit for the temperature change point can be set to rounding up the minimum temperature plus the change point step size, and the maximum search limit can be set by rounding down the maximum temperature minus the change point step size.

Each possible temperature change point is looped through, and for each model, the matrix of coefficients, $\boldsymbol{\beta}$, can be calculated from Eq. (6).

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{6}$$

The $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices for each of the 5 model types are shown.

$$\textbf{2P}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_n \end{bmatrix} \tag{7}$$

$$\textbf{3P-C}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & (T_1 - B_2)^+ \\ 1 & (T_2 - B_2)^+ \\ \vdots & \vdots \\ 1 & (T_n - B_2)^+ \end{bmatrix} \tag{8}$$

$$\textbf{3P-H}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & (B_2 - T_1)^+ \\ 1 & (B_2 - T_2)^+ \\ \vdots & \vdots \\ 1 & (B_2 - T_n)^+ \end{bmatrix} \tag{9}$$

$$\textbf{4P}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & (T_1 - B_3)^+ & (B_3 - T_1)^+ \\ 1 & (T_2 - B_3)^+ & (B_3 - T_2)^+ \\ \vdots & \vdots & \vdots \\ 1 & (T_n - B_3)^+ & (B_3 - T_n)^+ \end{bmatrix} \tag{10}$$

$$\textbf{5P}, \quad \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & (T_1 - B_3)^+ & (B_4 - T_1)^+ \\ 1 & (T_2 - B_3)^+ & (B_4 - T_2)^+ \\ \vdots & \vdots & \vdots \\ 1 & (T_n - B_3)^+ & (B_4 - T_n)^+ \end{bmatrix} \tag{11}$$

The root mean squared error (RMSE) for each model with a specific change point can be calculated, and the actual temperature change point corresponds to the model with the lowest RMSE. The RMSE definition is shown in Eq. (12), where n is the number of data points, p is the number of parameters in the model, $\boldsymbol{Y}$ the array of dependent variable data, $\hat{\boldsymbol{Y}}$ is the array of predicted dependent variable data, $\boldsymbol{X}$ is the matrix of the independent variable data, and $\boldsymbol{\beta}$ is the matrix of the model coefficients. A flowchart of the procedure for determining the temperature change point for a 3P-C, 3P-H, or 4P model is shown in Fig. 2. A 4P model has one change point and requires one loop in the change point calculation. The 5P model requires two nested loops since it has two change points.

$$\text{RMSE} = \sqrt{\frac{\sum_n^{i=1}(Y_i - \hat{Y}_i)^2}{n-p}} = \sqrt{\frac{\boldsymbol{Y^T Y} - \boldsymbol{\beta^T X^T Y}}{n-p}} \tag{12}$$

For the 5P model, there are two separate temperature change points. The search ranges need to be slightly modified, and two nested "for" loops will be necessary. The number of loops will be a triangular number, relating to Eq. (13), where n is the number of different change points tested in the outer loop.

$$\frac{n(n+1)}{2} \tag{13}$$

As an example, with a search range of 45 °C [81 °F] and a step size of 0.15 °C [0.27 °F], the number of loops would be 45,451. The regression process for this search range, including calculating parameters such as CV, R^2, and coefficient t-statistics, took 1.0 s on average for an Excel VBA application and 5.1 s on average for a Matlab script, using a desktop computer with an Intel i5 processor (3.10 GHz) and 4 GB of RAM.

2.2. Selecting the change point model type

The algorithm makes a decision on whether to use a 2P, 3P-Heating, 3P-Cooling, or 5P model based on three separate tests. The first test checks whether the slopes of the model are appropriate for the model shape. A second test checks if the coefficients are statistically significant beyond a specified threshold. The third test checks if enough data points are in each temperature section of the change point model.

2.2.1. Shape test

Only certain change point model shapes are expected with actual HVAC systems. For example, take a building using electricity for cooling and heating. The electricity use will increase as

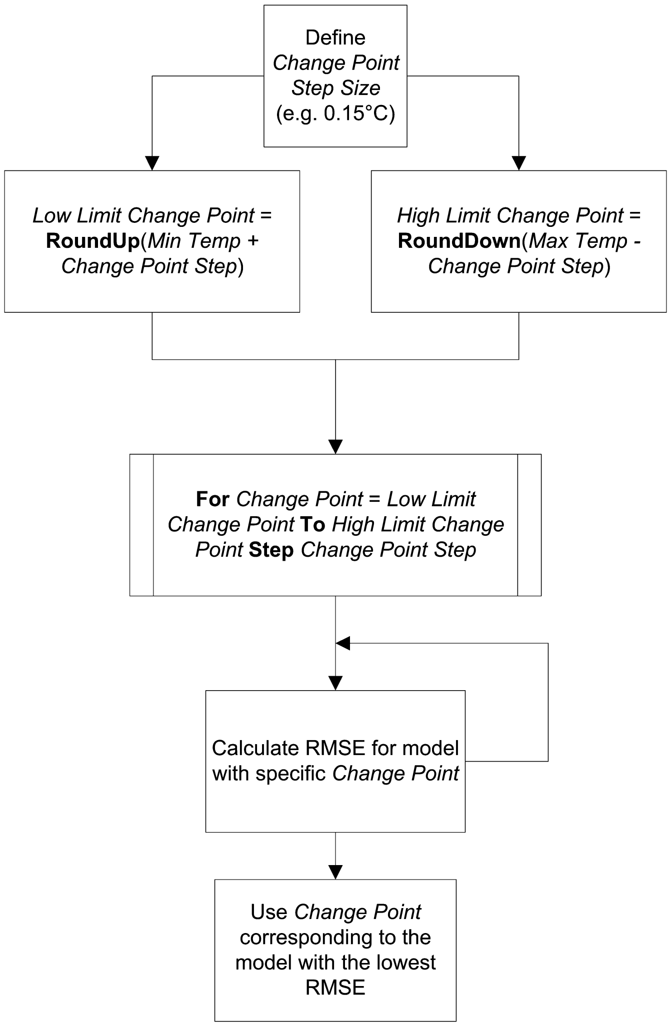

Fig. 2. Flowchart for determining the temperature change point in the 3P-C, 3P-H, and 4P model. The 5P model would require another for loop.

the temperature increases because of an increased cooling load. In the same way, the electricity use will increase as the temperature decreases when additional heating is necessary. There likely is also a minimum level of electricity use at moderate temperatures when heating and cooling loads are low. This results in a U-shaped change point model.

In an ideal sense, for a single zone building, an outdoor air temperature corresponding to when the internal heat gains are equivalent to the net heat loss through the building envelope may be considered a balance point temperature. As the outdoor air temperature changes from this balance temperature, more energy use for air conditioning is expected. From this logic, a positive slope should only be in a temperature range above a balance temperature and a negative slope should only be in a temperature range below a balance temperature.

An upside-down U-shape therefore does not qualify because it begins with a positive slope which should only occur at temperatures greater than a balance temperature. The shape then levels out and has a negative slope, which should only occur in temperatures below a balance temperature.

The other mathematical possibilities for 5P shapes are not expected from real systems. The permitted and non-permitted 5P shapes are shown in Fig. 3. Similarly, only certain shapes for the 4P, 3P-C, and 3P-H models are expected. The allowed shapes are shown in Figs. 4 through 6.

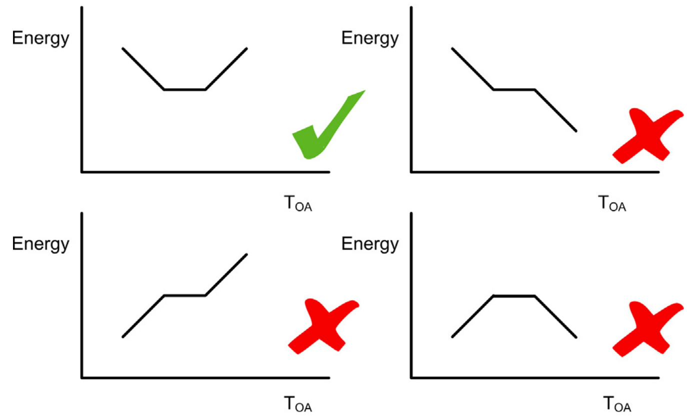

Fig. 3. Permitted and non-permitted shapes for the 5P model.

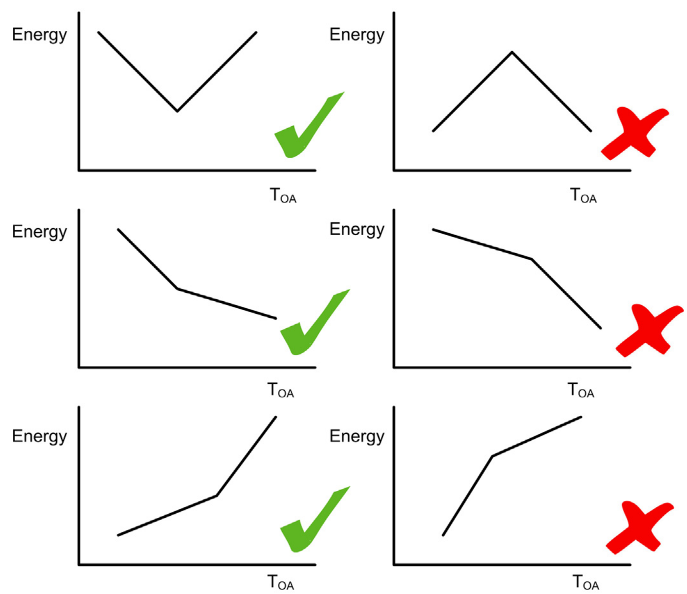

Fig. 4. Permitted and non-permitted shapes for the 4P model.

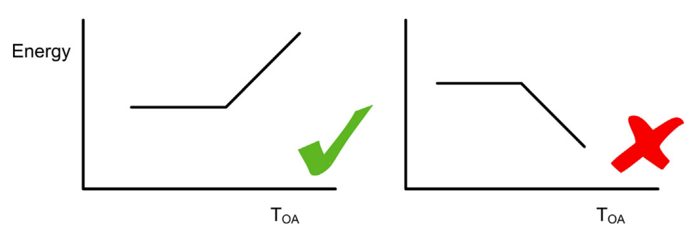

Fig. 5. Permitted and non-permitted shape for the 3P-C model.

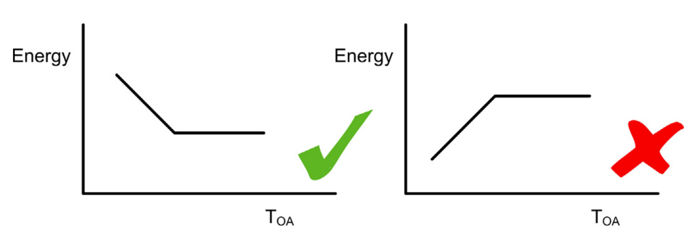

Fig. 6. Permitted and non-permitted shape for the 3P-H model.

One way to understand why the specific 4P shapes are disallowed is to imagine replacing the two linear sections with an exponential curve. The curvature should have an asymptote at zero energy. All the shapes on the right hand side of Fig. 4 would cross the x-axis (no energy consumption) if continued, resulting in negative energy consumption predictions.

The 2P model, a simple linear regression, can be appropriate with either a positive or a negative slope.

2.2.2. Significance test

Checking the significance of coefficients is common in regression analysis. Assuming the statistical assumptions underlying the linear least squares regression are met, a t-test can be used to determine whether a specific coefficient is statistically different from 0. If a coefficient is not found to be significant, a simpler model (less parameters) may be more appropriate.

The t-statistic and p-value for coefficients is an output from many statistical packages. If a regression package is not available, or if doing a personal implementation, it is easier to calculate the t-statistic than the p-value, since calculating the p-value involves integrating the probability distribution function of the t distribution. If the $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ matrix is denoted as

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} c_{00} & c_{01} & \cdots & c_{0k} \\ c_{10} & c_{11} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ c_{k0} & \vdots & \vdots & c_{kk} \end{bmatrix} \tag{14}$$

the t-statistic for coefficient i is then

$$t_i = \frac{\beta_i}{s\sqrt{c_{ii}}} = \frac{\beta_i}{\sqrt{(\mathrm{SSE}/(n-p))}\sqrt{c_{ii}}} = \frac{\beta_i}{\sqrt{((\boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y})/(n-p))}\sqrt{c_{ii}}} \tag{15}$$

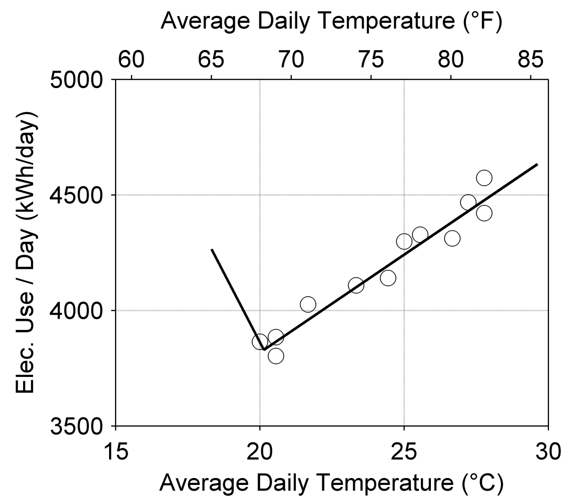

Fig. 7. Example data set showing the necessity of the third test.

where s is the estimated standard deviation of the model error ε, SSE is the sum of the squared errors, n is the number of data points, and p is the number of parameters of coefficients. Explanations and examples of matrix mathematics for multiple linear regression can be found in [45].

A t-statistic threshold of 2.0 worked well for determining the significance of parameters. This algorithm was developed expecting a full year of monthly data being used for creating a baseline model. With $n = 12$ and a t-statistic of 2, the critical two-tailed p-value equates to approximately 0.073 for models with 2 *regression* parameters (2P, 3P-C, and 3P-H models) and approximately 0.060 for models with 3 *regression* parameters (4P and 5P models). This is slightly more conservative than the more common p-value of 0.05.

This work used a t-statistic threshold because it is computationally efficient to implement. If available or desired, a p-value threshold can be used instead of a t-statistic threshold.

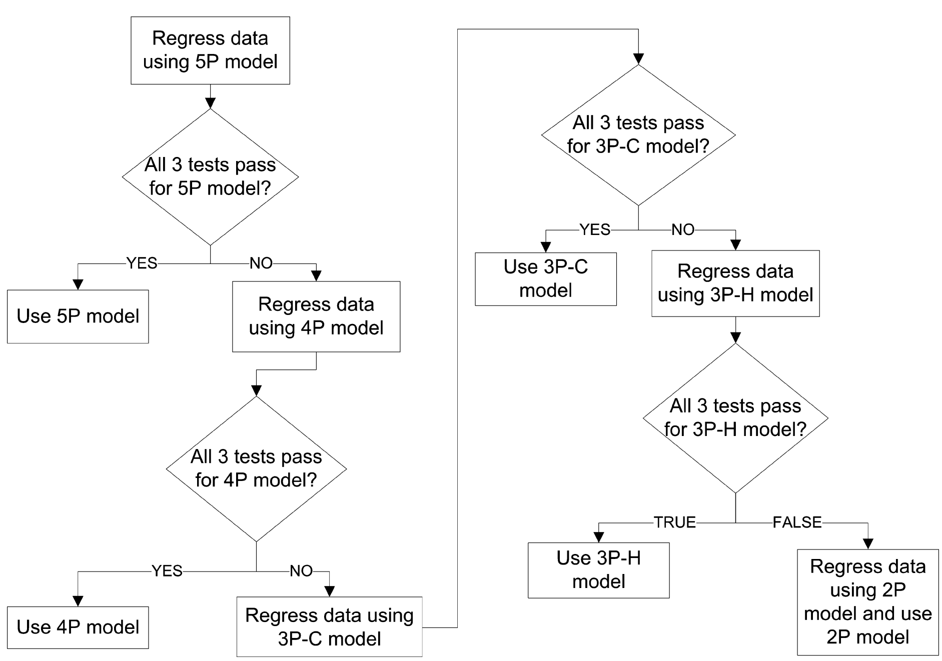

Fig. 8. Flowchart of algorithm for determining change point model type.

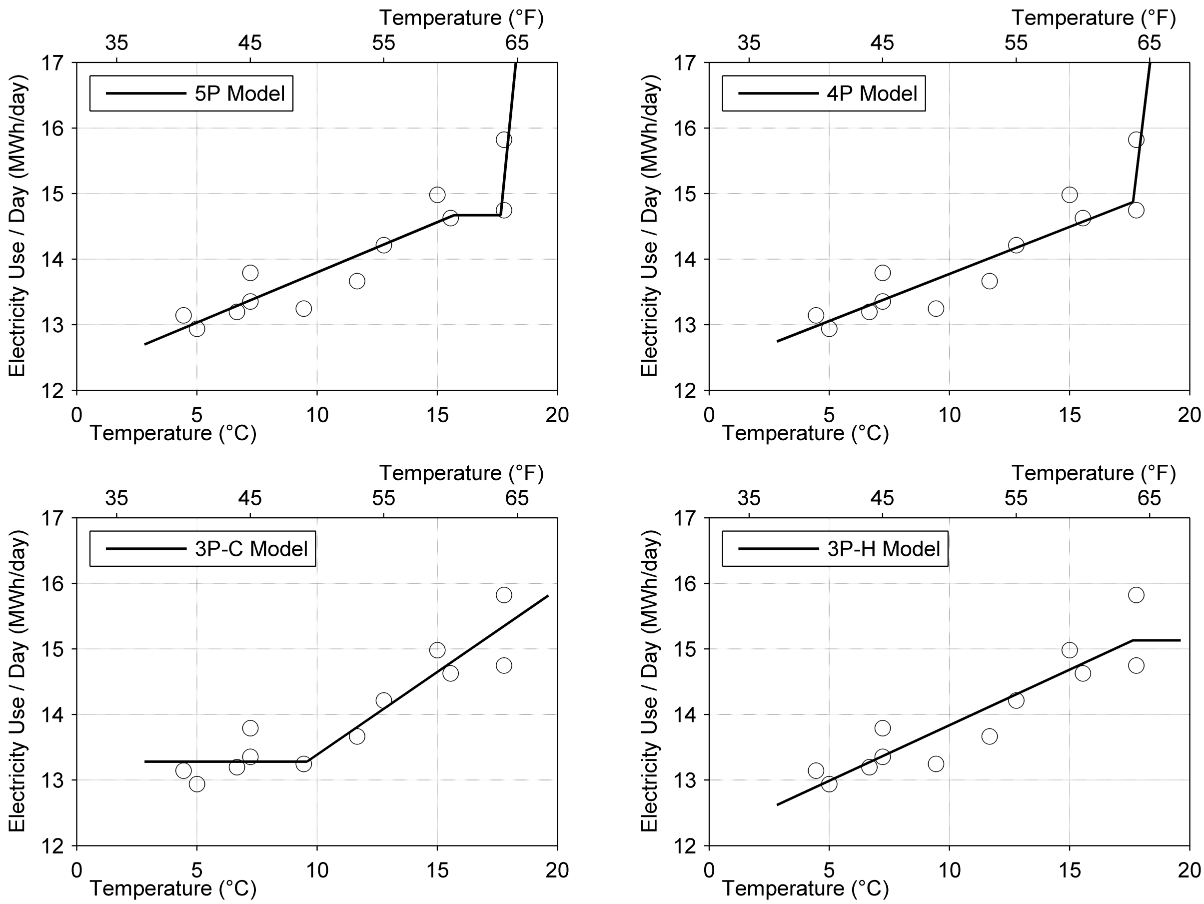

Fig. 9. 4 different change point model shapes fit to monthly average daily electricity consumption for a synthetic large office building simulated with Seattle weather.

2.2.3. *Data population test*

The final test checks whether enough data points are in each temperature section of the change point model. When using a year of monthly data, the change point model fit may end up having the correct shape and significant slopes but the slope may be far from reasonable. An example dataset where this problem is evident is shown in Fig. 7. In Fig. 7, the slope of the model for temperatures below 20 °C [68 °F] is not physically plausible. This problem often occurs when the maximum or minimum temperature in the dataset is near the physical balance temperature of the building. Essentially, one of the HVAC modes of operation is underrepresented when this occurs and unreasonable changes in slope could occur. For this reason, the final test checks if at least 3 data points are within the sloped regions of the change point models.

2.2.4. *Decision process*

The algorithm begins by regressing the monthly data using the model with the most parameters, the 5P model. By virtue of having the most parameters, if the model is appropriate as determined by the 3 tests, it will have the lowest RMSE of the possible models (although not necessarily the lowest coefficient of variation (CV) because of the influence the number of parameters has on the CV calculation). If the 5P model passes all three of the tests, the algorithm is finished and the 5P model is returned. If any of the 3 tests fail, the next most complex model, the 4P model, is tested. If the 4P model passes all 3 of the tests, the algorithm finishes and returns the 4P model. In the same manner, the 3P-C and 3P-H models are tested next. The order between the 3P-C and 3P-H model is arbitrary.

If the 5P, 4P, 3P-C, and 3P-H model all fail, a 2P (simple linear regression) model is returned. A flowchart describing the process is shown in Fig. 8.

3. Results

The algorithm was tested using synthetic monthly energy consumption data available from the U.S. Department of Energy (DOE). The synthetic data was in the form of EnergyPlus reference buildings. The reference buildings were created based on information derived from the Commercial Building Energy Consumption Survey (CBECS). The EnergyPlus files and related weather files can be found at http://energy.gov/eere/buildings/new-construction-commercial-reference-buildings [46].

The advantage of using the DOE reference buildings is a large combination of realistic data sets for different building types located in different climates can be created. There are 16 different building types such as large office buildings, warehouses, secondary schools, and hotels. 16 weather files covering nearly all the climate zones defined by [47] are also available. This results in 256 different synthetic data sets for testing algorithms.

This was reduced to 48 possible data sets by considering only the outpatient hospital, medium office building, large office building, large hotel, secondary school, and warehouse buildings with weather files from Miami, Seattle, Chicago, and Fairbanks. These locations were chosen to cover a large spectrum of U.S. weather, from a hot and humid climate in Miami to a cold climate in Fairbanks. Electricity and natural gas use were the dependent energy variables. The summary of the building types, climates, and energy variables tested are shown in Table 1. All of these 48 combinations were tested. Tests were completed by individually running all the different models (5P, 4P, etc.), having experienced modelers visually determine which model would be most appropriate, and then running the automated algorithm and checking whether the model chosen matched the model chosen by the experienced modelers. The modelers have worked with several thousand building years of data from hundreds of buildings in the course of developing a number of techniques incorporated in the IPMVP and the ASHRAE Guideline 14.

In all cases, the algorithm picked a reasonable model shape.

3.1. Large office building Seattle

An example of the algorithm process is presented for the reference large office building using weather data from Seattle, Washington. The large office building is simulated with a gas boiler for heating, and 2 water cooled chillers for cooling. The expected model shape for electricity use would be either 4P or 3P-C.

The results of fitting a 5P, 4P, 3P-C, and 3P-H model are shown in Fig. 9. The algorithm first checked the result of the 5P model fit to the electric data. The 5P model fit did not pass the "shape" test because the slope of the B_2 coefficient was negative. In other words, the model does not have the U-shape expected from a proper model. The 5P model would also have failed the third test because only 2 data points are in the high temperature region.

The algorithm then checked the results of the 4P model fit. As seen in Fig. 9, the shape test passed because the high temperature slope is greater than the low temperature slope, and they have the correct sign. The t-statistic for the B_1 coefficient was 2.66 (high temperature region), and the t-statistic for the B_2 coefficient was −6.87 (low temperature region). Both coefficients would pass the significance test because the absolute value was greater than the predetermined threshold of 2.0. The final test eliminated the 4P model. The final test checked if at least 3 data points were in each sloped section of the change point model. For the 4P model, only 2 data points are in the high temperature region (temperatures greater than the calculated change point of 17.64 °C [63.75 °F]). Similar to the example shown in Fig. 7, the test eliminates a poor 4P model fit because the slope in the high temperature region is unrealistically high.

The algorithm could have checked either the 3P-C or the 3P-H model next. If the 3P-H model was checked first, it would fail the shape test. The 3P-C model passes all three of the tests: shape, significance, and data population. The results of the 3P-C model would be returned from the algorithm.

3.2. Outpatient hospital in Seattle

An intriguing test case occurred with the outpatient hospital simulated with Seattle weather. The expected model shape for electricity use consumed by the outpatient hospital was 5P because electricity use should increase with increased chiller use in warmer weather and increased electrical energy use due to electrical reheat in colder weather. The maximum monthly average temperature across the year for Seattle however, is close to the physical balance temperature of the building. The algorithm returned the 3P-H model shown in Fig. 10. Though in a physical sense the model should be 5P, the 3P-H case will result in the least amount of error if new data comes in outside the range of temperatures used in the baseline.

Table 1
Different synthetic data set possibilities.

Building types tested	Climate locations	Energy variables
Outpatient hospital	Miami	Whole building electricity
Medium office building	Seattle	Whole building natural gas
Large office building	Chicago	
Secondary school	Fairbanks	
Large hotel		
Warehouse		

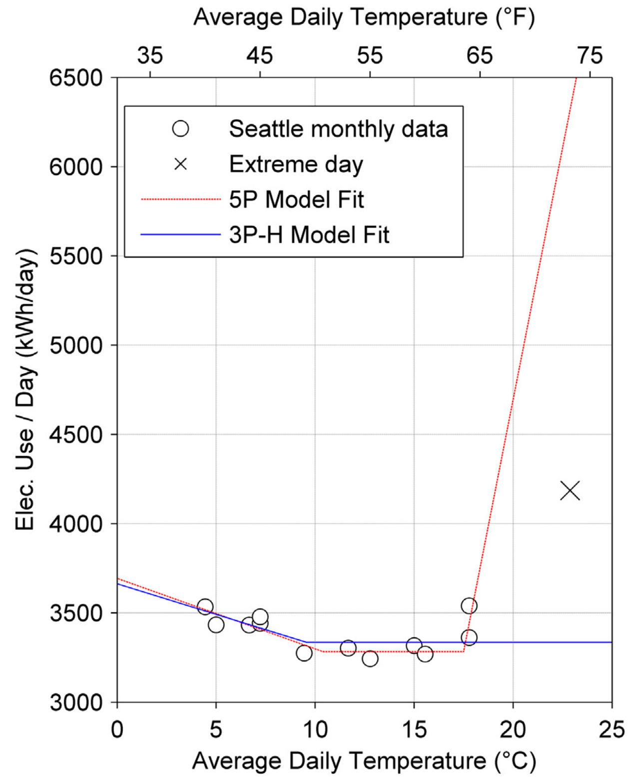

Fig. 10. 5P and 3P-H models fit for monthly average daily electricity consumption for a synthetic outpatient hospital in Seattle.

As an example, when a single hot day is taken from the same weather file, the 3P-H results in the least amount of error, although in the absolute sense it was not the appropriate model. For a day with an average temperature of 22.9 °C [73.2 °F], the energy use was simulated to be 4186 kW. The 3P-H model under predicted the energy use by approximately 850 kWh, and the 5P model over predicted the energy use by approximately 2140 kWh.

This case shows the difficulty in automating the selection of a model having less than 3 data points on either side of the physical change point temperature of the building. This algorithm is conservative since it will usually return a model where the energy use remains constant (like the 3P-H model in Fig. 10). It is an example of how making predictions outside the range of data used to create the model can result in large errors. This problem is lessened with higher resolution energy data, such as daily, hourly, or sub hourly.

3.3. Complete final results

The complete set of final results returned from the algorithm are shown in Figs. 11 and 12. Fig. 11 shows the results for the electricity energy component and Fig. 12 shows the results for the natural gas energy component. Each of the 6 buildings is shown on a separate plot, and the results using the four different weather locations are shown as different series. The variation in data for energy use at similar outdoor air temperatures can be attributed to other differences in weather such as humidity and solar load.

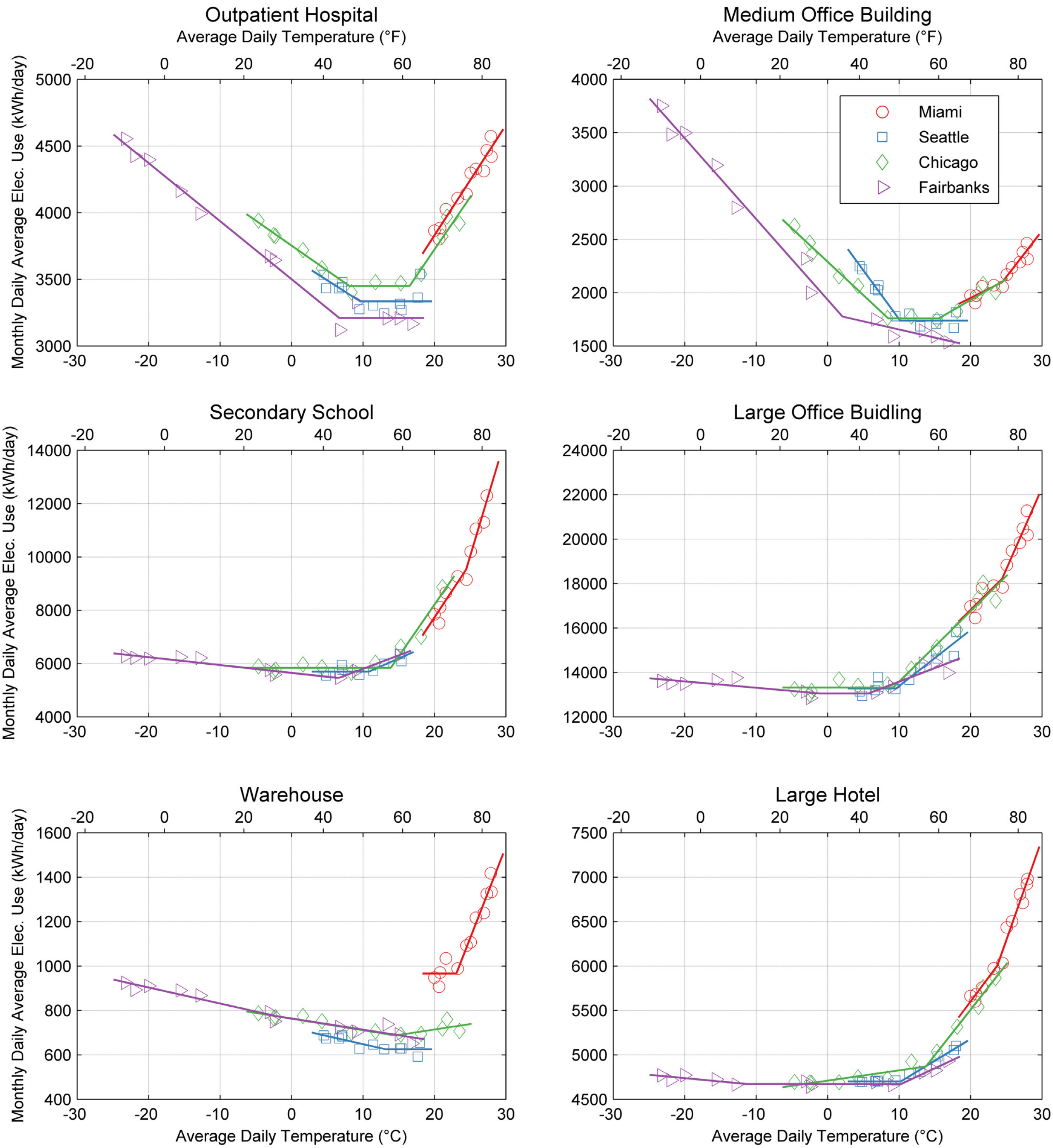

Fig. 11. Complete results for electricity consumption component.

The complete energy signature of the building becomes apparent when a large set of different weather files are used. For example, looking at the outpatient hospital electricity profile, the HVAC system clearly uses electricity for cooling and heating, which would relate to a 5P model shape. Yet due to weather differences, Miami results in a 2P model, Chicago a 5P model, and Seattle and Fairbanks result in a 3P-H model. This is an example of how predicting energy use outside of the baseline weather range can result in poor predictions and why selecting a proper model shape using monthly data can be difficult to automate. Having a limited range of temperatures for the baseline data not only causes difficulties in this automation process, but also causes difficulties in finding any inverse model that is most physically reasonable for a particular building and HVAC system [37,48].

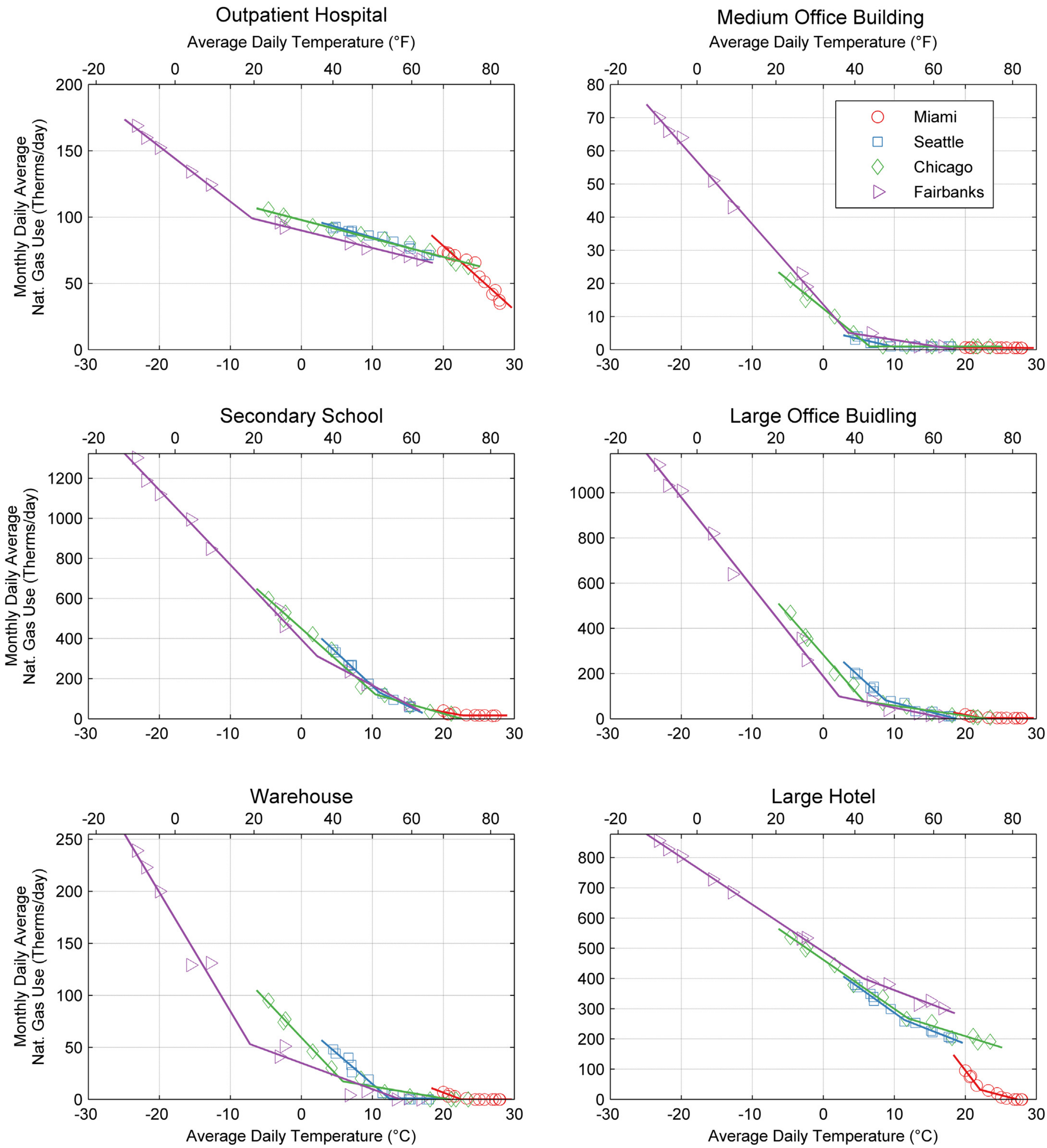

Fig. 12. Complete results for natural gas consumption component.

4. Discussion and conclusions

An algorithm for automating the process of selecting a temperature based change point model for monthly whole building energy use was presented. The algorithm was tested with synthetic EnergyPlus data for an outpatient hospital, medium office building, large office building, large hotel, secondary school, and warehouse, with weather data from Chicago, Miami, Seattle, and Fairbanks. The algorithm selected a reasonable model of the 5 different change point model types employed for each of the 48 cases tested.

Testing of this algorithm was limited to monthly energy consumption data. Future work will investigate adapting the current algorithm for daily or hourly energy consumption data. It is hypothesized that the proposed algorithm would be adequate for smaller time scales will minimal modification.

The intended use of the algorithm is to automate the process of selecting a temperature dependent change point model for monitoring and verification purposes. An analyst can simply check the results to make sure the building does not require a different model type. The results should be verified because some buildings are not modeled well with temperature as the only independent variable. In some cases, a calibrated simulation or multiple linear regression may be a better option. Future work will include adapting to models with multiple independent variables beyond outdoor air dry-bulb temperature, and using monthly, weekly, and hourly data.

References

[1] ASHRAE, Guideline 14-2002, Measurement of Energy and Demand Savings, 2002.
[2] International Performance Measurement & Verification Protocol Committee, International Performance Measurement & Verification Protocol, Concepts and Options for Determining Energy and Water Savings, vol. 1, 2002.
[3] A. Rabl, A. Rialhe, Energy signature models for commercial buildings: test with measured data and interpretation, Energy Build. 19 (1992) 143–154.
[4] S. Hammarsten, A critical appraisal of energy-signature models, Appl. Energy 26 (1987) 97–110.
[5] J. Sjögren, S. Andersson, T. Olofsson, Sensitivity of the total heat loss coefficient determined by the energy signature approach to different time periods and gained energy, Energy Build. 41 (2009) 801–808.
[6] H. Masuda, D.E. Claridge, Statistical modeling of the building energy balance variable for screening of metered energy use in large commercial buildings, Energy Build. 77 (2014) 292–303.
[7] ASHRAE, ASHRAE Handbook, Fundamentals, 2001.
[8] M.F. Fels, PRISM: an introduction, Energy Build. 9 (1986) 5–18.
[9] J.H. Eto, On using degree-days to account for the effects of weather on annual energy use in office buildings, Energy Build. 12 (1988) 113–127.
[10] J.M. MacDonald, D.M. Wasserman, Investigation of Metered Data Analysis Methods for Commercial and Related Buildings, Oak Ridge National Laboratory, 1989.
[11] S. Katipamula, T.A. Reddy, D.E. Claridge, Development and application of regression models to predict cooling energy consumption in large commercial buildings, Solar Eng. (1994) 307.
[12] J.C. Lam, K.K.W. Wan, D. Liu, C.L. Tsang, Multiple regression models for energy use in air-conditioned office buildings in different climates, Energy Convers. Manag. 51 (2010) 2692–2697.
[13] S. Asadi, S.S. Amiri, M. Mottahedi, On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design, Energy Build. 85 (2014) 246–255.
[14] J.S. Hygh, J.F. DeCarolis, D.B. Hill, S. Ranji Ranjithan, Multivariate regression as an energy assessment tool in early building design, Build. Environ. 57 (2012) 165–175.
[15] A. Aranda, G. Ferreira, M.D. Mainar-Toledo, S. Scarpellini, E. Llera Sastresa, Multiple regression models to predict the annual energy consumption in the Spanish banking sector, Energy Build. 49 (2012) 380–387.
[16] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V.I. Hanby, Regression models for predicting UK office building energy consumption from heating and cooling demands, Energy Build. 59 (2013) 214–227.
[17] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy Build. 57 (2013) 302–312.
[18] R.Z. Freire, G.H.C. Oliveira, N. Mendes, Development of regression equations for predicting energy and hygrothermal performance of buildings, Energy Build. 40 (2008) 810–820.
[19] A. Dhar, T. Reddy, D. Claridge, Modeling hourly energy use in commercial buildings with Fourier series functional forms, J. Solar Energy Eng. 120 (1998) 217–223.
[20] A. Dhar, T. Reddy, D. Claridge, A Fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable, J. Solar Energy Eng. 121 (1999) 47–53.
[21] Z. Hai-xiang, F. Magoules, New parallel support vector regression for predicting building energy consumption, in: IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making (MDCM), 2011, pp. 14–21.
[22] H.X. Zhao, F. Magoulès, Parallel support vector machines applied to the prediction of multiple buildings energy consumption, J. Algorithms Comput. Technol. 4 (2010) 231–249.
[23] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Appl. Energy 86 (2009) 2249–2256.
[24] Y. Zhang, W. Qi, Interval forecasting for heating load using support vector regression and error correcting Markov chains, in: International Conference on Machine Learning and Cybernetics, vol. 2, 2009, pp. 1106–1110.
[25] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build. 37 (2005) 545–553.
[26] L. Xuemei, L. Jin-hu, D. Lixing, X. Gang, L. Jibin, Building cooling load forecasting model based on LS-SVM, in: Asia-Pacific Conference on Information Processing, vol. 1, 2009, pp. 55–58.
[27] L. Xuemei, D. Yuyan, D. Lixing, J. Liangzhong, Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering, in: International Conference on Computer and Communication Technologies in Agriculture Engineering (CCTAE), vol. 1, 2010, pp. 438–441.
[28] Z. Hou, Z. Lian, An application of support vector machines in cooling load prediction, in: International Workshop on Intelligent Systems and Applications, 2009, pp. 1–4.
[29] J.A. Rodger, A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings, Expert Syst. Appl. 41 (2014) 1813–1829.
[30] C. Buratti, M. Barbanera, D. Palladino, An original tool for checking energy performance and certification of buildings by means of artificial neural networks, Appl. Energy 120 (2014) 125–132.
[31] K. Li, H. Su, J. Chu, Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: a comparative study, Energy Build. 43 (2011) 2893–2899.
[32] S.L. Wong, K.K.W. Wan, T.N.T. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Appl. Energy 87 (2010) 551–557.
[33] Y. Heo, V.M. Zavala, Gaussian process modeling for measurement and verification of building energy savings, Energy Build. 53 (2012) 7–18.
[34] M.C. Burkhart, Y. Heo, V.M. Zavala, Measurement and verification of building systems under uncertain data: a Gaussian process modeling approach, Energy Build. 75 (2014) 189–198.
[35] J.K. Kissock, T.A. Reddy, D.E. Claridge, Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings, J. Solar Energy Eng. 120 (1998) 168–176.
[36] J.K. Kissock, J. Haberl, D.E. Claridge, Development of a toolkit for calculating linear, change-point linear and multiple-linear inverse building energy analysis models, ASHRAE Research Project 1050-RP, 2002.
[37] B. Abushakra, M. Paulus, T.A. Reddy, V. Singh, Measurement, modeling, analysis, and reporting protocols for short-term M&V of whole building energy performance, ASHRAE Research Project 1404-RP, ASHRAE, 2014.
[38] V. Singh, T.A. Reddy, B. Abushakra, Predicting annual energy use in buildings using short-term monitoring: the dry bulb temperature analysis (DBTA) method (RP-1404), ASHRAE Trans. 120 (2013) 397–405.
[39] V. Singh, T.A. Reddy, B. Abushakra, Predicting annual energy use in buildings using short-term monitoring and utility bills: the hybrid inverse model using daily data (HIM-D), ASHRAE Trans. 119 (2013) 169–180.
[40] C. Reynolds, P. Komor, M. Fels, Using monthly billing data to find energy efficiency opportunities in small commercial buildings, in: Proceedings of the ACEEE 1990 Summer Study on Energy Efficiency in Buildings, vol. 10, 1990, pp. 10.221–10.232.
[41] J.K. Kissock, Methodology to Measure Retrofit Energy Savings in Commercial Buildings, Department of Mechanical Engineering, Texas A&M, 1993.
[42] R. Crawford, R. Dykowski, S. Czajkowski, A segmented linear least-squares modeling procedure for nonlinear HVAC components, ASHRAE Trans. 97 (1991) 11–18.
[43] D. Ruch, D.E. Claridge, A four-parameter change-point model for predicting energy consumption in commercial buildings, J. Solar Energy Eng. 114 (1992) 77–83.
[44] J.K. Kissock, W. Xun, R. Sparks, D.E. Claridge, EModel Software, Texas A&M University, Energy Systems Laboratory, 1994.
[45] W. Mendenhall, T. Sincich, A Second Course in Statistics: Regression Analysis, 6th ed., Pearson Education, Upper Saddle River, NJ, 2003.
[46] U.S. Department of Energy, http://energy.gov/eere/buildings/new-construction-commercial-reference-buildings, 2014.
[47] R.S. Briggs, R.G. Lucas, Z.T. Taylor, Climate classification for building energy codes and standards: Part 2-zone definitions, maps, and comparisons, ASHRAE Trans. 109 (2003) 122.
[48] B. Abushakra, Short-Term Monitoring, Long-Term Prediction of Building Energy Use, Department of Mechanical Engineering, Texas A&M, 2001.