AN EXPANDED AND UPDATED FRAMEWORK

FOR WHOLE-FACILITY ENERGY CONSUMPTION

STATISTICAL MODELS OF COMMERCIAL BUILDINGS

A Dissertation

by

HONGXIANG FU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David E. Claridge
Co-Chair of Committee,	Juan-Carlos Baltazar
Committee Members,	Bryan Rasmussen
	Jeff S. Haberl
Head of Department,	Andreas A. Polycarpou

August 2021

Major Subject: Mechanical Engineering

Copyright 2021 Hongxiang Fu

ABSTRACT

A significant portion of energy consumption occurs in buildings today. Accurate and easy-to-implement methods are needed to calculate building energy consumption for a wide range of applications, including efficiency assessment, consumption projection, and measurement and verification, to name a few. There are a number of approaches for building energy estimation but the statistical methods have remained popular. As the availability and quality of building energy data continue to improve, the methodologies behind building energy calculation also require updates. This work proposes three new technologies to bring contemporary mindsets to the application of whole-building energy consumption statistical models.

The first is a specialised model formulation for the heating hot water consumption for commercial buildings with constant volume reheat systems. It has been observed that the heating consumption of this system type has an unexpected local increase with an increase in ambient temperature caused by dehumidification and reheat. The proposed new method can improve model fit with statistical significance and remove the local trend in the residuals.

The second is the use of domestic cold water use or non-HVAC electricity use as an occupancy proxy for building energy models. It is found that combining domestic cold water use with a clustering technique was able to improve model fit by 2.9 percentage points of the CV-RMSE, on average, on top of 14.2% from the traditional weekday-and-weekend method. In the study it was found that other methods, i.e., the use

of electricity use as occupancy proxy and the additional of a linear term, were not able to improve the model fit consistently.

Finally, a procedure was proposed to examine all data separation or grouping possibilities automatically and comprehensively with pre-defined elementary day-types through a series of lack-of-fit *F*-tests. This procedure suggests a best separation that balances between model accuracy and simplicity. It was tested on measured energy consumption data of 76 case study commercial buildings. The proposed method weighed simplicity more heavily than traditional statistical complexity-penalising metrics. The new method improved the CV-RMSE by 7.6 percentage points, on average, and helped extract information to help better understand the buildings' energy consumption patterns.

DEDICATION

To Cats.

ACKNOWLEDGEMENTS

I would like to thank my advisory committee. I thank my advisor and committee chair Dr. Claridge. He offered me not only academic guidance but also wisdom of life from which I will continue to benefit. I thank my supervisor and committee co-chair Dr. Baltazar. His expertise in this field has ironed out many details in this dissertation and his amicability made my work life in the ESL an enjoyable one. I thank my committee member Dr. Rasmussen. He is the first professor I interacted in depth when I started my Ph.D. study and he helped me survive the first year filled with unknowns. I thank also my committee member Dr. Haberl. I was along side him in many ASHRAE committee meetings and his dedication inspired me immensely.

I would also like to thank my family. I thank my mom YUAN Hong 袁红 and my dad FU Dong 付东 for their unwavering support through out all adventures of my life. I thank my maternal grandpa YUAN Shiyu 袁世余 and grandma XIONG Xiuzhen 熊秀珍 as their videocalls give me strength across the Pacific Ocean. I thank my paternal grandpa Prof. FU Shuyi 傅树义 for always being my shining beacon. I thank my paternal grandma ZHUANG Chengluan 庄成栾 for watching over me from heaven.

There is a long list of people that I have attempted but failed to include in this page including all my incredible friends. From the bottom of my heart:

THANK YOU!

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. David E. Claridge (chair, Department of Mechanical Engineering), Dr. Juan-Carlos Baltazar (co-chair, Department of Architecture), Dr. Bryan Rasmussen (Department of Mechanical Engineering), and Dr. Jeff S. Haberl (Department of Architecture).

The data are provided by Dr. Yasuko Sakurai and her team from Utilities and Energy Services at Texas A&M University.

The clustering analysis in Section 4 was performed and partially written by Mr. Shinwoo Lee.

Mr. Pei, Hongming from the Department of Information, China Bank of Communications Sichuan Provincial Branch and Dr. Zhang, Yunfeng from Microsoft provided clarifications concerning the statistical concepts used in Sections 4 and 5.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Graduate Research Assistantship from the Energy Systems Laboratory – Texas A&M Engineering Experiment Station.

NOMENCLATURE

Abbreviations

4P-CP	4-parameter change-point (model)
ACal	Academic calendar
ANN	Artificial neural networks
ANOVA	Analysis of variance
ARIMA	Autoregressive integrated moving average
BP	Break-point
CHW	Chilled water
CP	Change-point
CVRH	Constant volume reheat
CV-RMSE	Coefficient of variance – root mean square error
DBSCAN	Density-based spatial clustering of applications with noise
DCW	Domestic cold water
DOW	Day of week
DT	Day-type
EBCx	Existing building commissioning
EDT	Elementary day-type
ELE	Electricity
HHW	Heating hot water
IPMVP	International Performance Measurement and Verification Protocol
kNN	*k*-th nearest neighbour

LOF	Lack-of-fit
MLR	Multiple linear regression
M&V	Measurement and verification
OAT	Outdoor air dry-bulb temperature
O&M	Operation and maintenance
PCA	Principal component analysis
SLR	Simple linear regression
SSE	Sum of squares error
SVM	Support vector machine
TDB	Dry-bulb temperature

Symbols

N, n	Total or sub-sample size
p	Number of regression parameters
R^2	Coefficient of determination
R^2_{adj}	Adjusted coefficient of determination
T_{CL}	Cooling coil air leaving temperature

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

In 2019, the United States used 16% of end-use energy in the residential sector, 12% in the commercial sector [1, 2], implying how energy efficiency in buildings have made and still can make substantial impact on energy conservation efforts as a whole. Building energy calculation is an important technique to estimate or model the building's energy consumption for the purpose of design optimisation, baselining, savings verification, model predictive control, data quality assurance, and fault detection, to name a few. The ASHRAE Handbook – Fundamentals [3] classified energy calculation methods as shown in Figure 1.1. The statistical method is the focus of this dissertation. As a data-driven approach, it has been recognised to be simpler to use and able to capture the buildings as-built performance more accurately than the forward approach [3-7].

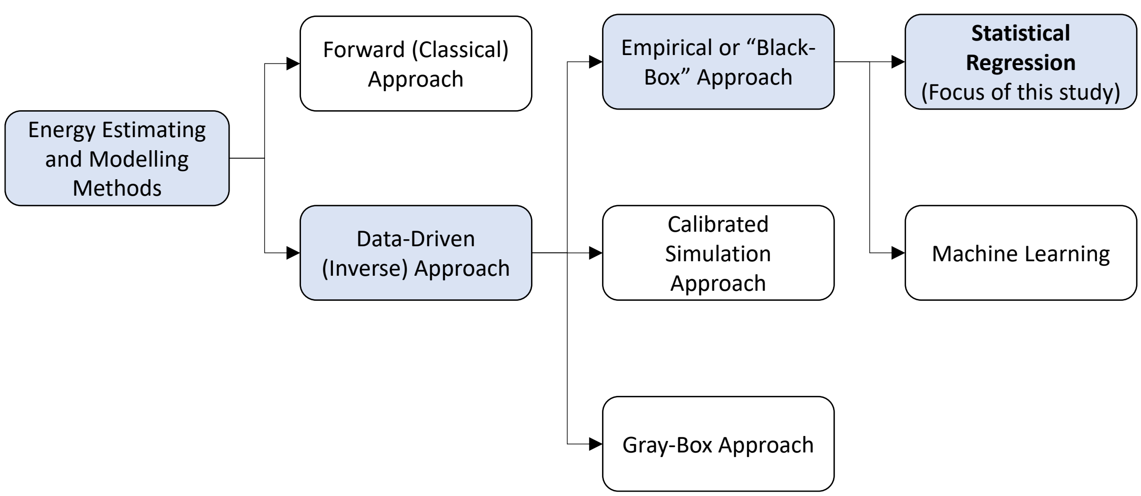

Figure 1.1 Classification of energy estimating and modelling methods, summarised and adapted from [3] and [8].

Whole-building energy calculation methods are also classified by guiding documents of Measurement and Verification (M&V), one of its most important applications. Both EVO IPMVP [9] and ASHRAE Guideline 14 (2014) [10] provide a standard set of energy savings calculation procedures. They include component isolation, whole-facility before-and-after measurements, and calibrated simulation [11, 12]. The focus of this work falls under the whole-building approach (IPMVP Option C and ASHRAE Guideline 14-2014 Section 5.1). One of the many values of this approach is its ability to quickly assess the building energy use pattern without detailed information about its system configurations, occupancy schedule, space type, etc., which reduces the cost of this activity. An accurate description of the building and its system is vital for the forward approach and calibrated simulation. However, in many cases it is unavailable, inaccurate, or difficult to verify. A purely data-driven analysis circumvents such impediments, making it less expensive. Statistical models are also preferred as a simple indication of energy efficiency for a building or a plant compared to a detailed simulation programme [13].

1.2. Purpose and Objectives

Since these guidelines for building energy modelling [9, 10] were first established, the availability and quality of building energy data have vastly improved. More and more often, energy analysts have access to daily or sub-daily energy data instead of relying solely on monthly utility bills to perform the analysis. In the meantime, the computing devices also rapidly evolved making methods and algorithms formerly difficult or expensive to execute now easy and affordable. With these factors in mind, this research aims to provide an expanded and updated framework for whole-facility energy consumption statistical models of commercial buildings in the following aspects:

1. Expand and specialise the base functional form of the whole-building energy statistical model;
2. Investigate and compare the inclusion of a proxy occupancy variable; and
3. Automatically and comprehensively search for optimal day-typing of the energy data.

1.3. Methods

This is a journal-article style dissertation consisting of four parts that are published journal articles or unpublished manuscripts. All of them have been revised and adapted to fit the format of this dissertation.

- Section 2 is based on a published literature review [14].
- Sections 1.3.1 and 3 are based on an extended abstract presented at an ASHRAE conference [15] and a published journal article [16].
- Sections 1.3.2, 1.3.3 and 4 are based on a conference paper presented at an ASHRAE conference [17] and an unpublished manuscript.
- Section 5 is based on an unpublished manuscript.

This section will also introduce three methodologies that will be applied throughout the following sections.

1.3.1. Change-Point Model

Figure 1.2 is an example of a generic change-point (CP) model regressed to measured HHW consumption data. The formulation in Equation (1) is the four-parameter version proposed by Ruch and Claridge [18]. The CP model has been successful as it has a sound physical interpretation and appropriately describes the relationship between energy consumption and TDB. This formulation describes one linear relationship for $T < T_{CP}$ and another linear relationship for $T > T_{CP}$, but the overall model is nonlinear due to interaction between regression parameters [19].

$$y = \beta_1 + \beta_2(x - T_{CP})^- + \beta_3(x - T_{CP})^+ \tag{1}$$

where β_1 – intercept, β_2 – left slope, β_3 – right slope, T_{CP} – change-point temperature, $()^-$ – the value in the parentheses is set to zero if it is positive, and $()^+$ – the value in the parentheses is set to zero if it is negative.

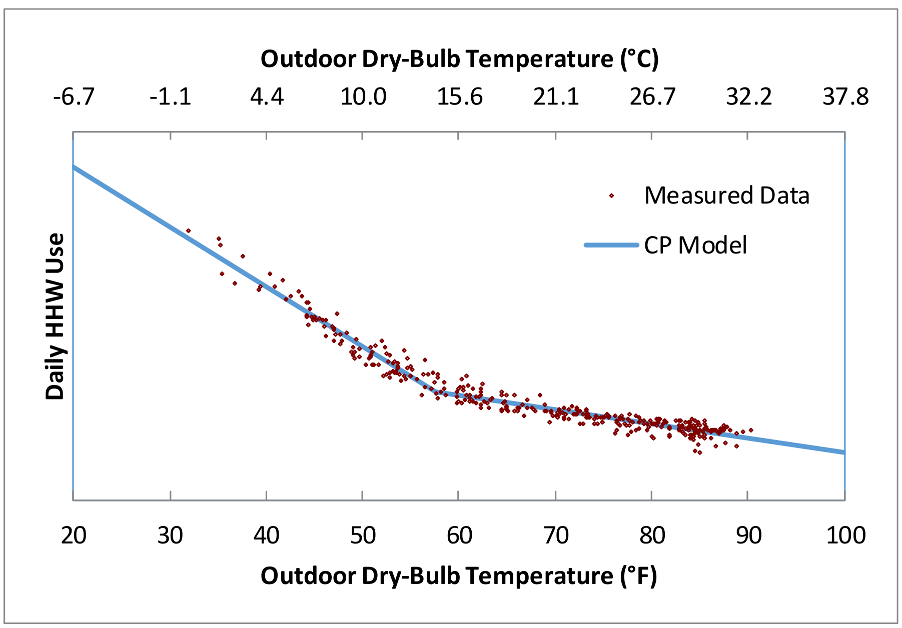

Figure 1.2 Example of a CP Model with One Year of Daily HHW Consumption Data

1.3.2. Model Series

When the data are separated into multiple groups, say *k* groups, different regression models can be fitted to these data subsets. The *k* subset models are then written in a compact form expressed in Equation (2).

$$\hat{y}_i = \sum_i \tau_{ik} \cdot (\beta_{0k} + \beta_{1k}(x_i - \beta_{3k})^- + \beta_{2k}(x_i - \beta_{3k})^+) \tag{2}$$

where $\tau_{i,k} = \begin{cases} 1, \text{ day } i \in \text{group } k \\ 0, \text{ day } i \notin \text{group } k \end{cases}$ is an indicator variable.

This expression will be called a *model series* in this work. It essentially means that the *k*-th subset model is chosen for day *i* according to whether day *i* belongs to cluster *k*. The purpose of expressing all the subset models in a compact form is again so that models with different numbers of parameters can be compared under a uniform framework.

1.3.3. DBSCAN Clustering Technique

There are a variety of clustering methods such as *k*-means, *k*-prototype, hierarchical clustering, spectral clustering, and so on. In this work, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is chosen for multiple reasons. First of all, it has a record of excellent clustering performance [20]. Second, it is highly flexible and not restricted by the shape or placement of the clusters as long as the constituent data points are relatively close together. Third, like many machine learning methods, it has hyperparameters that must be predetermined and then tuned by the user. But DBSCAN is relatively simple to use as it has only two of them to tune: ϵ and minPts. Also, it does not require users to predetermine the number of clusters, which is particularly valuable when the user does not have prior knowledge of the number of clusters in the data.

The DBSCAN method was introduced by Ester et al. [21], and it has a straightforward working principle which will be demonstrated using Figure 1.3.The algorithm scans every data point with a user-defined distance ϵ, marked by circles centred on each point. When the number of points within ϵ of one or more other points is equal to or larger than minPts, a cluster is formed, which either (1) creates a new cluster if these points are not yet clustered or (2) expands a cluster if the point in question already belongs to one. In the figure, because four points are found within distance ϵ from point A, reaching the required minPts = 4, a cluster is formed and the points are marked red. If the number of points found within ϵ of one or more points is less than minPts, as is the case for point C, a cluster cannot be formed. However, since C can be reached by D which can successfully form a cluster by having four points within the specified distance, it is also included in the red cluster. Point N cannot reach enough other points nor can it be reached by other clusters. Points of this type of are considered outliers.

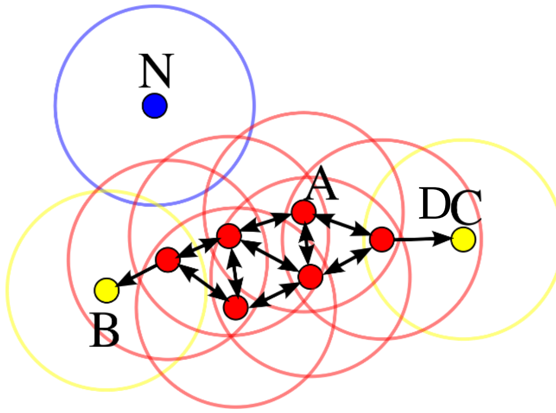

Figure 1.3 Example DBSCAN graph with ϵ Marked by Circles and minPts = 4. Adapted from [22].

The two hyperparameters must be chosen and tuned by the user. To start with, this work sets minPts = 4 as recommended by other researchers [21]. When a clustering result is unreasonable or unsatisfactory with the set value, the value of minPts is increased. Increasing minPts improves the robustness of clustering by not allowing a small number of outliers or noise to form their own cluster. The minPts values selected for different samples in this work range between 10 and 20. The other hyperparameter ϵ is found with the assistance of the k-nearest neighbour (kNN) graph. A kNN distance is the distance of a data point to its k-th nearest neighbour. Once k is decided (usually k = minPt – 1), it is a known property of each data point within the sample. A kNN graph plots the kNN distance of all data points in ascending order. Figure 1.4 is an example kNN graph with k = 25 and shows how the optimal ϵ is found based on the “shoulder” of the graph.

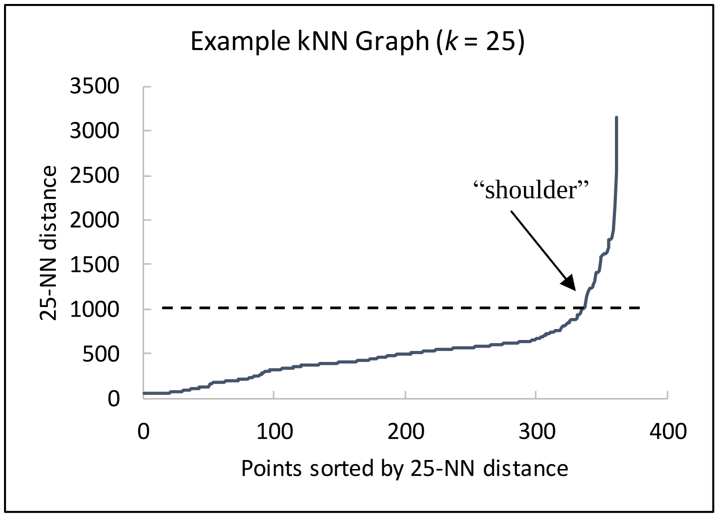

Figure 1.4 Example kNN graph with *k* = 25. Based on the position of the "shoulder", the optimal ϵ is around 1000.

The clustering analysis in this work was applied to samples of CHW-DCW data and CHW-ELE data. Before running clustering, obvious outliers are removed manually. Implementation is done with the programming language R and a package "fpc" written by Hahsler et al. [23].

2. LITERATURE REVIEW

This review will summarise the previously published articles related to statistical models for the whole-facility approach with a focus on more recent works in commercial buildings. The following sections will first summarise the literature already published in related areas. Secondly, the development of statistical modelling methods and theories for whole-building energy consumption will be discussed. Afterwards, the recent works applying these techniques will be described. Finally, an overall summary will be given and some future research directions will be suggested.

2.1. Existing Reviews

The topic of building energy models and M&V in general has received continued attention since its beginning in the 1980’s. Review articles found by this work are summarised in Table 2.1. The articles collected here include not only reviews of academic literature but also reviews of technical reports.

Table 2.1 Existing Reviews

Year	Citation	Scope
1984	Hirst [24]	Federal Residential Conservation Service
1987	Hammarsten [25]	Energy signature
1989	MacDonald and Wasserman [26]	Metered energy data analysis
2004	Abushakra et al. [27]	Load diversity factor for simulation
2006	Reddy [28]	Calibrated simulation
2012	Zhao and Magoulès [4]	Energy consumption prediction
2014	Coakley et al. [29]	Calibrated simulation
2016	Abushakra and Paulus [30]	Prediction using short-term data
2016	Harish and Kumar [31]	Methods with control applications
2017	Verhelst et al. [32]	Calibrated simulation
2017	Herrera et al. [33]	Weather data for building simulation
2017	Deb et al. [5]	Time series forecasting
2017	Molina-Solana et al. [34]	Data science for energy management
2018	Amasyali and El-Gohary [6]	Data-driven machine learning algorithms
2019	Bourdeau et al. [7]	Data-driven energy prediction

The oldest review article found was written by Hirst [24] in 1984. It reviewed the federal Residential Conservation Service (RCS) organised in 1981 and addressed the issue of whether the measures that led to savings were due to the RCS programme or whether they would have been implemented in the absence of RCS. Hammarsten [25] in 1987 reviewed several academic papers and technical reports on energy signature models. The energy signature was defined as a set of parameters estimated from a statistical analysis that describe the building's energy performance. The reviewed studies were performed in the USA, Sweden, Switzerland, and Italy. It found the energy signature model useful for energy consumption prediction using time-steps of one day or longer, but advised caution in using it for building parameter estimation. MacDonald and Wasserman [26] in 1989 reviewed analysis methods for measured energy data in then-published literature for annual total energy and energy intensity comparisons, simple linear regression and component models, multiple regression models, building simulation programmes, and dynamic thermal performance models.

Four articles are relevant to IPMVP Option D (calibrated simulation). Abushakra et al. [27] reviewed methods for generating typical load shapes required for energy simulation as part of ASHRAE Research Project 1093 with the aim of developing a library of diversity factors and schedules for typical commercial buildings. Reddy [28] reviewed published papers on building energy simulation programs as a part of ASHRAE Research Project 1051. He summarised the calibration methods used by the authors, capabilities of the tools utilised, as well as the relevant error and uncertainty issues. Coakley et al. [29] reviewed various methods utilised by practitioners to calibrate the models. They pointed out that there is no consensus reached on a standard calibration procedure and the process usually relied on user knowledge. Harish and Kumar [31] covered all methods – simulation, statistical models, and machine learning while mostly focused on methods with potential for control applications rather than M&V. Verhelst et al. [32] specifically reviewed the model selection for ongoing commissioning programs. They observed a large diversity of models and procedures used in ongoing commissioning implementations and the potential for cross-domain (namely among benchmarking, fault detection and diagnosis, and model-based control) reuse of these models is limited. On a related topic, Herrera et al. [33] reviewed the creation and handling of weather data for building simulations. Caution was especially advised for the production of weather files to express the urban micro-climate.

On the side of IPMVP Option C, whole-facility modelling approach, three recent articles reviewed machine learning methods in this area. Deb et al. [5] reviewed time series forecasting techniques used in building energy prediction. The authors noted that the infrastructure cost for a system of extensive data collection, monitoring, and control network poses a challenge for deployment of these methods. Molina-Solana et al. [34] reviewed data science used for building energy management in general including both machine learning and statistical methods. It addressed the application of techniques such as classification, regression, clustering, association rules, sequence discovery, anomaly detection, and time series analysis. Amasyali and El-Gohary [6] reviewed data-driven building energy prediction models. They pointed out the said method’s limitation in extrapolation and the lack of physical meaning for the parameters. These two review articles which focus on machine learning methods are parallel to this work which will focus on statistical models.

Apart from Deb et al. [5], three other articles emphasised the prediction performance of models. Zhao and Magoulès [4] classified modelling methods into five categories: elaborate simulation, simplified simulation, statistical methods, ANN, and SVM. The statistical methods were found to have the least complexity and the least accuracy. Abushakra and Paulus [30] reviewed efforts using short-term simulated or measured data to project long-term energy use. They found that short-term data from “swing periods” could outperform longer-term data taken from only summer or only winter months. Bourdeau et al. [7] used six categories: autoregressive models, statistical regressions, *k*-nearest neighbours, decision trees, SVM, ANN. They speculated that future data-driven modelling efforts would likely be categorised as grey-box methods, which mitigates the black-box approach’s reliance on large amounts of data and lack of generality.

2.2. Methodology Development

As the availability of building energy data kept increasing with the advancement and accessibility of metering technologies, researchers and practitioners have developed methods to extract information from measured data for various purposes. This review will summarise papers that contributed to the establishment of building energy statistical models.

Claridge [35] in 1998 comprehensively discussed methods for analysing historic temperature-dependent electricity use data including linear regression models, calibrated simulations, Fourier series models, and ANN models, among others. Among those, PRISM [36] was the first tool to determine a weather-adjusted index of building energy consumption which routinely provided the standard error. Reddy [37] performed very early studies on the predictive power of ARIMA and MLR models. Reddy et al. [38] described the residual behaviours and the reasons regression models did not work ideally. Specifically they pointed out the errors brought in from model mis-specification, namely the inclusion of irrelevant variables, exclusion of important variables, inappropriate assumption of model's linearity, and inappropriate assumption of model's order.

2.2.1. Model Formulation

When graphical data analytics was relatively new, researchers started to explore visualisation of building energy data [39-42]. Those works gave valuable insights into the functional forms for building energy models.

A building's energy use can be broken down to cooling, heating, and weather-independent electricity. The former two are strongly dependent on outdoor air dry-bulb temperature, often simply referred to as OAT. The CP model has been widely used for them as it successfully captures the thermostatic on and off behaviour of most buildings [18, 36, 43, 44]. Its most flexible form has four parameters [18, 45]. This model form was first pointed out by Schrock and Claridge [46] and the theoretical basis for its broad applicability was developed by Kissock et al. [45]. This model is set as the standard for the statistical approach in ASHRAE Guideline 14 (2014) [10]. Many research works are built upon the CP model.

The use of CP models was first made possible by relevant development in statistics [47]. At an early stage, Fels [36] used a method where the “balance-point temperature” in the degree-day method was treated as a variable in PRISM. Kissock et al. [48] and Haberl et al. [49] developed the inverse modelling toolkit for ASHRAE research project RP-1050 [50] and reported use of the grid-search algorithm as well as methods to calculate the uncertainty. For 3-parameter CP models specifically, Paulus [51] derived an explicit solution for the regression problem instead of relying on a grid search. Paulus et al. [52] also proposed a method to automatically select the appropriate model form among the CP model formulations for whole-building energy use. Monthly datasets were used to demonstrate the automated process which selects a simpler model form when higher-order parameters are not statistically significant.

The idea of using linear segments to model non-linear energy data was quickly generalized and specialized by several authors. Mathieu et al. [53] divided the hourly models into six separate and connected linear segments. This method pays more attention to capturing the “load shape” in each temperature range without emphasis on the physical significance of each parameter. Fu et al. [15] and Fu et al. [16] proposed using three linear segments to capture the heating hot water profile of buildings with constant volume reheat systems.

This model was also found applicable to other aspects of building systems. Crawford et al. [54] built segmented models for components (valves etc.) instead of the whole building. The procedure can decide how to segment the data by cutting it into halves and seeing whether it improves through iteration. Kim and Haberl [55] proposed the use of a revised 3P model for building domestic water use. It found a meaningful relationship between indoor water use and OAT. This study also noted that this is likely affected by its atypical space. It also found the inclusion of a precipitation parameter improves the outdoor part (irrigation) of water consumption while it does not improve the indoor part of it. Tereshchenko et al. [56] reported that a compound piecewise regression model outperformed other modelling techniques such as support vector machine, partial least squares regression, least absolute shrinkage, and selection operator in predicting hourly heating and domestic hot water use.

2.2.2. Independent Variables

Identifying the independent variables is not only the starting point but the most significant part of statistical modelling. Previous works usually included more variables with differentiation of system types [43, 57] while more recent papers tend to be univariate. There are three types of independent variables typically considered for most commercial building energy models: weather, occupancy, and time. For industrial facilities a variable related to production is also considered [58]. This section will discuss the former two and time effects will be discussed in its own section.

Excluding a related independent variable or including an unrelated one could bring substantial bias to the model [38]. Granderson and Price [59] compared different model techniques considering OAT, periodicity (time effects), or both. Those models that consider both were found to have improved prediction performance over others and similar among themselves. Heo and Zavala [60] used ambient temperature, supply air temperature, time stamp, occupancy (CO2 level), and humidity as independent variables in the proposed Gaussian process model. However, using such an extensive list may prove troublesome in a classic statistical model which will be demonstrated in the following sections.

Unique to industrial buildings, terms with a production variable need be considered for addition to the model (such as [61]). Kissock and Eger [62] and Abels et al. [63] applied both weather and production variables for industrial facilities. The method disaggregated savings into weather-dependent, production-dependent and independent components to gain additional insight into the nature and effectiveness of individual measures. Similar to normalised weather, Lammers et al. [64] proposed normalising industrial production with "normalized energy intensity".

2.2.2.1. Weather

Weather is obviously an influential factor for HVAC-related building energy use, but multiple weather variables play a role, such as dry-bulb temperature, humidity, wind speed, solar radiation, etc. Although it seems plausible to include as many related variables as possible, these weather variables are usually correlated and cause multicollinearity in regression models [38, 65]. Fortunately, the selection of weather variables can be aided by tools provided by statistics and physical understanding of the building system.

Several papers used PCA to help select the most influential independent parameters. Djuric and Novakovic [66] used PCA to model electricity and heating hot water use. Among cases it studied, it pointed out that while there is generally a strong correlation between heating hot water use and OAT, in some cases the operation schedule could be the strongest driving parameter. Ruch et al. [67] used PCA to re-construct multiple independent variables for electricity use modelling and prediction of supermarket energy use in central Texas, USA to compare with MLR. In that study, the independent variables were dry-bulb temperature, absolute humidity, solar radiation, and sales. PCA was helpful in providing significantly more stable models (with lower parameter standard error) and mildly superior predictions. The paper also discussed how independent variables' influence changed in the two parts of a 4-parameter CP model. However in general, researchers prefer to use OAT as the sole independent variable for building energy models [45].

Influencing variables also vary by energy type. If a facility uses district cooling and heating, its electricity use typically should not have a strong correlation with weather variables. Even though cooling energy use is strongly influenced by both dry-bulb temperature and humidity, ASHRAE Guideline 14-2014 [10] specifically recommended against using both of them in a multivariate model because their multicollinearity would jeopardise the stability of the model and hinder its predictive power. On the other hand, ignoring humidity leads to heteroscedasticity as humidity's role is present only when OAT is higher [18]. Under typical operational conditions, only when the OAT is over roughly 15°C or 60°F is it possible for the cooling coil to be under the wet condition. The high-temperature side of the model would have larger errors in an OAT-univariate model because humidity, an influencing factor, is omitted. Some researchers proposed using enthalpy as the sole independent variable representing both as a work-around [68, 69]. However, Masuda et al. [68] pointed out that while improving the high-temperature side, the low-temperature side of the model would be worsened because humidity, now a non-influencing factor, is inappropriately included. Li and Baltazar [70] and Li et al. [71] further proposed using a term called "operational effective enthalpy" to maintain model performance on both ends. This revised enthalpy variable breaks outdoor air enthalpy into its sensible and latent components and only considers the latent part when the system is expected to dehumidify.

Apart from energy consumption modelling, there is a marked lack of literature regarding statistical models for electric demand. Kim and Haberl [72] pointed out that there has not been agreement on what temperature, e.g. average or maximum, should be used for electric demand modelling. Shonder and Hughes [73] performed M&V analysis using the monthly electricity bills for a facility in Louisiana, USA. In the study they found that for the pre-retrofit period, the electric demand depended on the monthly minimum temperature during winter months and on the maximum during summer. In the post-retrofit period however, because the winter months no longer required supplemental electric heating, the whole year depended only on the monthly maximum temperature. This study is an example of the complexity of choosing an appropriate independent variable for the electric demand.

2.2.2.2. Occupancy

Occupancy is important for models of hourly or sub-hourly data, but its use is limited in the literature on statistical models. Possible reasons include: (1) its measurement difficulty, (2) its lack of necessity for daily or monthly models, and (3) the viability of modelling it as a part of operational schedule. Some studies have estimated occupancy itself, including work performed by ASHRAE MTG.OBB [74] and the International Energy Agency's Annex 66 [75]. Relevant studies were mostly used to estimate electric or domestic water demand (e.g. [76, 77]) instead of the weather-dependent energy forms addressed in this work. There are typically two indirect methods [78]. One is to use an operation schedule which reflects when the building is expected to be occupied or not. This will be discussed in Section 2.2.3 Time Effects. The other method is adopting a variable closely related to occupancy as an alternative. Abushakra and Claridge [79] found that using a linear transformation to derive the occupancy level from lighting and equipment load data yielded comparable results to directly using a detailed walk-through occupant survey. This method was adopted by Abushakra and Paulus [80] for energy consumption prediction. Brown et al. [81] used water consumption as a proxy for building occupancy in some non-domestic buildings, including schools, offices, libraries, and some commercial premises and industrial units and shops in Leicester, UK. It pointed out that no publication was found using this approach in similar areas at the time. Fu et al. [17] explored the use of the domestic cold water consumption data as proxy for occupancy level. It applied the domestic water data in a clustering analysis to represent multiple occupancy levels of a classroom building to developed different statistical models and resulted in significantly improved model fit.

2.2.3. Time Effects

Time plays a crucial role in building energy models. There are two types of time effects in building energy data. The first is autocorrelation which can be caused by the building's thermal mass effects and brought in from weather influence. The second is the periodicity typically built into the system operation schedule. Both present significant challenges to model accuracy and uncertainty estimation. These effects, however, can be substantially mitigated by using data of lower resolution. Katipamula et al. [82] suggested that daily data are preferable for retrofit savings determination, while hourly models are the best for O&M purposes, which would now be called commissioning purposes.

2.2.3.1. Lag and Autocorrelation

If not addressed properly, inflated autocorrelation may mislead the energy analyst to have undue confidence in the model's uncertainty [38, 65]. These effects are stronger in hourly data. Daily data tend to still have at least a first-order autocorrelation but it is significantly reduced [38, 83]. Tereshchenko et al. [56] reported the lag of hourly hot water data was 14 hours in one case study. Dhar et al. [84] obtained results that were favourable and comparable to other methods including ANN and wavelets using a nonlinear temperature dependent Fourier series model for hourly energy use. Ruch et al. [85] used a hybrid ordinary least squares and autoregressive model and found it provided favourable short-term projection and robust estimate of uncertainty.

2.2.3.2. Periodicity and Scheduling

Periodicity is usually caused by the diurnal or weekly schedule of system operation. It can be seen as a type of long-term autocorrelation with larger lags and may be addressed by multiple approaches.

First, a profile or statistical diversity factor can be used to describe the periodic load. As early as 1986, Reiter [86] studied archetypical diurnal and seasonal load shapes in the commercial sector for lighting, equipment, heating, and cooling end-use consumption. Claridge et al. [87] described a method to capture the profile of hourly electricity consumption accurately on a weekday or a weekend day using a diversity factor. This method was reviewed by Abushakra et al. [27] for building simulation but can also be applied to statistical models [88, 89].

A second approach is simply creating separate models for different operation conditions. Eto [90] considered different operation modes that the system changes between as multiple more-or-less steady-state conditions. Many papers have applied this principle to develop separate statistical models for different hours, days, or control regimes [67, 91-97] as well as calibrated simulations [98, 99]. This scheduling process can be performed simply by identifying a day's calendar feature (i.e. weekday, weekend, holiday, etc.) or by using statistical tools such as clustering. Mathieu et al. [53] developed multiple piecewise linear models for all 168 hours of a week to analyse temperature-dependent 15-min electricity data. Clustering techniques have been popular for separating data into different groups. Jalori and Reddy [78] and Seem [100] used different clustering algorithms to help identify which group each day belongs to and remove outliers from the data. Li et al. [101] applied Canonical Variate Analysis to cluster electricity consumption data in an application to identify abnormal utility consumption and notify the building operator in real time.

2.2.4. Uncertainty

Reddy et al. [38] and Reddy et al. [83] provided detailed discussions on sources of errors in building energy regression models. Reddy and Claridge [102] provided a framework for calculating uncertainty in energy savings from CP models but at the same time suggested against using a hard cut-off threshold for energy models but rather consider the ratio of the expected uncertainty to the total savings. There are two main challenges in the uncertainty calculation for statistical building energy models: autocorrelation and non-linearity. There has not been a consensus for an analytical solution to these challenges.

Autocorrelation has been discussed in section 2.2.3. Reddy et al. [83] suggested adopting the Cochrane-Orcutt procedure as a remedial approach to overcome the first-order serial correlation effects in daily data. Ruch et al. [65] proposed a hybrid ordinary least squares and autoregressive model for uncertainty calculation but cautioned that the autoregressive component has no use in prediction.

Many authors proposed alternative methods to overcome the non-linearity challenge. Several authors proposed simply ignoring the uncertainty of the change point itself in CP models in order to apply classic linear statistical methods [102-104]. Shonder and Im [105] and Heo and Zavala [60] proposed the use of Bayesian inference or Gaussian process modelling as a consistent framework for estimating savings and savings uncertainty in a more systematic way. Authors from Lawrence Berkeley National Laboratory used cross validation to measure uncertainty of these non-linear models [106-109]. In this method, in the baseline, part of the data was used for training and the resulting predictions were compared to the remaining data for validation. The model uncertainty is then calculated via this comparison.

2.2.5. Training Set Length

A full year of measurement is recommended to provide a complete picture of a building's energy profile for model training. Using shorter periods of training data often leads to degradation of the model performance, such as reported in [108]. However, sometimes shorter datasets must be used to project annual energy use due to constraints of time and cost. Researchers generally agree that when the selected training period has a large enough temperature span, it can provide fairly good results [82, 92, 110-112]. On the other hand, if the selected period centres on summer or winter, datasets as long as six months would still perform poorly. Abushakra and Paulus [30] in 2016 reviewed various studies that investigated how short datasets could be applied for long-term energy use as part of ASHRAE Research Project RP-1404. This research project used hourly data, daily data, and a hybrid of daily (short-term) and monthly data (long-term) to construct statistical models [80, 113, 114]. These studies suggested that the closer the average temperature of the model period was to the average temperature of the whole year (the projected period), the better their performance. Counterintuitively, having more data did not always help – if the additional data pulls the average temperature away from the annual average, it diminishes the prediction accuracy.

This practice is also subject to other sensitivities. For example, VAV systems with higher outdoor air intake [115] and heating energy use [82, 92] may be more susceptible to prediction error.

2.3. Applications

Benefiting from quick and inexpensive development, statistical models are applied to building energy analysis across all sectors: residential, commercial, and industrial. This section will review the articles that applied the statistical approaches in their applications, mainly in assessment, M&V, and data screening. Many of them overlap with those already discussed in the previous section concerning the theoretical development. A list of reviewed articles and their applications is provided in Table 2.2.

2.3.1. Assessment

Comparing whole-building energy consumption, either "horizontally" among different but similar buildings or "vertically" between one building's pre- and post-retrofit periods, is no simple task. Because the energy use is affected by many factors, it must be normalised in some way to make a reasonable comparison. The "horizontal" comparison is usually conducted for performance assessment to identify building "outliers" whose energy behaviour is different from others with similar characteristics. Perez et al. [116] studied the energy consumption data of 45 similar houses in Austin, Texas, USA and compared their CP model slopes to screen buildings that have energy savings potential by targeting houses with the largest magnitude of slopes. Burak Gunay et al. [117] applied CP model, regression tree, and ANN to hourly building energy intensity of 35 office buildings in Ottawa, Ontario, Canada. CP model parameters were compared to identify building outliers. Other multivariate methods also used wind speed, solar radiation, and an on-off schedule indicator as independent variables. The three methods were largely in agreement identifying outlier buildings, but multivariate models provided more insight into the potential cause.

2.3.2. Measurement and Verification

The “vertical” comparison is measurement and verification (M&V). The ability to appraise how much energy is saved accurately is fundamental to any energy conservation measure [118]. This task is completed by M&V which is a major application of statistical data-driven models. Turner et al. [119] summarized results of the Texas LoanSTAR (Loans to Save Taxes And Resources) program which was a $98.6 million capital retrofit program for building energy efficiency. This program heavily contributed to energy guidelines such as NEMVP [120], IPMVP [9], and ASHRAE Guideline 14 [10]. Acquisition and monitoring of building energy data is indispensable in these long-term savings projects [44, 121]. Figure 2.1 shows part of the process of an existing building commissioning (EBCx) project. During the step where the project is handed off to the owner, the energy savings must be documented following accepted M&V protocols such as IPMVP [9] and ASHRAE Guideline 14 [10].

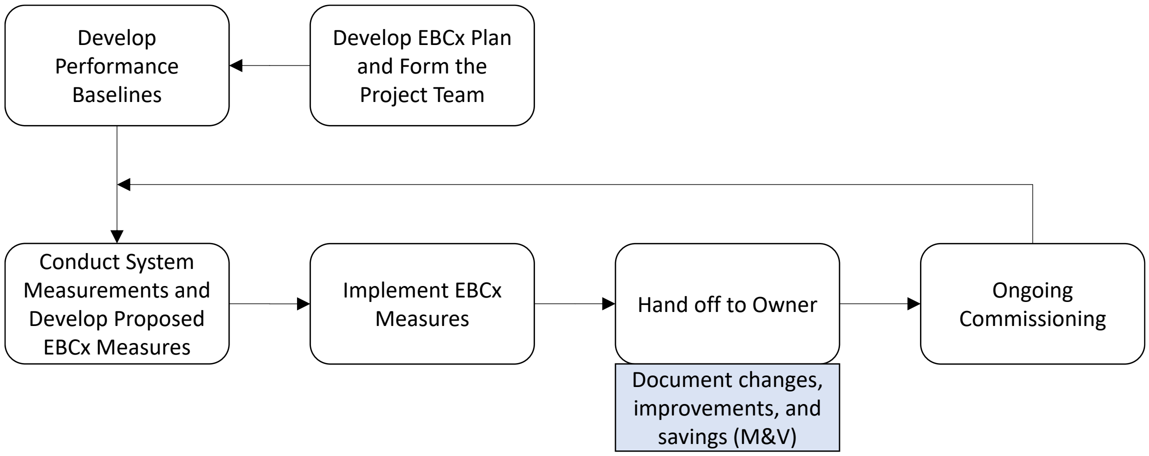

Figure 2.1 Outline of part of the EBCx process, adapted from Claridge et al. [122].

Shonder and Hughes [73] pointed out the importance of normalizing occupancy, weather, and energy price, among other factors when calculating energy cost savings in M&V. Earlier researchers used a Normalized annual Consumption (NAC) index to compare a building's energy consumption under different weather conditions [36, 123-126]. Fels and Goldberg [123] studied its use in natural gas consumption in townhouses and aggregated residential building groups. Stram and Fels [125] applied this method to residential cooling energy and combined cooling and heating energy. Ruch and Claridge [126] applied it to commercial buildings. Most of the works above prefer to use OAT as the sole independent variable, but Kissock et al. [127] cautioned that the change in indoor air temperature and internal gain may have a large impact on calculated energy savings and should be routinely monitored for savings estimation.

Shin et al. [128] reported the savings of a renovated net-zero energy building with multiple calculation methods, including unadjusted savings, CP regression model, and calibrated simulation. The calculated savings ranged from 37% to 50%. The

ASHRAE/CIBSE/USGBC Performance Measurement Protocols for Commercial Buildings [129] was applied by researchers in a field test where the CP model was used to analyse electricity use, electric demand, and natural gas consumption [72, 130, 131].

2.3.3. Data Screening

Statistical models are sometimes applied for energy data screening, either to identify faulty data points or irregular operation. These applications often require models of less variability. Shao and Claridge [132] proposed the energy balance load as a variable that is more stable than consumption of individual energy type and is independent from air-side system configuration. Relying on its stability, anomalies detected in energy balance load can more sensitively direct to problems of energy consumption data. Ji et al. [133] proposed using enthalpy in place of dry-bulb temperature to plot the energy balance to account for humidity influence. Masuda et al. [134] further developed statistical control limits of the energy balance load. As heteroscedasticity is one of the major challenges of cooling energy modelling that is also present in energy balance load, the work used the bin method to calculate the variable variance. Re-constructing the independent variable is also an approach to address heteroscedasticity and improve model performance and Masuda et al. [68] discussed outdoor air enthalpy in place of OAT to improve high temperature performance (and thus high-enthalpy) for cooling energy and energy balance models to account for the influence of humidity.

Li et al. [101] applied a combination of autoregressive model, Canonical Variate Analysis, and Linear Discriminate Analysis for FDD of a building’s electricity consumption. Some non-parametric methods were also applied in this area. Seem [135] separated data by day of week and marked outliers by statistical distribution. Brown et al. [81] identified ‘failure modes’ of building energy systems by inspecting plotted energy data visually.

2.3.4. Other Applications

Researchers also explored other applications for statistical models. Some researchers used whole-building energy data to determine building characteristics inversely [136, 137].

Table 2.2 List of Articles Discussing Applications of Statistical Approaches for Building Energy Consumption

Year		Model Type*	Building Sector	Energy Type	Energy Data			Independent Variable(s)***	Purpose or Application
					Source	Length**	Resolution		
1986	[36]	PRISM	Residential	NG	Measured	1 Y	Monthly	OAT	M&V
1986	[123]	PRISM	Residential	NG	Measured	1 to 4 Y	Monthly	OAT	M&V
1986	[124]	PRISM	Residential	ELE	Measured	7 M to 2 Y	Monthly	OAT	M&V
1988	[90]	DD, VBDD	Commercial	ELE, NG	Synthetic	12 Y	Monthly	DD	Assessment
1989	[37]	Thermal Network, MARMA, Time Derivative	Residential	ELE	Measured	290 to 562 H	Hourly, 10-min	OAT, Solar, ELE (non-cooling)	Projection
1993	[126]	NAC	Commercial	ELE, CHW, HHW	Measured	1 Y	Daily	OAT	Assessment
1994	[138]	Bayesian	Commercial	ELE, CHW, HHW	Measured	4 M	Hourly	OAT, Hum., Solar, WS	Projection
1994	[89]	Bin, CP	Commercial	ELE, CHW, HHW	Measured	11 M	Daily	OAT, TS	M&V
1996	[139]	CP	Commercial	CHW, STM	Measured	1 Y	Monthly	OAT	M&V
1996	[140]	Bayesian	Commercial	ELE, CHW, HHW	Measured	1 Y	Hourly	OAT, Solar, WS, TS, Hum.	Projection
1996	[94]	MLR	Commercial	ELE, CHW, HHW	Measured	1 Y	Hourly	OAT, Solar, WS, Hum.	Projection
1997	[111]	MLR, Fourier	Commercial	ELE, CLG, DHW, IL	Synthetic	4 W	Hourly	OAT, Hum., Enth., Solar, IL, TS	M&V
1997	[104]	VBDD, CP	Commercial	ELE, NG	Measured	1 Y	Monthly	DD, OAT	M&V
1998	[84]	Fourier	Commercial	ELE, CHW, HHW	Measured	4 to 35 M	Hourly	TS	M&V
1998	[141]	CP	Commercial	ELE, CHW, HHW	Measured	1 Y	Daily	OAT	M&V
1998	[45]	CP, NAC	Commercial	CHW, HHW	Measured	13 M	Daily	OAT	M&V
1998	[127]	CP	Residential	ELE, HTG	Measured	5 M	Daily	OAT, DT	M&V
1998	[96]	CP	Commercial	ELE, NG	Measured	2 Y	Monthly	OAT	Assessment
2004	[61]	CP	Industrial	ELE, NG	Measured	1 Y	Monthly	OAT, Production	Assessment, M&V

Table 2.2 Continued

Year		Model Type*	Building Sector	Energy Type	Energy Data			Independent Variable(s)***	Purpose or Application
					Source	Length**	Resolution		
2006	[62]	CP, VBDD	Industrial	Fuel	Measured	1 Y	Monthly	DD, OAT, Production	M&V
2006	[132]	N/A	Commercial	EB	Measured, Synthetic	1 Y	Daily	OAT	Screening
2006	[73]	VBDD, SLR	Commercial	ELE, Demand	Measured	12 - 21 M	Monthly	DD, Max or Min TDB	M&V
2007	[135]	N/A	Commercial	Elec. Power	Measured	1 M	Daily (Peak)	TS	Screening
2008	[133]	N/A	Commercial	EB	Measured, Synthetic	1 Y	Daily	Enth.	Screening
2008	[58]	CP, VBDD	Industrial	Fuel	Measured	1 Y	Daily	OAT, DD, Production	M&V
2008	[134]	CP	Commercial	EB	Measured	1 Y	Daily	OAT	Screening
2009	[110]	CP, SLR	Commercial	ELE	Measured	1 Y	Hourly	OAT	M&V
2009	[68]	N/A	Commercial	EB	Measured	1 Y	Daily	Enth.	Screening
2010	[101]	CVA	Commercial	ELE	Measured	5 M	Half-Hourly	CV, TS	Screening
2011	[63]	CP	Industrial	ELE, NG	Measured	1 Y	Monthly	OAT, Production	Assessment
2011	[64]	CP	Industrial	ELE, NG	Measured	1 Y	Monthly	OAT, Production	M&V
2011	[53]	PWLR	Commercial	Elec. Power	Measured	5 M	15-min	OAT, TS	Projection, Demand Response
2012	[60]	Gaussian	Commercial	ELE, CHW	Measured, Synthetic	1 Y, 37 D	Daily, Hourly	OAT, TS, Occ., Hum., Supply air temp., CO_2	M&V
2012	[136]	N/A	Commercial	EB	Measured	1 Y	Daily	OAT	Characteristics
2012	[105]	Bayesian, DD, CP	Residential	NG, ELE	Measured	1 Y	Monthly, Daily	DD, OAT	M&V
2013	[113]	CP	Commercial	ELE, CHW, HHW	Measured, Synthetic	1 - 12 M	Daily	OAT, IL, Hum.	Projection
2014	[55]	CP	Commercial	DHW	Measured	1 Y	Monthly	OAT, Precipitation	Assessment
2014	[114]	CP	Commercial	ELE, CHW, HHW	Measured, Synthetic	1 - 12 M	Daily	OAT	Projection

Table 2.2 Continued

Year		Model Type*	Building Sector	Energy Type	Energy Data			Independent Variable(s)***	Purpose or Application
					Source	Length**	Resolution		
2015	[78]	Clustering	Commercial	ELE	Measured, Synthetic	1 Y	Hourly	TS	Screening
2016	[80]	CP	Commercial	ELE, CHW, HHW	Measured, Synthetic	2 W	Hourly	OAT, IL, Hum.	Projection
2016	[69]	CP, DD	Commercial	CHW	Measured	1 Y	Daily	Enth., TS	Projection
2017	[116]	CP	Residential	ELE	Measured	1 Y	Daily, Hourly	OAT	Assessment
2017	[142]	CP	Commercial	ELE	Measured	1 Y	Daily	OAT	M&V
2019	[117]	CP, Regression Trees, ANN	Commercial	CLG, HTG	Measured	1 Y	Hourly	OAT, WS, Solar, Occ.	Screening
2019	[128]	CP	Commercial	ELE	Measured	1 Y	Daily	OAT	M&V

* Abbreviations exclusively used in the Energy Type column: NG, natural gas; ELE, electricity; CHW, chilled water; HHW, heating hot water; STM, steam; CLG, cooling; DHW, domestic hot water; IL, internal load; HTG, heating; EB, energy balance; Elec. power, electric power.
** Abbreviations exclusively used in the Length column: Y, year; M, month; W, week; H, hour.
*** Abbreviations exclusively used in the Independent Variable(s) column: Hum., humidity; WS, wind speed; TS, time stamp; Enth., enthalpy; IL, internal load; CV, canonical variable; Occ., occupancy; temp., temperature.

2.4. Summary

2.4.1. Literature Chronology

Publication year ranges of the reviewed articles are summarised in Figure 2.2. Only academic articles directly discussing the statistical approaches are included in this figure and articles discussing related areas (e.g., simulation, machine learning) and guideline documents etc. are excluded. It shows a clear surge of publications from 1996 to 2000. Much slower growth was seen from 2000 to 2010. 41% of the reviewed academic articles were published in 2010 or later and 58% in 2000 or later. The oldest reviewed article was published 1980 but this article was concerned with statistical theories on which CP models are based [47]. Other than this, the oldest articles directly related to building energy models were published in 1984 [24, 41].

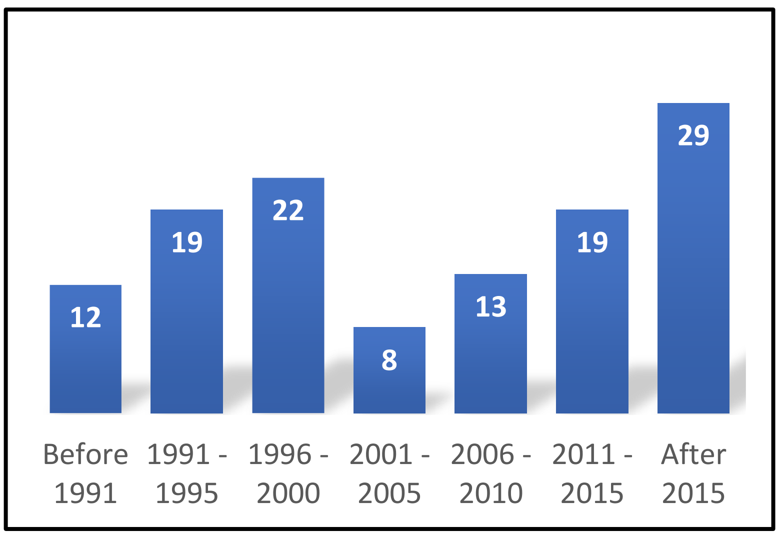

Figure 2.2 Number of articles discussing statistical approaches to analyse whole-building energy consumption data by publication year range.

2.4.2. Future Research Directions

This review has identified several areas that are not sufficiently addressed in the literature. The items below show one common theme which is that statistical models for building utilities have a heavy reliance on OAT as the sole independent variable.

- **Electric demand**: There has not been much work modelling electric demand using statistical models. One challenge pointed out in the literature regards the parameter that should be used as the driving variable [72]. Average daily or monthly temperature which is used in usual energy consumption models seems inappropriate for electric demand and some form of local maximum should be used instead. At the present time there is no agreement on the most appropriate parameter.
- **Power factor**: The power factor is another quantity that is often considered in retrofit or commissioning projects and sometimes charged on utility bills. However, no literature has been found discussing the modelling and baselining of power factor.
- **Domestic water consumption**: Building water savings are covered in ASHRAE Guideline 14 (2014) [10] as a utility consumed by buildings even though it is not a form of energy. Again, because a building's water use is not typically weather dependent (with notable exceptions in [55, 56]), at the present time there is no established solid M&V method for water savings similar to energy. Existing literature relies on occupancy to predict water use [76].

- **Occupancy variable**: Occupancy was difficult to measure directly but there has been rapid development [74, 75]. Abushakra and Claridge [79] reported that using a factor derived from measured light and equipment electricity consumption in the model yielded results comparable to using actual measured occupancy. Since then even with the rapid progress in methodology and techniques of occupancy measurement, there has been limited progress applying these for whole-building statistical models [81].
- **Specialised formulation**: Many of today's statistical models rely primarily on a change-point formulation. The four-parameter functional form has been accepted as a standard for commercial buildings but is not a universal solution to all problems. Researchers recently provided examples where the classic model form may fall short and an improved form or an alternative method was needed [15, 16, 53, 89].

3. BREAK-POINT STATISTICAL MODEL FOR BUILDING HEATING HOT WATER CONSUMPTION WITH CONSTANT VOLUME REHEAT SYSTEMS

3.1. Overview

A very early use of the change-point concept was by Fels [36], where the "break-even temperature" for use in the classic degree-day method was treated as a variable. While this was not called a CP model, this expanded degree-day model was equivalent to a three-parameter CP model. The CP method was made easier to implement thanks to the grid-search method [47, 48]. It was soon used in savings calculations for gas and electricity consumption of residential houses [123-125]. This piecewise-linear form was later upgraded to a four-parameter CP model by Ruch and Claridge [18] for better flexibility. CP models proved an accurate and cost-effective method in M&V applications and were tested for commercial buildings [91, 119, 141], industrial buildings [58, 143], and also HVAC component modelling [54].

Reddy et al. [38] pointed out that model misspecification is an important source of uncertainty. In the search for an accurate regression form for whole-building energy consumption, researchers made efforts in expressing the energy use based on physical understandings of building systems and operations. Katipamula et al. [144] and Katipamula et al. [92] found that when daily data are used, dry-bulb temperature alone can account for more than 80% of variation, whilst dew-point temperature, internal gain, solar gain, etc. are not as influential. Reddy et al. [43] provided closed-form steady-state functions for major air-side HVAC system types in order to obtain accurate regression models of monitored energy use. Chonan [140], who participated in ASHRAE's Great

Energy Shootout with minimal background in HVAC systems, used a Bayesian nonlinear regression method, but noted that even when applying a “universal” solution to the energy prediction problem, understanding of the physical system would be immensely helpful.

Earlier works seldom separate cooling and heating from the total electricity consumption, usually due to metering availability. The three energy types, electricity, cooling, and heating, present different challenges in energy modelling. While electricity usually only has a weak dependence on weather when it does not drive an on-site plant, both cooling and heating are heavily weather-dependent. Typically, cooling energy modelling poses a dilemma between neglecting humidity with risk of model misspecification and heteroscedasticity, and including a humidity term but tolerating inflated prediction uncertainty due to collinearity. Researchers proposed different approaches to address this problem, such as using enthalpy [133] or a transform of enthalpy [70, 71] as the independent variable. In the meantime, heating energy modelling mostly focuses on finding the most appropriate form to track the load shape. Kissock et al. [145] found that heating energy models may be more susceptible to bias caused by short data periods than cooling energy models.

The heating energy model of this work will use whole-building HHW data where it is supplied by a district plant. A different and improved formulation inspired by the CP model is proposed and tested in this work for buildings using a CVRH system. These systems tend to have a rather different heating energy use profile than what the CP model typically aims at. Perhaps due to the success of the classic CP model, these different profiles are often seen as results of low-quality data or malfunctioning controls. In the meantime, as CVRH systems are designed for perfect humidity control in exchange for very poor energy efficiency due to substantial simultaneous heating and cooling, buildings with such system types tend to have great savings potential [57]. It is therefore valuable to develop and improve modelling methods for these buildings.

3.2. Theory and Methodology

3.2.1. Constant Volume Reheat System Operation

A typical single-duct CVRH system has a cooling coil and a heating coil in the air-handling unit, and reheat coils in terminal boxes. The only way to dehumidify the air in a CVRH system is by cooling the air below its dew point for the moisture content to condense. A typical T_{CL} used for dehumidification is 55°F (13°C). However, this air would overcool the occupied zones under most load conditions; thus the reheat coil brings its temperature back up to a suitable level, say 65°F (19°F). The need for dehumidification is seasonal. Many buildings therefore simply rely on the outdoor air TDB to reset T_{CL} for its humidity control sequence. Since reheat load would only be present when T_{CL} is low, the setpoint of T_{CL} would influence reheat load, thus total heating and cooling load of the building.

In order to display the load shape of heating coil load (both primary heating and reheat load) of a CVRH system under these conditions, calculations were conducted with temperature bins in steady state. The calculations are based on the Simplified Energy Analysis Procedure by Knebel [146]. The results are shown in Figure 3.1 . Figure 3.1 (a) shows a control sequence where T_{CL} remains constant for all values of TDB. The resulting heating coil load has a typical change-point shape. If a simple reset is used for T_{CL} where it is set to a lower temperature when TDB is higher than 60°F (16°C), as shown in Figure 3.1 (b), heating coil load would unexpectedly increase at this point because of the activation of dehumidification. More commonly, the control sequence gradually decreases T_{CL} as TDB rises and there is a transition band for both T_{CL} and heating coil load, shown in Figure 3.1 (c). In either case of T_{CL} reset, the heating coil load shape exceeds the scope of the classic CP model and needs a different model.

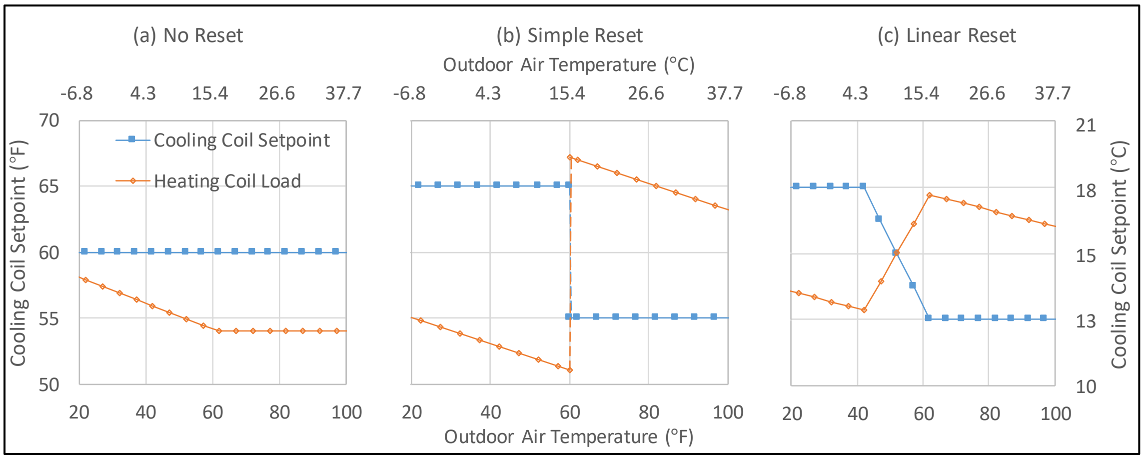

Figure 3.1 Cooling Coil Air Leaving Temperature Setpoint and Calculated Heating Coil Load as Functions of Outdoor Air Dry-Bulb Temperature of a Generic CVRH System. The Unit for Heating Coil Load is Arbitrary.

3.2.2. Break-Point Formulation

A new formulation is therefore proposed to describe the relation shown in Figure 3.1 (a) and (b). Inspired by the change-point models, this method aims to find the break-points and use straight lines for each segment. There are multiple ways to do so. Equation (3) is adopted for its clear physical interpretation. Some other equivalent formulations will be addressed in the discussion section.

$$E = \begin{cases} \beta_1 + \beta_2 T, \ T \leq T_L \\ \beta_3 + \beta_4 T, \ T \geq T_R \\ \frac{E_L - E_R}{T_L - T_R}(T - T_R) + E_R, \ T_L < T < T_R \end{cases} \quad (3)$$

where E – energy use, T – outdoor air temperature, β_i – the i-th regression parameter, T_L and T_R – left and right break-point temperatures, $E_L = \beta_1 + \beta_2 T_L$, and $E_R = \beta_1 + \beta_2 T_R$.

Demonstrated in Figure 3.2 , this formulation uses three straight lines. One is to the left of the first break-point T_L and one to the right of T_R. These two are expressed in the first two lines of the model. The third line of the model simply connects the two break-points; (T_L, E_L) and (T_R, E_R), and this part presents no new parameters. Similar to the change-point models, a grid-search method is needed for the break-point model to converge properly. An ad hoc tool is developed because there is no double grid-search tool at the ready.

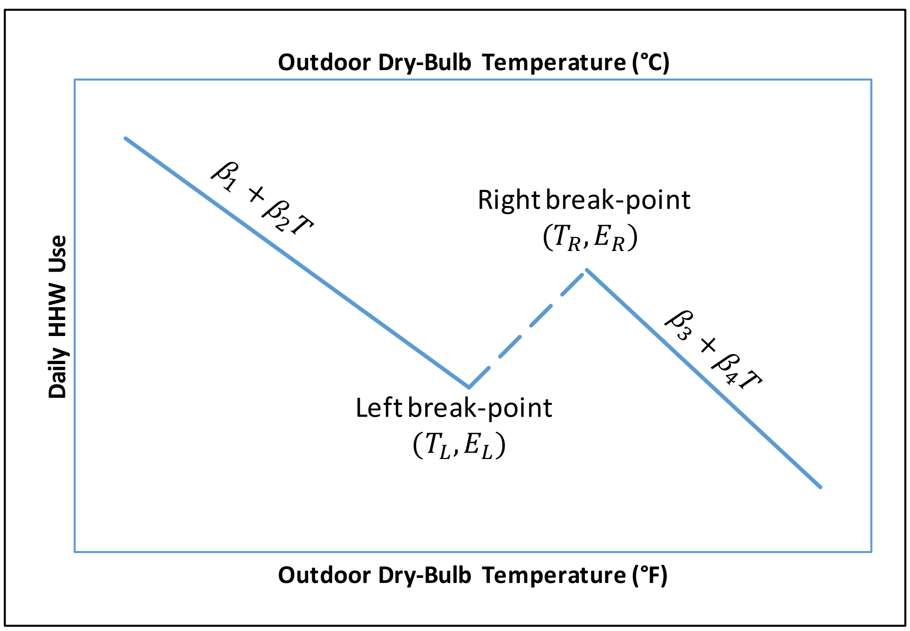

Figure 3.2 Shape of a BP Model

3.3. Case Study

Nine sets of whole-year measured heating hot water consumption data were used to test this method. These data sets come from seven on-campus buildings. Two of the buildings are used twice with different data years. The data sets were cleaned where days with a reported meter fault or irregular operation were removed. All these buildings are primarily served by constant volume reheat systems and have heating energy profiles that are not appropriate for the classic change-point models. Table 3.1 lists some basic information of these buildings. The sample covers a variety of building types: two of them have mixed use of offices and classrooms, two others have mixed use of laboratories and classrooms, three are residence halls, and the last one is a sports service building with an indoor practice facility. The majority of the buildings are aged, ranging from the oldest one built in 1932 to the second newest one in 1981. The newest building is also the only sports building which was built in 2009.

Table 3.1 Basic Information of the Sample Buildings

Building	Year Built	Area (ft^2)	(m^2)	Primary Building Use
A	1933	10,872	1,010	Offices and Classrooms
B	1932	38,313	3,543	Laboratories and Classrooms
C	1960	35,458	3,293	Laboratories and Classrooms
D	1981	53,493	4,970	Residence Hall
E	1979	59,196	5,499	Residence Hall
F	1979	53,719	4,991	Residence Hall
G	2009	58,356	5,421	Sports

The BP model is fitted to the nine data sets. Figure 3.3 shows the regression results using measured whole-building HHW consumption data over a one-year period. Both the classic CP model and the proposed BP model are plotted for comparison. Some details of the fitted models are presented in Table 3.2 and the goodness-of-fit metrics as well as the comparison between the two models are presented in Table 3.3.

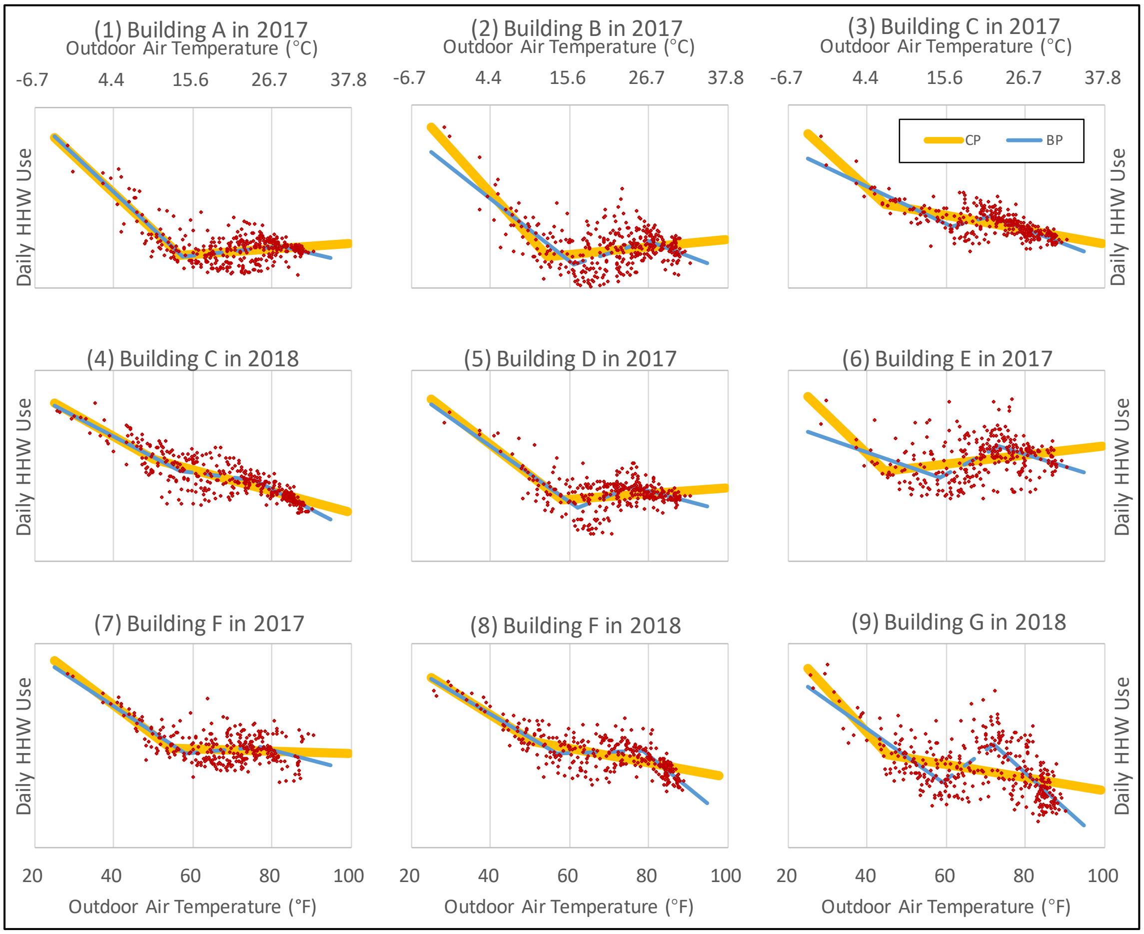

Figure 3.3 Measured Whole-Building Whole-Year HHW Use and the Comparison of CP Models and BP Models.

Table 3.2 Details of Fitted Break-Point Models (BP) of the Sample Building Datasets

Model	Building	Data Year	Sample Size (days)	T_L (°F)	T_L (°C)	T_R (°F)	T_R (°C)
(1)	A	2017	358	57.3	14.0	82.3	27.9
(2)	B	2017	358	61.0	16.1	81.0	27.2
(3)	C	2017	320	67.3	19.6	71.3	21.8

Table 3.2 Continued

Model	Building	Data Year	Sample Size (days)	T_L (°F)	T_L (°C)	T_R (°F)	T_R (°C)
(4)	C	2018	352	56.8	13.8	75.8	24.3
(5)	D	2017	338	62.0	16.7	76.0	24.4
(6)	E	2017	311	58.4	14.7	73.4	23.0
(7)	F	2017	300	58.0	14.4	77.0	25.0
(8)	F	2018	325	56.7	13.7	78.7	25.9
(9)	G	2018	362	59.1	15.0	72.1	22.3

Table 3.3 Comparison with Fitted Break-Point Models (BP) and Four-Parameter Change-Point Models (4P-CP)

Model	BP model			4P-CP model			Lack-of-fit test	
	R^2	R^2_{adj}	CV-RMSE	R^2	R^2_{adj}	CV-RMSE	F Score	p-Value
(1)	0.567	0.561	28.8%	0.549	0.548	29.2%	4.5	0.011
(2)	0.405	0.396	39.6%	0.371	0.369	40.8%	10.4	<0.001
(3)	0.613	0.606	14.3%	0.556	0.554	15.4%	21.7	<0.001
(4)	0.751	0.747	13.1%	0.721	0.720	13.8%	15.3	<0.001
(5)	0.452	0.443	17.8%	0.381	0.379	19.0%	20.2	<0.001
(6)	0.196	0.182	17.5%	0.103	0.100	18.4%	14.9	<0.001
(7)	0.439	0.430	12.6%	0.418	0.416	12.8%	5.1	0.007
(8)	0.683	0.678	12.0%	0.609	0.608	13.3%	31.4	<0.001
(9)	0.501	0.494	24.7%	0.316	0.315	28.7%	49.9	<0.001

3.4. Discussion

All cases shown in Figure 3.3 demonstrate that the proposed BP formulation successfully captures the shape of the HHW consumption and appears to have better fit than the classic four-parameter CP model.

3.4.1. Break Points

The key challenge of using the BP model is finding the two BP’s. Similar to CP models, the grid-search method is recommended for its non-uniformity. The two BP’s are found by exhaustive search and the other four parameters are calculated by least-squares simple linear regression. Figure 3.4 is a flowchart demonstrating this procedure. The model with the smallest sum of squares error found in the grid is then chosen.

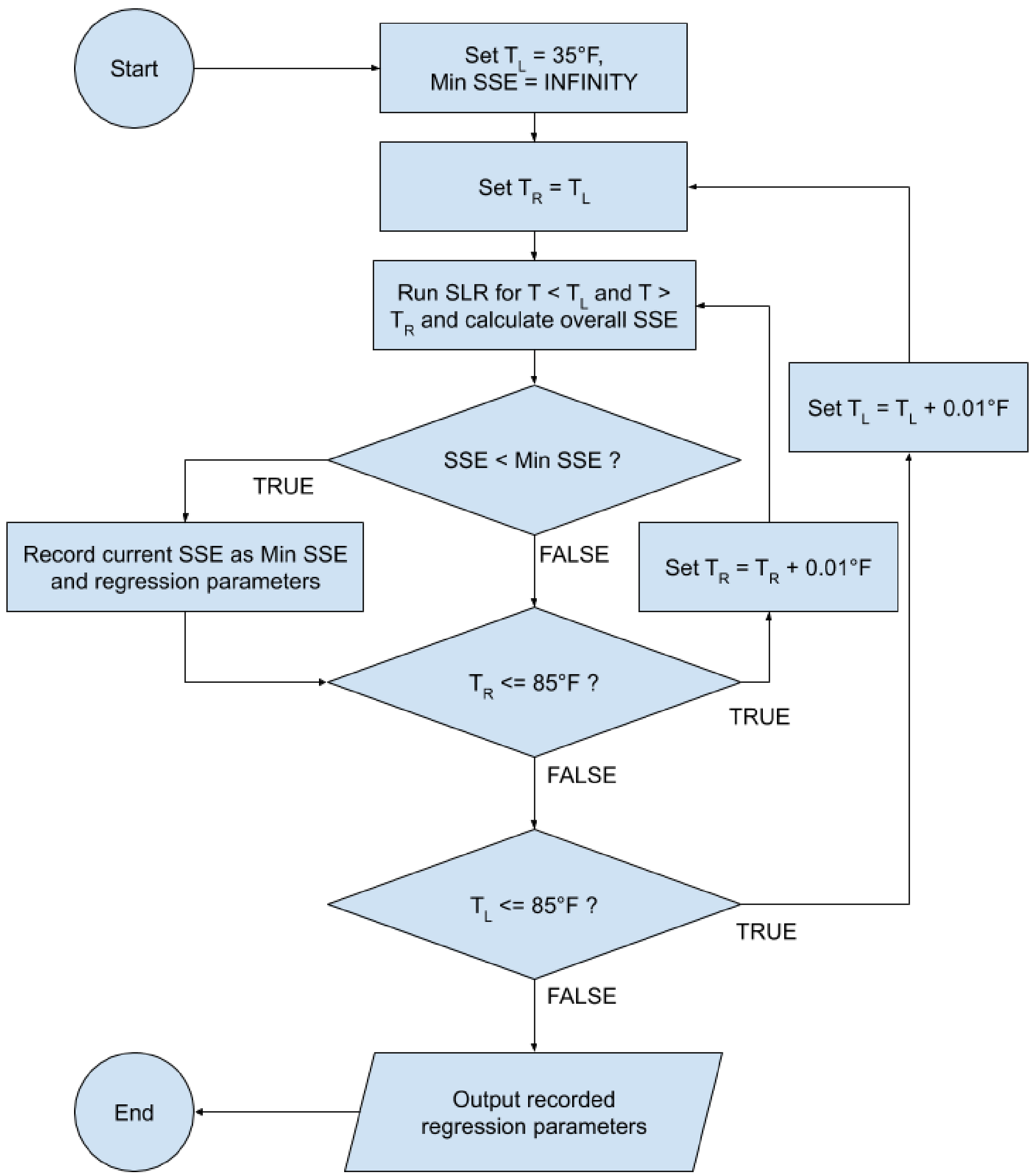

Figure 3.4 Flowchart for the Break-Point Searching Algorithm Between 35°F and 85°F with Step of 0.01°F.

3.4.2. Residual Behaviour

The plots in Figure 3.3 show that the two model formulations tend to have some degree of agreement on the left side where outdoor air temperature is low. However, the

BP model, being able to take one more turn, outperforms the CP model in tracking the decreasing HHW use when the temperature is high on the right side.

Figure 3.5 provides a closer look at this improvement by examining the residual behaviour. The residuals are divided into three sections by the break-points. The slopes of the CP model residuals are calculated in each section and plotted on the plots as well. The same residual slopes of the BP models for these sections are zero by definition. The slopes of the CP model residuals are all close to zero in the left section. However, all nine CP models show a clear slope in the right-most section on the residual plots. This is true for six of the nine models in the middle section as well. These regional slopes are problematic if the model needs to be used for extrapolation at the higher-temperature end. It is eliminated by BP model which has a more accurate linear component to the right.

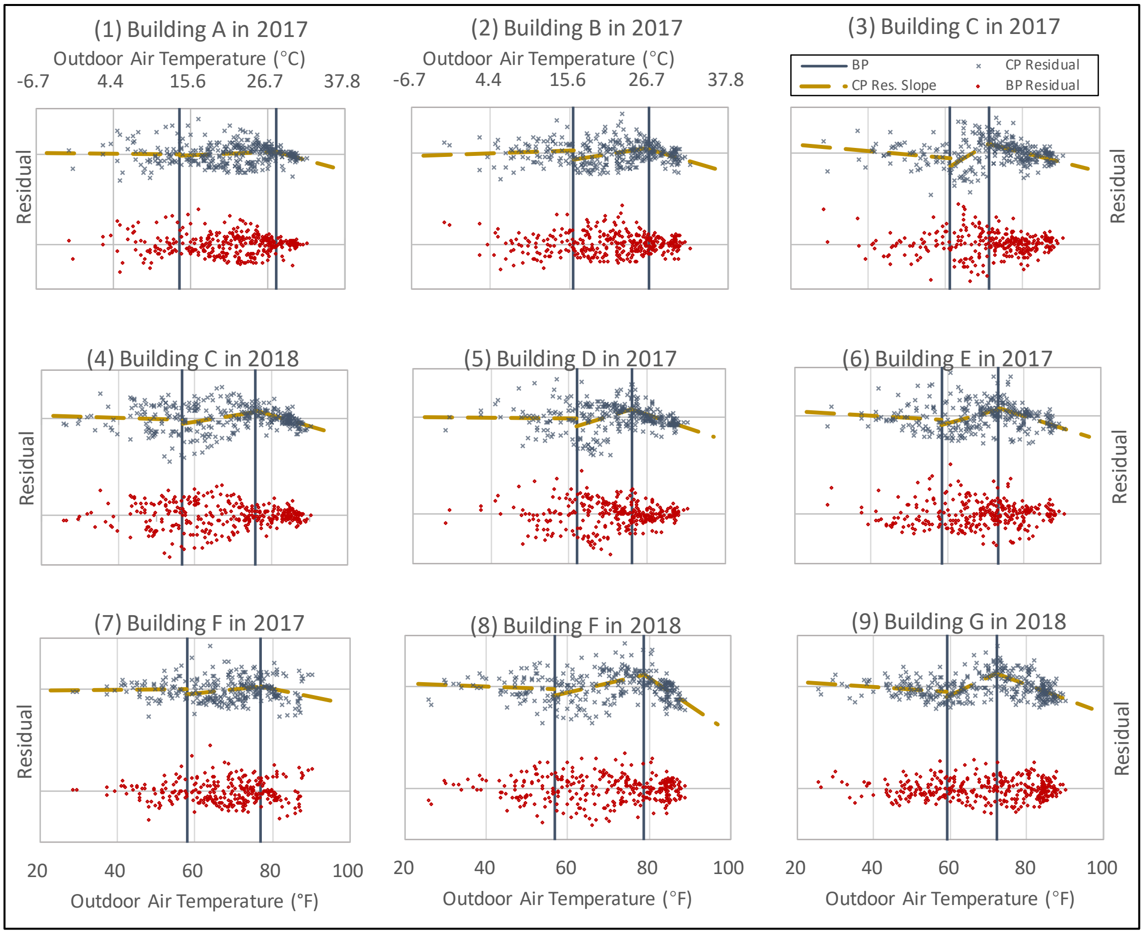

Figure 3.5 Regression Residual of Each CP and BP Model Plotted Side by Side, with One Horizontal Zero-Line Crossing Through Each and Local Slopes of CP Model Residuals Divided by Break-Points.

3.4.3. Improvement of Fit

Table 3.3 shows that all BP models produced more favourable R^2 and CV-RMSE values compared to 4P-CP models. The R^2 of BP models ranges from 0.19 to 0.78 and is largely affected by the quality of the data sets themselves. Also, because the data used in this study do not have a strong overall trend, the flatness of data also results in poorer R^2 values. CV-RMSE of BP models varies from 12% to 42%, where seven out of nine are lower than 25% which is the ASHRAE Guideline 14 recommended threshold [10].

Notably, the CP version of model (9) does not reach this threshold while its BP model is improved to fall within the Guideline 14 threshold. Adjusted R^2 values are also provided in Table 3.3 to provide slightly more accurate comparison between the models as the two models have different numbers of parameters, namely that the BP model has six and the CP model has four.

Although all BP models yielded better results judged by the metrics in Table 3.3, some of them seem marginal. Lack-of-fit tests with 95% confidence level ($\alpha = 0.05$) were therefore conducted to provide additional clarity. The test treats the BP model, transformed to an equivalent formulation of Equation 4, as an improvement to the CP model with two additional parameters. Compared to Equation 1, Equation 4 adds a term for β_4 and splits T_{CP} to T_L and T_R, effectively adding two more parameters. This formulation was not used for regression, because in the case of simple reset (See models (3) and (4) shown in Table 3.2 and Figure 3.3), the middle section would have infinite slope and the regression parameter β_4 cannot converge.

$$E = \beta_1 + \beta_4 T + \beta_2 (T - T_L)^- + \beta_3 (T - T_R)^+ \tag{4}$$

The test results of all nine model pairs yielded p-values less than $\alpha = 0.05$, indicating that the BP models have significantly improved the regression results from the CP models [147].

3.5. Summary

The proposed break-point formulation has been tested on measured whole-building heating hot water energy use data and proved suitable for the sample of buildings tested. The models successfully captured the operational characteristics of the studied cases. Although the average change in CV-RMSE of 1.2 percentage points appears marginal, the improvement of each sample was statistically significant with lack-of-fit tests' *p*-values far smaller than 0.05. This improvement is not only related to regression metrics, but also in the implications for enhanced reliability in diagnostics and energy consumption prediction due to the improvement residual behaviour.

4. COMPARING THE DOMESTIC WATER AND ELECTRICITY CONSUMPTION AS OCCUPANCY PROXY IN COMMERCIAL BUILDING COOLING ENERGY MODELLING

4.1. Overview

Occupants play a crucial role in buildings' energy consumption, but this role may have been underestimated in building energy models. Traditionally, occupants are considered as part of the internal gain in load calculation [148]. Authors have pointed out that these considerations have been much simplified in energy codes or standard modelling practices, while occupant behaviour and interactions with the building (such as opening and closing windows, changing the thermostat settings) have significant influence on the building's energy consumption [149, 150]. How to model occupant behaviour has received much research attention in recent years (such as [74, 75, 151]). Many papers go into great detail with the measurement and projection of occupancy per se so that simulation programmes can provide better results with better occupancy input. This work, however, will explore how occupancy information can be used in the faster and more affordable whole-building energy statistical models.

Before occupancy sensing technologies became as advanced and affordable as they are today, researchers and practitioners typically resorted to two approaches to indirectly consider occupancy in their modelling efforts [78]. The first approach is to consider the change of occupancy as the change of schedule, i.e., assuming the change of occupancy is predictable by date and time. ASHRAE Research Project 1093 reviewed methods of creating such typical load diversity factors [27]. This method is very popular in calibrated simulation and is an embedded functionality in many simulation software packages. The second approach, which this work adopts, is finding an alternative variable that can represent the occupancy. Abushakra and Claridge [79] reported that using a linear transformation of the lighting and equipment load data in the model yielded comparable results to using real occupancy data. Brown et al. [81] used water consumption as a proxy to help identify building system failure and reported that no literature was found using water data to represent occupancy.

This work will explore the use of two types of utilities thought to useful as an occupancy proxy in whole-building analysis under two different approaches. The two utilities to be tested are DCW and non-HVAC ELE. They will both be tested to function as the occupancy proxy using the following two approaches:

1. The linear approach: directly add the occupancy proxy as an additional linear term to the regression formula.
2. The clustering approach: apply clustering analysis to the occupancy proxy to separate data into multiple groups, then build separate regression models for each separated subset of data.

A case study building is used to test these methods against some more conventional methods, such as separating the energy data into weekdays and weekends. The case study demonstrates that the combination of DCW and the clustering approach resulted in significantly better model performance in different scenarios, while other proxy-approach combinations performed inconsistently. This study is valuable because it provides a fast and easy-to-implement process to improve the energy consumption models of buildings with modellable occupancy patterns where conventional modelling practice may have neglected occupancy. Although occupancy sensor technologies have been rapidly emerging and becoming more and more affordable, numerous buildings with savings potential may not have these sensors. A proxy method will improve their pre-retrofit baseline models and thus improve the accuracy of their savings calculation in measurement and verification.

4.2. Methodology

This work will test and compare two measured building utilities, DCW and ELE, for their suitability for use as a building occupancy proxy under two approaches - linear and clustering.

4.2.1. The Proxy Variable

There is no landscape irrigation consumption under the metered whole-building DCW data used in this work. This makes the dataset more directly related to the presence of the occupants in the building who are the only consumers of the water. Otherwise, special treatment is needed to filter out the irrigation portion to improve the model fit [55].

The building used in the case study uses district cooling and heating and its building-level ELE consumption does include heating or cooling energy. Its ELE data will therefore only exhibit very limited dependence on OAT caused by smaller temperature-dependent loads such as pumps and fans in its secondary HVAC systems.

4.2.2. Approach 1: Additional Linear Term

The first approach is adding the occupancy as an additional linear term to Equation (1) introduced in Section 1.3.1. Equation (1) can be expressed in an equivalent form as Equation (5).

$$y = \beta_0 + \beta_1(x - \beta_3)^- + \beta_2(x - \beta_3)^+ + v \tag{5}$$

where *v* represents the residual.

In the regression analysis, ideally, the residual series *v* should be random and no more information can be extracted from it. If not, a meaningful regression model can be established in the form of Equation (6).

$$\hat{v} = \beta_4 z + c \tag{6}$$

where *z* represents a new independent variable.

"Meaningful" regarding Equation (6) means that the regression coefficient β_4 is statistically significant. Merging Equation (6) into Equation (5) and combining the intercept terms β_0 and c, one has the CP model with an added linear term as Equation (7). This method is intuitive and straightforward, but it is worth noting that bringing in one more regression coefficient is bound to reduce the regression error of the model. However, this reduction may not always be "meaningful". This issue will be addressed in section 4.2.4.

$$\hat{y} = \beta_0 + \beta_1(x - \beta_3)^- + \beta_2(x - \beta_3)^+ + \beta_4 z \tag{7}$$

where *z* represents the occupancy proxy variable.

4.2.3. Approach 2: Clustering of the Occupancy Proxy

The second approach to treat the proxy variable is clustering. This is based on the assumption that several relatively distinct occupancy levels can be identified, and the proxy variable can be clustered into multiple groups to reflect those occupancy levels. A separate regression model will then be built on the data subset of each cluster. This is similar to separating data into weekdays and weekends. However, the clustering analysis will be more flexible and able to identify more than two different levels of occupancy.

After the data are clustered into multiple groups using the clustering technique introduced in Section 1.3.3, they are fitted to a model series as expressed by Equation (2) in Section 1.3.2.

4.2.4. Goodness-of-Fit Metrics

Once the data are separated into multiple clusters, it is no longer appropriate to compare the single model with the original whole dataset with the subset models. A formulation like Equation (2) then provides a reference for the calculation of regression error across the whole dataset. This will further allow comparison between the model series with the whole-dataset model or with another model series that separates the same dataset in a different way, or across approaches.

Two statistical goodness-of-fit metrics are used in this work to measure and compare the models. They are the adjusted R^2, defined in Equation (A.3), and the coefficient of variation of the root mean square error (CV-RMSE), defined in Equation (A.4). A model is a good fit to the data when its adjusted R^2 is large and close to one and its CV-RMSE small and close to zero [152]. These two metrics are chosen because they penalise the models for having too many regression parameters p. This is important because when more regression parameters are included in the model, the regression error (in terms of the sum of squares error) would naturally be smaller, causing the ordinary R^2 to inflate [152]. As this work will compare a number of data treatment methods with wildly different numbers of parameters (see Table 4.3), the number of parameters should be normalised in the comparison.

4.3. Case Study

A university classroom building in Texas is used to test the proposed methods. It was built in 1991 with a floor area of 237,147 ft^2 (22,032 m^2). Its space usage per the university record is shown in Table 4.1 . This building is suitable for this study in the sense that it has no research lab space. Such spaces usually require large amounts of outdoor air which tends to drown out other influences on the energy consumption.

Table 4.1 Space Types of the Case Study Building

Space Type	Classroom	Office	Research Lab	Office Service*
Area Percentage	14.8%	24.8%	0.0%	6.8%
Space Type	Conference Room	Clerical Support	Special Use**	Non-Assignable***
Area Percentage	2.0%	1.2%	20.2%	30.2%

* This category includes file rooms, closets, record rooms, office supply rooms, etc.
** This category includes work rooms, study rooms, computer rooms, office storage, etc.
*** This category includes mechanical rooms, restrooms, locker rooms, public circulation areas, lobbies, foyers, elevator stairways, etc.

Utility meters are already installed on this building for billing purposes. For this study, its CHW consumption, DCW consumption, and ELE consumption data are used. These data are measured at the whole-building level with hourly or sub-hourly frequency. Together with the utility data, concurrent OAT data from the nearest airport are also collected. All these data are aggregated to the daily resolution for this study. There is no direct occupancy measurement in this building. Four full-year datasets from 2017 to 2020 were analysed in this study to test and compare the proposed methods. Results for two of these four years are chosen to be inspected in greater detail: a “normal” year of 2019 and the year 2020 when the COVID-19 pandemic forced the building to close and the occupancy level drastically decreased.

4.3.1. Clustering Results

The results of the clustering for each individual year at each proxy variable are shown in Figure 4.1. It is notable that three clusters were consistently formed when DCW was used as the proxy. For the ELE proxy, mostly there were only two clusters. Due to the campus closure from mid-March of 2020 caused by the COVID-19 pandemic, the clustering results of 2020 shown in Figure 4.1 (d) and (h) are visibly different from other years. This will be discussed in section 4.4.2.

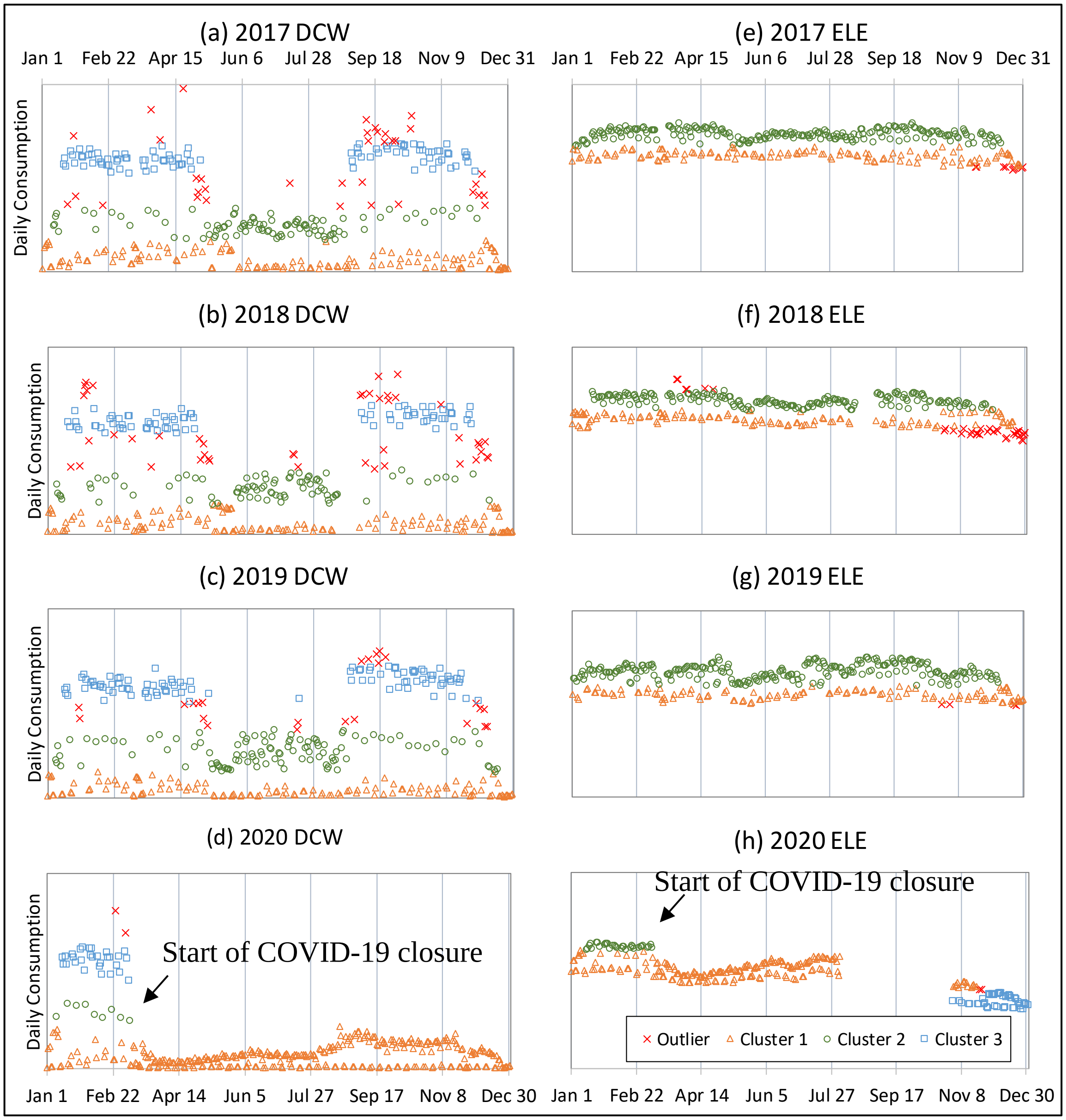

Figure 4.1 Clustering Results in Time Series by Year and by Proxy Type. Numbers and Units of the *y*-Axes are Hidden Because They are Arbitrary. The *x*-Axes are Shared by Subplots on the Same Columns and the *y*-Axes are Shared by Those on the Same Rows. The Arrows Point Out the Spring Break on Each Subplot.

In order to validate the clustering results, the university academic calendar is used to determine whether patterns can be found. The days are classified in four categories in the academic calendar:

- Academic Days (Aca.): During the fall or spring semesters, the classes are in session. both the students and the faculty and staff are present in the building.
- Summer (Smr.): During the summer short semesters, fewer classes are in session. Fewer students are present in the building, but all the faculty and staff still are.
- Break (Brk.): During the breaks between semesters (both long and short), no class is in session. No students are present in the building, but all the faculty and staff still are.
- Holiday (Hol.): During the school holidays, neither students nor the faculty and staff (except those essential) are present in the building.

The results of clustering are arranged according to academic calendar as shown in the Table 4.2 . From 2017 through 2019, the clustering results demonstrate consistent correlation with the academic calendar. In the DCW-CHW clustering, the clusters of the high DCW consumption mostly correspond to Aca. M-H. The cluster of intermediate DCW consumption predominantly falls under Aca. Fr and Smr. WD. In addition, this intermediate cluster takes some portions of Brk. WD. The lowest DCW cluster matches the Aca. WE, Smr. WE, Brk. WE, and Hol. Together with the intermediate DCW consumption cluster, this clusters also take up some portions of Brk. WD. The clusters with high ELE consumption tend to correspond to Aca. WD, Smr. WD, and some fractions of the Brk. periods. The low ELE cluster usually fit the Aca. WE, Smr. WE, Brk. WE, Hol., and the remaining portion of Brk. WD. The mixed results for Brk. WD using both proxies could be caused by the fact that the building usage is less consistent under this day-type. Overall, the high and intermediate DCW clusters combined approximately matches the high ELE cluster. That is to say, the ELE clustering failed to further distinguish differences within the Aca. WD. This implies that clustering with DCW can provide a higher resolution occupancy proxy than ELE.

Table 4.2 Clustering Results Arranged per Academic Calendar

Proxy Type	**DCW**								**ELE**							
Calendar	**Aca.**			**Smr.**		**Brk.**		**Hol.**	**Aca.**			**Smr.**		**Brk.**		**Hol.**
Category*	M-H	Fr	WE	WD	WE	WD	WE		M-H	Fr	WE	WD	WE	WD	WE	
Year 2017								(Outlier 36 / Total 365)**								(9 / 365)
Cluster 1	7	2	59	1	7	2	59	1	8	4	64	1	20	7	14	9
Cluster 2	5	19	5	48	5	19	5	48	122	27	0	49	0	31	0	0
Cluster 3	94	1	0	0	94	1	0	0	-	-	-	-	-	-	-	-
Year 2018								(44 / 342)								(34 / 346)
Cluster 1	5	2	57	0	20	19	14	17	11	8	48	1	20	16	12	7
Cluster 2	2	17	5	48	0	14	0	0	102	20	0	50	0	17	0	0
Cluster 3	78	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
Year 2019								(24 / 361)								(4 / 365)
Cluster 1	5	0	60	0	22	5	16	18	3	0	55	0	19	7	16	16
Cluster 2	4	26	1	50	0	32	0	0	124	29	5	53	3	31	0	0
Cluster 3	97	0	0	1	0	0	0	0	-	-	-	-	-	-	-	-
Year 2020								(2 / 365)								(2 / 277)
Cluster 1	0	0	52	0	19	0	17	12	44	16	32	50	20	19	10	6
Cluster 2	36	11	10	49	1	25	5	0	31	2	0	0	0	1	0	0
Cluster 3	36	4	2	0	0	1	0	0	0	0	11	0	0	1	4	11

* Under each calendar category, the days are also divided to weekdays (WD) and weekend days (WE). For the academic days alone, the weekdays are further divided to Monday through Thursday (M-H) and Friday (Fr).
** The outlier is a point that is excluded by the clustering algorithm. The total sample size is the number of data points used in the clustering process for this calendar year. For each calendar year, there may be points that are excluded due to data missing or reported meter fault.

4.3.2. Regression Results

Regression analysis was conducted with the following methods:

a. None: OAT-univariate model with no data separation. This method is only a reference.

b. WD/WE: OAT-univariate model with data separated to weekdays and weekend days, two subsets total. This method is a widely-used strategy.

c. DOW: OAT-univariate model with data separated to seven days of a week, seven subsets total. This method is a saturated strategy.

d. DCW-C: OAT-univariate model with data separated according to DCW clusters.

e. DCW-L: Multivariate model with OAT and DCW with no data separation.

f. ELE-C: OAT-univariate model with data separated according to ELE clusters.

g. ELE-L: Multivariate model with OAT and ELE with no data separation.

All regression results for each individual year with each method are listed in Table 4.3 . Because the clustering process flags certain data points as outliers, to maintain a consistent dataset, methods (a) through (e) use the same sample as the DCW clustering results while methods (f) and (g) use the ELE clustering results.

The method (d) DCW-C outperformed all other methods in three out of four data years. In years 2018 and 2019, it is the clear favourite by substantial margins. It is especially meaningful that this setting defeated the saturated method (c) as well. A detailed comparison of the proposed methods (d) through (g) will be given in section

4.4.1 Comparing Different Proxies and Approaches. The only exception of method (d)'s superiority is the year 2020, which will be discussed in detail in section 4.4.2 Influence of the COVID-19 Pandemic.

Table 4.3 Regression Results of Each Individual Year with Different Methods. The Highest R^2_{adj} and the Lowest CV-RMSE of Each Individual Year are Highlighted in Bold.

Setting	N	p*	R^2_{adj}	CV-RMSE	N	p	R^2_{adj}	CV-RMSE
	Year 2017				**Year 2018**			
(a) None	329	4	0.672	22.2%	298	4	0.804	19.6%
(b) WD/WE	329	8	0.896	12.5%	298	8	0.900	14.0%
(c) DOW	329	28	0.915	11.3%	298	28	0.908	13.5%
(d) DCW-C	329	12	**0.917**	**11.2%**	298	12	**0.942**	**10.7%**
(e) DCW-L	329	5	0.842	15.4%	298	5	0.911	13.2%
(f) ELE-C	356	8	0.909	11.4%	312	8	0.909	12.6%
(g) ELE-L	356	5	0.844	14.9%	312	5	0.904	13.0%
	Year 2019				**Year 2020**			
(a) None	337	4	0.717	27.8%	363	4	0.819	19.6%
(b) WD/WE	337	8	0.906	16.0%	363	8	**0.893**	**15.1%**
(c) DOW	337	28	0.913	15.4%	363	28	0.890	15.3%
(d) DCW-C	337	12	**0.949**	**11.8%**	363	12	0.833	18.8%
(e) DCW-L	337	5	0.869	18.9%	363	5	0.866	16.8%
(f) ELE-C	361	8	0.902	16.3%	275	12	0.874	16.9%
(g) ELE-L	361	5	0.914	15.2%	275	5	0.887	16.0%

* p is the number of parameters of the whole model series, i.e., the parameters of all component models combined.

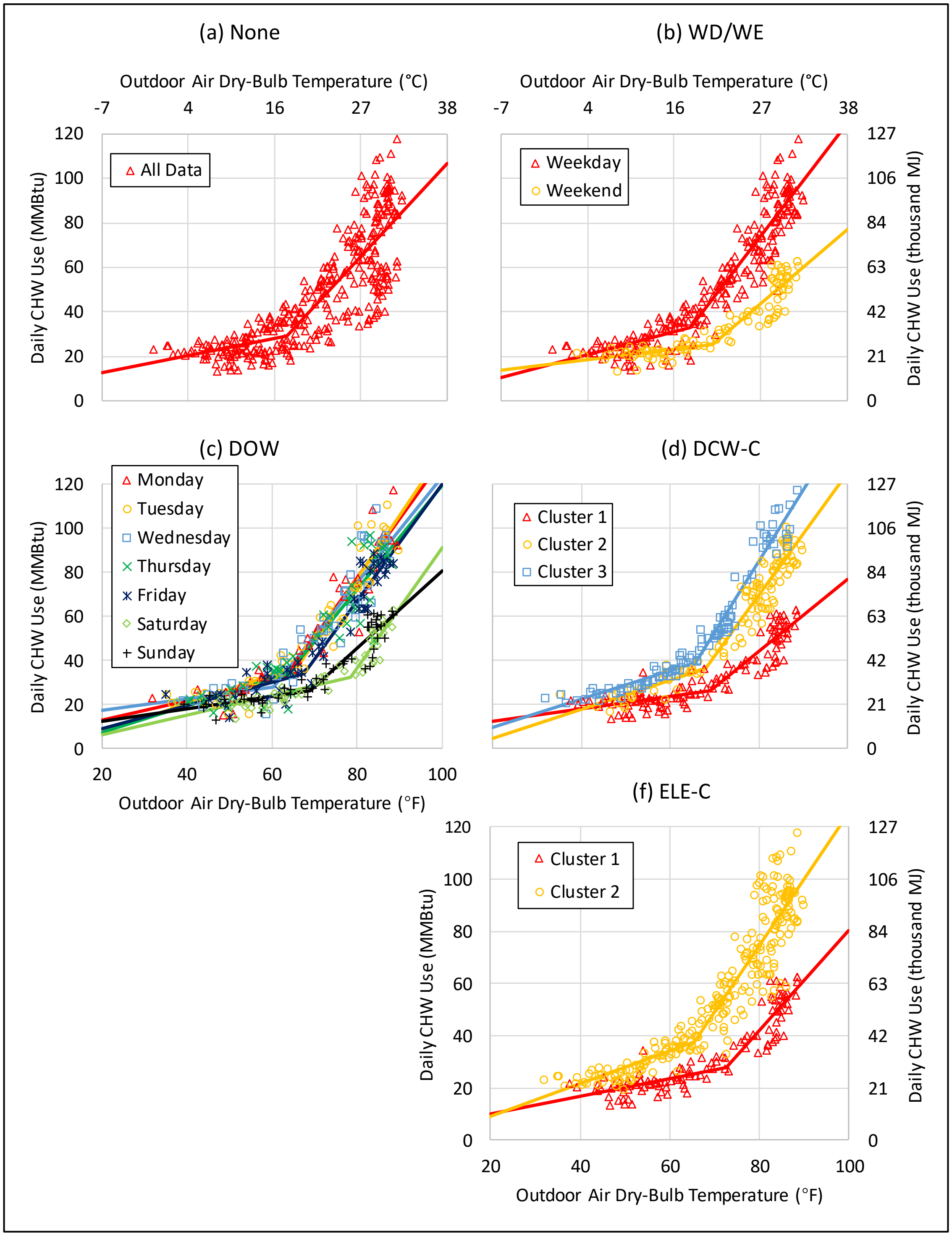

Figure 4.2 Regression Results by Method for the Year 2019. The *x*-Axes are Shared by Subplots on the Same Columns and the *y*-Axes are Shared by Those on the Same Rows.

Figure 4.2 shows the regression results for the year of 2019 for methods (a) through (d) and (f). The linear regression methods (e) and (g) are not shown here because they cannot be meaningfully displayed and compared on a 2-D scatter plot. Upon initially inspecting Figure 4.2 (a) and (b), it is immediately clear that the dataset requires some level of data separation to produce meaningful regression results. Figure 4.2 (d) shows that the DCW clustering mainly managed to further separate the weekday subset shown in Figure 4.2 (b) to two levels. In the meantime, the ELE clustering shown in Figure 4.2 (f) failed to make this improvement.

4.4. Discussion

4.4.1. Comparing Different Proxies and Approaches

This section will set the year 2020 aside, which has its own section. Under the proposed methods with two proxies, DCW and ELE, and two approaches, clustering and linear addition, the combination of DCW and clustering (method d) always outperformed other methods. The margins of CV-RMSE are substantial for 2018 and 2019 even over the saturated method (c), being 2.8 percentage points and 3.6 percentage points respectively. As pointed out in section 4.3.2 Regression Results, the substantial improvement is attributed to the DCW clustering's success in breaking the weekday subset into two well-separated layers.

The performance of all the other three combinations (methods e, f, and g) are not consistent, sometimes worse than the standard weekday-and-weekend method (b). Also for the clustering approach, ELE as a proxy (method f) outperformed the standard method (b) in two of the three years and outperformed the saturated method (c) in only one year. In the meantime, since the ELE clustering only produced two data subsets and the clustering results are in good agreement with a simple weekday and weekend separation (see Table 4.2), this method is therefore not a valuable alternative to the standard method (b).

As for the linear approach, with the only exception of ELE in 2019, all of them produced worse results compared to the clustering approach using the same proxy. The DCW proxy always performed better in the clustering approach (method d) than the linear approach (method e). The reason is thought to be DCW's "layered" data distribution resulting in poorer correlation with CHW. In the meantime, the ELE proxy, whose data distribution is more continuous, always performed better in the linear approach (method g) than DCW (method e) did.

4.4.2. Influence of the COVID-19 Pandemic

The year 2020 is very different in the sense that the campus was closed in mid-March due to the COVID-19 pandemic. Seen in Figure 4.1 the campus was closed in mid-March 2020. In a "normal" year, influenced by the change of occupancy, both the DCW consumption and the ELE consumption would only briefly drop to a lower spring break level for a week before they return to normal. In 2020, however, the occupants did not return to the building because all classes and the vast majority of work moved online after the break. This, in fact, caused an even further decrease of both DCW and ELE. Both proxies slowly increased again as the building was gradually opened for partial capacity, but they remained much lower than the pre-pandemic level.

This break of pattern caused the year of 2020 to have very different results. The standard weekday-and-weekend method (b) performed the best for this year. The two methods of linear addition (e and g) outperformed the two methods with clustering (d and f). However, the regression results of all seven settings had CV-RMSE above 15% which is somewhat poor. In conclusion, although the data of this study has been organised according to the calendar year for convenience, common protocols regarding the choice of an energy baseline such as avoiding sudden and irregular changes should still be followed, because the proposed methods are unable to accommodate them automatically.

4.4.3. Influence of the Clustering Hyperparameters

Hyperparameters mean a set of parameters that are supposed to be decided by users before implementing the algorithm. Thus, the performance of machine learning depends upon the selection of hyperparameters. Clustering is categorised as unsupervised machine learning. While its counterpart, supervised learning, is performed with input that consists of a pair of dependent and independent variables, unsupervised learning is carried out only with independent variables. This lack of dependent variable, which could serve as a supervisor, makes it cumbersome to precisely measure the accuracy and goodness of a result and thus to verify the appropriateness of hyperparameter selection because there is no reference available to calculate the error from. Fortunately, the plots on which this work performs clustering are two-dimensional, so the validity of clustering result can be easily checked with visual inspection.

Although the use of the methodology of this work would be straightforward on an individual case level, it may be a more challenging task to fully standardise and automate the whole process due to the inherent nature of unsupervised machine learning: a lack of means to quantify metrics of error. Nonetheless, it would be more feasible to develop an *ad hoc* procedure that suggests a range of hyperparameters in which a reasonable clustering result is expected.

4.5. Summary

This section has explored the use of an occupancy proxy to aid the statistical models for whole-building cooling energy consumption. Two types of occupancy proxies were tried: DCW and ELE. Their potential has been tested under two approaches. The first approach is simply to add the proxy term as a linear addition. The second is to conduct a clustering analysis on the proxy and then build separate statistical models for each data subset.

Overall, using the DCW as the occupancy proxy in a clustering approach improved the model fit substantially. From the standard weekday-and-weekend data separation, the DCW clustering managed to further identify two well-separated groups in the weekday subset of data. This on average improved the CV-RMSE by 2.9 percentage points for the typical years when the average CV-RMSE values of weekday-and-weekend models were already as favourable as 14.2% on average. The other methods, using the DCW as an additional linear term or using ELE in both approaches failed to improve the model consistently.

It is also important to note the significance that the clustering algorithm produced results with physical meanings in Section 4.3.1. Normally, machine learning methods such as ANN are purely black-box and their parameters cannot be meaningfully interpreted. However, the results shown in this work show the potential of closing this gap.

The case study building was closed in mid-March 2020 due to the COVID-19 pandemic. This resulted in a sudden change of occupancy pattern and hence an abrupt drop in both occupancy proxies. Using the data from this year revealed that the proposed methods are not able to accommodate these events. Care must be taken when the analyst is selecting a suitable baseline period.

5. AUTOMATED AND COMPREHENSIVE DAY-TYPE GROUPING IN BUILDING ENERGY BASELINE MODELS WITH A SERIES OF STATISTICAL INFERENCES

5.1. Overview

Buildings sometimes have operation schedules that lead to varying but predictable fluctuations in their energy consumption levels. In such cases multiple diurnal or weekly levels are reflected in the energy consumption data and, if not addressed properly, they can substantially limit the performance of the energy model. Djuric and Novakovic [66] applied PCA to electricity and heating hot water consumption data and reported that, in some cases, the operation schedule turned out to be a stronger driving variable than the outdoor-air temperature. As the field renews its attention to the accuracy of non-linear energy models which often lack standard uncertainty calculation procedures [106-109], an energy analyst must find a best data separation strategy to improve the energy consumption model fit. If the energy model will be used to assist savings calculation or consumption projection, among many other applications, it must be a good fit to the original data available.

Prior works have extensively discussed several approaches to address a building's operational idiosyncrasies when the system switches modes. These approaches include applying a series of load diversity factors which is reviewed by Abushakra et al. [27] and developing a system type-specific model formulation when the reason for system mode shifts is understood (e.g., Fu et al. [16]). For statistical models, the data were often separated into subsets. The data separation procedure reported in many articles, like this work, is based on the time or the calendar. The approaches of all these articles are exploratory – for example, the data were separated according to evident calendar characteristics, such as days of a week, and then separate models were built to verify if the separation was effective. Since the data separation technique per se was not their focus, it is entirely possible that a slightly more complex data separation would have led to significant model improvement.

This work proposes a completely different and novel approach. After a group of elementary day-types are defined, the proposed method examines all data separation possibilities and suggests a best one that balances model accuracy and complexity. This approach is more comprehensive because it does not stop at the first effective possibility like an exploratory method often would, and thus can lead to the discovery of unobvious data separation that would significantly improve the model fit. The methodology and results of this work are valuable because the practice of data separation not only improves the model quality while saving the energy analysts' time, but also helps extract more information from the measured data which could assist control strategies that use forecasting and anomaly detection [100, 126]. Although this study is conducted in the context of statistical modelling and regression analysis of commercial building cooling energy consumption, it is likely to be applicable across other energy modelling approaches including calibrated simulation.

Eto [90] argued that as the building system changes among its several operational modes, they can all be modelled as separate steady-state conditions. This has been the central hypothesis widely applied by researchers who separated the measured data in building energy statistical models and regression analysis [53, 69, 91-94, 96, 128] as well as in calibrated simulations (e.g. [98]). Shin and Do [69] separated the measured energy consumption data of their case study buildings to weekdays and weekends which significantly improved the model fit. It is worth noting that there is no obvious weekday/weekend separation when one simply visually inspects the scatter plot. This highlights the necessity of investigating data separation even when it is not clear to see. Rabl and Rialhe [91] reported that by giving occupied and unoccupied days two different sets of regression parameters, the energy signature models were significantly improved. The standard error of a third of the improved models was only 90% of their counterparts. Katipamula et al. [92] reported that when using hourly energy data for energy models, further differentiating the hours of a weekday and a weekend day improved model fit. Wang and Claridge [93] reported in their case study that cooling energy regression models could achieve less than 0.5% annual prediction error when the AHUs operate 24 hours per day. However, if the building's AHUs have nighttime shut down, the error can be as high as 6.1%. This error was reduced to 0.6% by developing two separate models for the two control regimes. In another study by Katipamula [94], the authors reported that separating summer and winter data may yield better results. In a study analysing energy use in K-12 public schools, Landman and Haberl [96] found that even for monthly data analysis, separating the data into a school year group and a summer group can provide more accurate modelling. In a study by Mathieu et al. [53] hourly building electricity use data were divided for all 168 hours of a week to provide detailed insight into the building's energy behaviour. Apart from using the calendar, researchers also used other methods to separate data, such as a temperature cut-off point [67], the Tukey-

Kramer multiple comparison procedure [153], clustering analysis [17, 78, 100], canonical variate analysis [101].

To facilitate the energy analyst's decision-making, this work proposes an automated and comprehensive procedure to explore data separation possibilities that steers a middle course between the two extremes found in the literature. One extreme relies on simple and convenient methods based on the time or calendar or using a single cut-off point [53, 67, 91-94, 96]. The other relies on advanced techniques that many who work in the field may find inconvenient to implement or require extensive human intervention in the process [17, 78, 101, 153].

5.2. Methodology

In one of the simplest scenarios, the data only need to be separated into weekday and weekend groups. Yet some buildings may have more complex operation schedules, such as different Saturdays and Sundays, or Fridays with reduced loads. One apparent option is to build a separate model for each day of the week. More finely-divided data produce smaller regression error in general. However, an over-divided data set reduces the size of each sub-sample. The smallest sub-sample size of a model that completely separates all days of a week is 52, half that of a weekday/weekend model which is 104. Having too many components in a model also reduces its auditability. A simpler model, on the other hand, may still be preferable if its performance, albeit poorer, is found acceptable. It is then desirable to find a balance point between accuracy and simplicity. This can be an arduous task to undertake manually. This work thus proposes an automated procedure to find an optimised combination of pre-defined elementary day-types to develop a statistical model of building energy consumption.

5.2.1. Definitions

In order to describe building data modelling methods better, this sub-section defines some key concepts. The concept of the day-type (DT) is a way to mark a data point based on its calendar feature. An elementary day-type (EDT) is the finest element into which the data can be divided. This work uses two EDT schemes: day-of-week (DOW) and the academic calendar (ACal). The DOW scheme with all seven days of a week is self-explanatory. The ACal scheme is extracted from the academic calendar of the institution where the case study buildings are located. The ACal scheme is explained in Table 5.1 . Both schemes have seven elements.

Table 5.1 Elementary Day-Types Under the Academic Calendar Scheme

Academic Calendar (ACal)		Remarks	Expected Occupants	
			Students	Faculty & Staff
Academic Weekday	(AD)	In a fall or spring semester	Yes	Yes
Academic Weekend	(AE)			
Break Weekday	(BD)	Not an academic day, summer day, or holiday	No	Yes
Break Weekend	(BE)			
Summer Weekday	(SD)	In a summer semester	Some	Most
Summer Weekend	(SE)			
Holiday	(HL)	University-designated holidays	No	No

A *separation* is a way or plan to divide data into a number of groups. It can also be seen as a combination of EDT. For example, a weekday/weekend separation can be expressed in two ways: (a) it divides the whole dataset into two groups, or (b) it combines EDT's of Monday through Friday into one group and Saturday and Sunday into the second group. Each group contains a subset of the sample and provides a *model* *component* through separate regressions. All these model components together comprise a *model series* that again contains the whole dataset. A model series that directly uses all the EDT's, i.e. no EDT's are combined, is called a *saturated model*. A saturated model is the most finely-divided and most complex. It cannot be separated further. The definition of the model series allows a uniform framework for comparing models with different numbers of regression parameters which will be discussed in Section 5.2.3.

5.2.2. Exhaustive Search of Reasonable Data Divisions

It appears impractical to exhaust all possible data separations and combinations. However, with the aim of finding a better statistical model, those data separations that are obviously unreasonable can be easily eliminated. Shown in Figure 5.1 is an example of a meaningful and a meaningless EDT combination. The data in the figures were separated into three EDT's: A, B, and C. Figure 5.1 (a) shows a meaningful division where adjacent A and B were combined while Figure 5.1 (b) shows a meaningless separation where A and C cross over with B between them. It is evident that such a separation is bound to produce worse regression results. Upon discovering that the EDT's should be ordered based on their energy consumption levels, the problem at hand is reduced to whether each adjacent pair should be divided from each other. It is easy to see that for n ordered EDT's, there are 2^{n-1} possibilities. In this work, using the seven EDT's under both schemes only presents 64 meaningful divisions each, falling within a manageable range of computation intensity. However, this also reveals that this exhaustive search has exponential time complexity and may lack expandability.

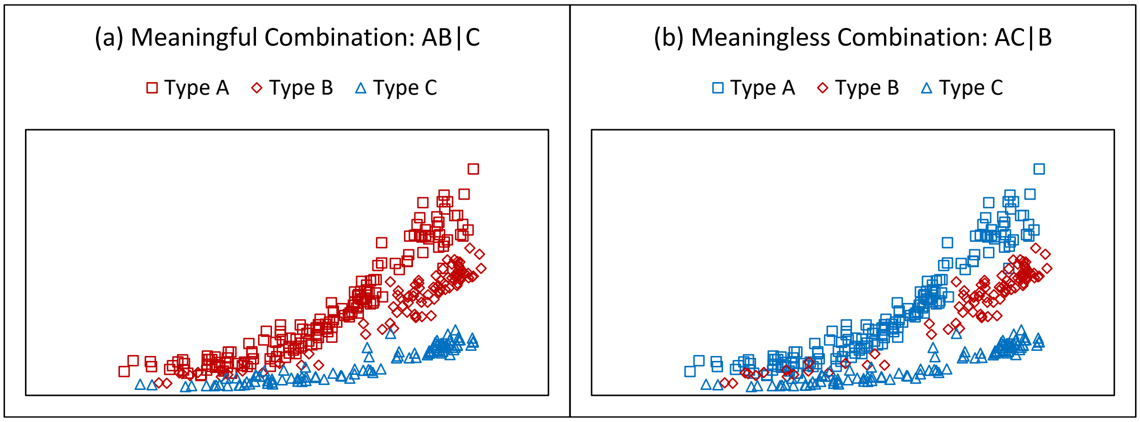

Figure 5.1 Example of (a) A Meaningful and (b) A Meaningless Data Separation of Three Elementary Day-Types.

5.2.3. Comparing the Model Series

Two factors are considered when different model series by different divisions are compared: the regression error and model complexity. More complex models tend to provide less regression error. However, if this improvement is marginal, the simpler model may be preferred.

The regression error is directly measured by the *SSE* defined in Equation (A.1). The complexity of the model is measured by the number of parameters *p*. It is the sum of the number of parameters required by each model component. This work will use the 4P-CP model as expressed in Equation (1) and the model series is expressed in Equation (2). Also important to note is that, as pointed out in section 1.3.1, even though the 4P-CP model is comprised of two linear components, its overall form is non-linear [19]. The change-point temperature cannot be found through ordinary least-squares but by a minimisation algorithm (such as a grid search). This will have implications for the behaviour of some goodness-of-fit metrics which will be discussed in Section 5.4.1.

5.2.4. Unbalanced Data Treatment

Between the two EDT schemes, the DOW scheme has virtually balanced sub-datasets, i.e. the sub-datasets representing each EDT are approximately the same size. It is also evident that all seven EDT sub-datasets of the DOW scheme should have virtually the same distribution of OAT (see Figure 5.2 (a)). However, this is not the case for the ACal scheme. For the year 2019, which is used in the case study, the sub-sample sizes of each EDT of both schemes are shown in Table 5.2 . The difference between the biggest sub-sample (academic weekday) with 156 days and the smallest sub-sample (break weekend) with only 16 days is substantial. These smaller EDT's with smaller sub-samples may produce less robust model components and should be preferred to be merged with other EDT's to form larger groups.

Table 5.2 Sub-Sample Sizes for Each Elementary Day-Type Under Both Schemes for the Year 2019

Day-of-Week (DOW)		Sample Size (day)	Academic Calendar (ACal)		Sample Size (day)
Monday	(Mo)	52	Academic Weekday	(AD)	156
Tuesday	(Tu)	53	Academic Weekend	(AE)	62
Wednesday	(We)	52	Break Weekday	(BD)	38
Thursday	(Th)	52	Break Weekend	(BE)	16
Friday	(Fr)	52	Summer Weekday	(SD)	53
Saturday	(Sa)	52	Summer Weekend	(SE)	22
Sunday	(Su)	52	Holiday	(HL)	18

Another inconvenient feature of the ACal scheme is that some EDT's would have wildly different OAT distributions, which is the sole independent variable for the model in this work. Shown in Figure 5.2 , the two summer sub-samples both heavily and narrowly occupy the high-temperature side. A model component that contains only summer components is likely to result in a fit that is not robust for extrapolation.

Therefore, instead of using the CP model as expressed in Equation (1), this work will force such a model component to a simple linear regression. From the analyst's perspective, these "summer" model components should be preferably merged with other EDT's to cover a wider OAT span in order to ensure model robustness.

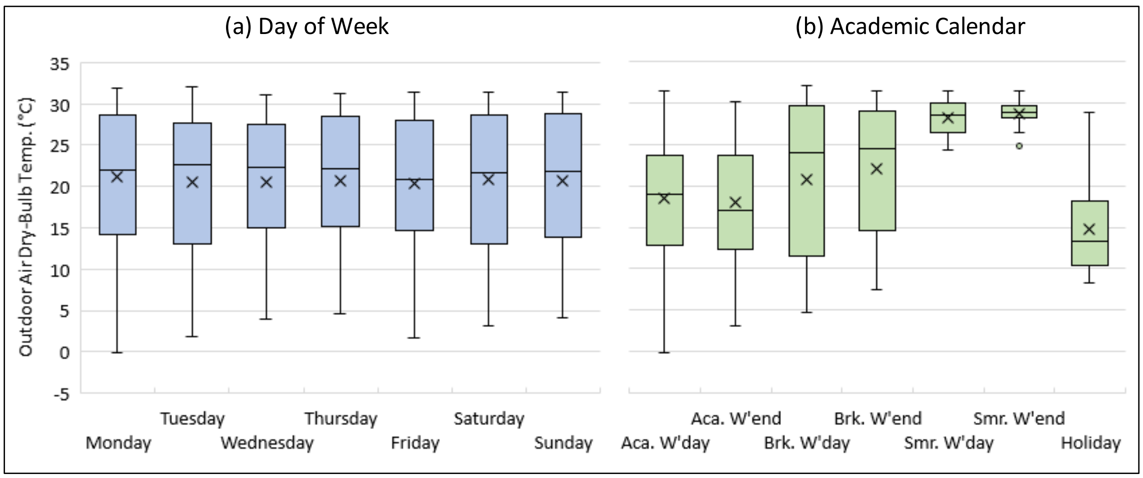

Figure 5.2 Daily Average Outdoor Air Dry-Bulb Temperature Distributions of Each Elementary Day-Type Under Both Schemes. The Middle Line Inside the Box Represents the Sample Average and the Cross Represents the Median. Both Plots Share the Same *y*-Axis.

5.2.5. Lack-of-Fit *F*-Test

The lack-of-fit *F*-test is part of the ANOVA. This section will demonstrate key parts of this procedure. More details can be found in many standard statistics textbooks (such as [147, 152]).

The LOF test compares two model formulations called the full model and the reduced model. The reduced model is nested within the full model, meaning one can obtain the reduced model by setting some of the regression parameters in the full model to specific values. On the other hand, the full model can be seen as the reduced model improved with additional regression parameters. Additional parameters will always result in a smaller *SSE*, but they also add to the complexity of the model. If the improvement is marginal, one may find the simpler model preferable. The LOF test provides an inference whether the reduced model is roughly as good as the full model, and thus helps answer the question:

- Does the simplicity of the reduced model outweigh its fit degradation? Or equivalently,
- Does the fit improvement brought by the additional parameters in the full model outweigh its increased complexity?

The LOF procedure involves computing the *F*-statistic or *F*-score and the corresponding *p*-value. The *F*-statistic is calculated by Equation (8).

$$F_{Lack} = \frac{SS_{Lack}/DF_{Lack}}{SSE_{Full}/DF_{Full}} \tag{8}$$

where

Lack sum of squares $SS_{\text{Lack}} = SSE_{\text{Reduced}} - SSE_{\text{Full}}$,

Lack degree of freedom $DF_{\text{Lack}} = p_{\text{Full}} - p_{\text{Reduced}}$,

Full error degree of freedom $DF_{\text{Full}} = n_{\text{Full}} - p_{\text{Full}}$,

With the *F*-statistic, one can look up the *p*-value with an $F_{\alpha,DF_{\text{Lack}},DF_{\text{Full}}}$ distribution, where α is the significance level, usually 0.05. The search for p-value can be easily done with any software with basic statistics packages. If the p-value is smaller than α, it rejects the reduced model for being too parsimonious even though it is simpler. Otherwise, the full model is not significantly better and the additional parameters are not justified. In this study this leads to the simpler model being favoured even though it has slightly larger error.

This work will use the saturated model as the full model and all other model series as the reduced model to conduct LOF tests. Because multiple LOF tests are conducted at the same time, in order to maintain a family-wise significance level of 0.05, the Bonferroni's correction is applied [154, 155]. For each individual test, their own significance level is divided by the total number of tests. With 64 models in total, there will be at most 63 LOF tests for each dataset, in which case the significance level of each single test is set to $0.05 \div 63 = 0.00079$. In rare cases in this study, the saturated model may not be the model with the lowest *SSE*. In such cases the model series with an even smaller *SSE* would be excluded in the tests. This phenomenon will be discussed in Section 5.4.1.

The procedure for selecting the best model series is as follows:

1. Conduct lack-of-fit *F*-tests for all reduced models against the saturated model. Eliminate all model series rejected in the tests. If all reduced models are eliminated, the saturated model is selected.
2. Among the model series not rejected, find the ones with the smallest number of parameters, p.
3. Finally, select the one with the smallest *SSE*.

5.2.6. Other Goodness-of-Fit Metrics

Four other classic goodness-of-fit metrics will also be used in this work for comparison. They are the *SSE*, R^2, CV-RMSE, and R^2_{adj}. Their definitions are given in

Equations A.1 through A.4 in Appendix A [147, 152]. Models with better performance tend to have a smaller *SSE* and CV-RMSE and a larger R^2 and R^2_{adj}. However, these four metrics consider slightly different aspects of model fit and do not always agree with each other. Among them, the *SSE* is rarely directly used by analysts. It is included because it is directly used in the LOF tests. Two of these metrics, CV-RMSE and R^2_{adj}, consider the number of parameters used in the model against them and they will be watched closely and be compared with the LOF test.

5.3. Case Study

5.3.1. Overview

CHW consumption data are used in this study to validate the proposed method. These data are measured at the whole-building level with hourly or sub-hourly frequency by an energy meter already installed on site for billing. The data are all aggregated to daily resolution for this study. A data quality-control audit has already been performed where outliers were removed when a meter fault is observed or suspected. Together with the energy-use data are local weather data from the nearest airport. The outdoor air dry-bulb temperature will be used as the sole independent variable in the regression analysis. In total, data from 76 institutional buildings on a university campus in Texas, USA are used in this work. The complete results for all case-study buildings are provided in Appendix B. Some of these buildings will be discussed in greater detail. Table 5.3 and Figure 5.3 show the complexity of the model series (in terms of number of model components or model parameters) favoured by each metric. For all cases under both schemes:

- *SSE* and R^2 always agreed.
- CV-RMSE and R^2_{adj} always agreed and never favoured a model series more complex than those chosen by *SSE* and R^2.
- LOF never favoured a model series more complex than those chosen by any other metrics.

Under the DOW scheme, the *SSE* and R^2 chose the most complex model series (with seven components) in all but one case. In 44.7% of the cases, CV-RMSE and R^2_{adj} preferred the model series with only two groups. Among them many are weekday-weekend separations. The proposed method, however, mostly suggested that data separation was not necessary and preferred the no-separation model series (with only one component) in 61.8% of the cases.

Under the ACal scheme where the data treatment is unbalanced, the results were similar but with some notable differences. The *SSE* and R^2 continued to recommend the most complex model series, albeit in only 85.5% of the cases. The CV-RMSE and R^2_{adj} also tended to support some more complex model series and in no case did they prefer a model with no separation. And last, LOF still favoured the simplest model series compared to all others, but its peak moved to two components and in one case recommended a relatively complex model with five components.

Across all cases where the LOF tests did suggest the data be separated, the CV-RMSE improvement from their no-separation counterparts was 10.9 percentage points on average under the DOW scheme (29 or 38.2% of the cases), 6.1 percentage points under the ACal scheme (61 or 80.3% of the cases), and 7.6 percentage points overall.

Table 5.3 Counts and Percentages of the Model Series and Their Number of Components Favoured by Each Metric Under Both Data Separation Schemes.

Number of model components	Day-of-Week						Academic Calendar					
	SSE or R^2		CV-RMSE or R^2_{adj}		LOF		SSE or R^2		CV-RMSE or R^2_{adj}		LOF	
1	0	0.0%	8	10.5%	**47**	**61.8%**	0	0.0%	0	0.0%	15	19.7%
2	0	0.0%	**34**	**44.7%**	20	26.3%	0	0.0%	1	1.3%	**29**	**38.2%**
3	0	0.0%	14	18.4%	5	6.6%	0	0.0%	12	15.8%	20	26.3%
4	0	0.0%	8	10.5%	4	5.3%	0	0.0%	9	11.8%	11	14.5%
5	0	0.0%	5	6.6%	0	0.0%	0	0.0%	**26**	**34.2%**	1	1.3%
6	1	1.3%	6	7.9%	0	0.0%	11	14.5%	20	26.3%	0	0.0%
7	**75**	**98.7%**	1	1.3%	0	0.0%	**65**	**85.5%**	8	10.5%	0	0.0%

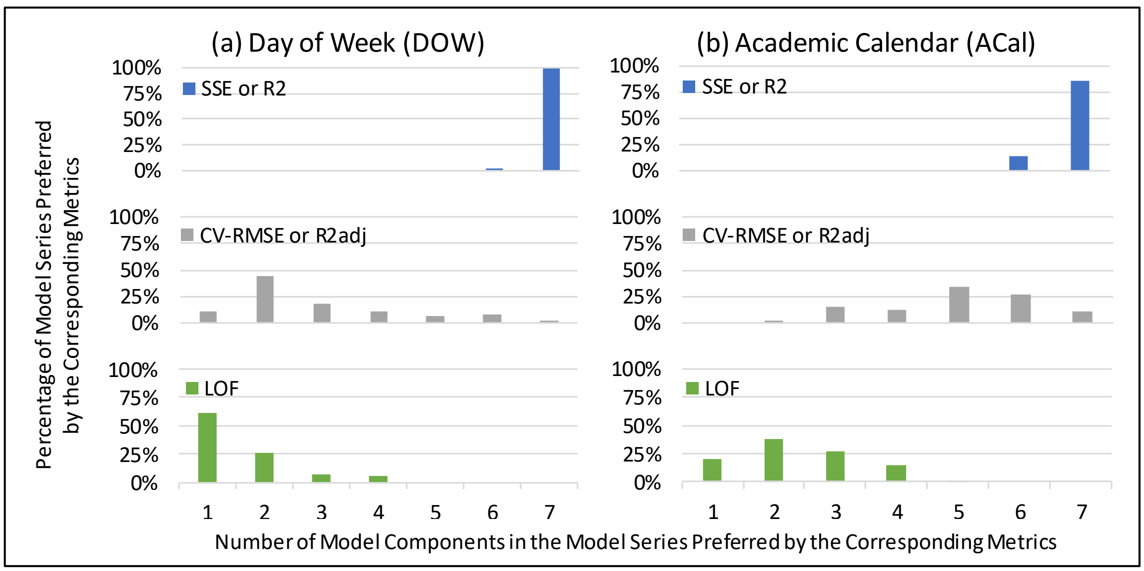

Figure 5.3 Percentages of the Model Series and Their Number of Components Favoured by Each Metric Chosen by Both Data Separation Schemes.

5.3.2. Selected Cases

5.3.2.1. No Separation

Out of all 76 case buildings, the proposed method favoured no separation in 47 cases (61.8%) under the DOW scheme. Figure 5.4 and Table 5.4 show the results for building #25 which falls within this category. This building mainly has research labs and offices. Figure 5.4 displays two different data separations for this building under DOW scheme: (a) no separation favoured and (b) weekday and weekend. In Figure 5.4 (a) it appears that all data are in one compact group. However, as is shown in Table 5.4 , CV-RMSE and R^2_{adj} are able to identify an extremely slight difference between the weekdays and weekend days which would improve CV-RMSE by 0.1 percentage point and R^2_{adj} by 0.001, favouring model series (b). Finally, the LOF test did not find this marginal improvement significant, favouring the utmost simplicity of a model with no data separation.

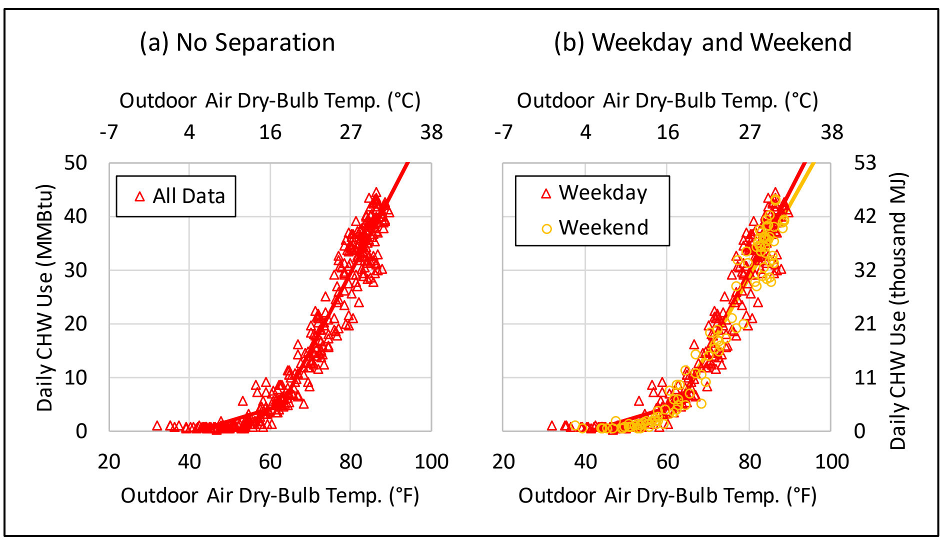

Figure 5.4 Building #25 Different Data Separation Results and Corresponding Model Series Under the Day-of-Week Scheme. Both Sub-Plots Share the Same *y*-Axis.

Table 5.4 Building #25 Data Separation Favoured by Different Metrics Under the Day-of-Week Scheme. The Most Favourable Metric is Displayed in Bold. All Metric Values are Rounded. For the LOF Inferences, a "Preferred" Model Series is Also "Not Rejected".

	Separation	g	p	SSE	R^2	CV-RMSE	R_{adj}^2	LOF Inference
(a)	No Separation	1	4	13001	0.948	17.9%	0.947	**Preferred**
(b)	Weekday/Weekend	2	8	12691	0.949	**17.8%**	**0.948**	Not Rejected
	All Separated	7	28	**12387**	**0.950**	18.1%	0.946	Full

5.3.2.2. Weekday and Weekend

Many buildings have a clear difference in their consumption levels between weekdays and weekend days. The proposed method favoured a weekday and weekend separation in 19 cases (25.0%) under the DOW scheme and 10 cases (13.2%) under the ACal scheme (with the holiday element grouped to either the weekday group or the weekend group). Shown in Figure 5.5 and Table 5.5 , the proposed method favoured a weekday-weekend separation for building #75. This building mainly has offices, classrooms, and teaching labs. Table 5.5 sees substantial improvements of CV-RMSE by approximately nine percentage points and R_{adj}^2 by approximated 0.09 in choosing a weekday/weekend separation over the no-separation model under both schemes. In the meantime, the CV-RMSE and R_{adj}^2 favoured a rather complex model series with six components: (b) six components – Mon and Tue combined and all other elements separated under the DOW scheme and (d) six components – academic weekend and break weekend combined and all other elements separated under the ACal scheme. Again, their improvement vis-à-vis their two-component counterparts is extremely marginal. Therefore, the choices of the LOF tests are well justified.

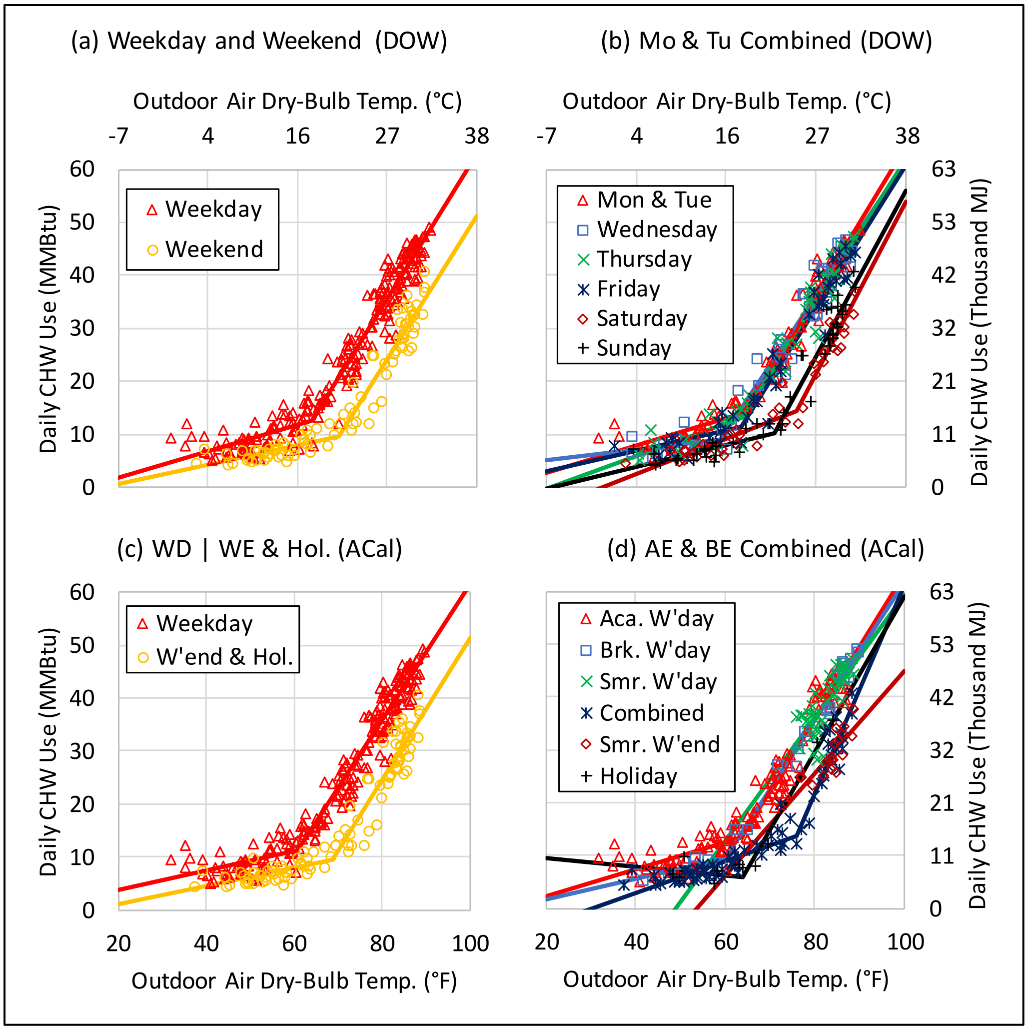

Figure 5.5 Building #75 Different Data Separation Results and Corresponding Model Series Under (a)(b) the Day-of-Week Scheme and (c)(d) the Academic Calendar Scheme. The Sub-Plots Share the Same *x*- and *y*-Axes.

Table 5.5 Building #75 Data Separation Favoured by Different Metrics Under Both Schemes. The Most Favourable Metric is Displayed in Bold. All Metric Values are Rounded. For the LOF Inferences, a "Preferred" Model Series is Also "Not Rejected".

	Separation	g	p	SSE	R^2	CV-RMSE	R^2_{adj}	LOF Inference
	No Separation	1	4	7605	0.870	21.3%	0.869	Rejected (both)
Day-of-Week								
(a)	Weekday/Weekend	2	8	2473	0.958	12.2%	0.957	**Preferred**
(b)	Mo & Tu Combined	6	24	2131	0.964	**11.7%**	**0.961**	Not Rejected
	All Separated	7	28	**2117**	**0.964**	11.7%	0.960	Full

Table 5.5 Continued

Separation		g	p	SSE	R^2	CV-RMSE	R^2_{adj}	LOF Inference
Academic Calendar								
(c)	WD | WE & HL	2	8	2344	0.960	11.9%	0.959	**Preferred**
(d)	AE & BE Combined	6	20	2127	0.964	**11.6%**	**0.961**	Not Rejected
	All Separated	7	24	**2108**	**0.964**	11.6%	0.961	Full

5.3.2.3. Seasonal Change

Some buildings' energy consumption level changes seasonally. Building #60 is a residence hall. Its CHW consumption does not change along a weekly cycle, but is slightly higher on academic days (both weekdays and weekend days) than all other days. The results for this building are shown in Figure 5.6 and Table 5.6 . In Figure 5.6 the two model components are superposed on all data. Although the data points are relatively tight and difficult to differentiate simply by visual inspection, academic days tend to appear on the upper side of the data cloud in Figure 5.6 (a) and all other days tend to appear on the lower side in Figure 5.6 (b). Table 5.6 shows that this subtle difference led to meaningful model improvement from a no-separation model, especially by 1.8 percentage points in CV-RMSE.

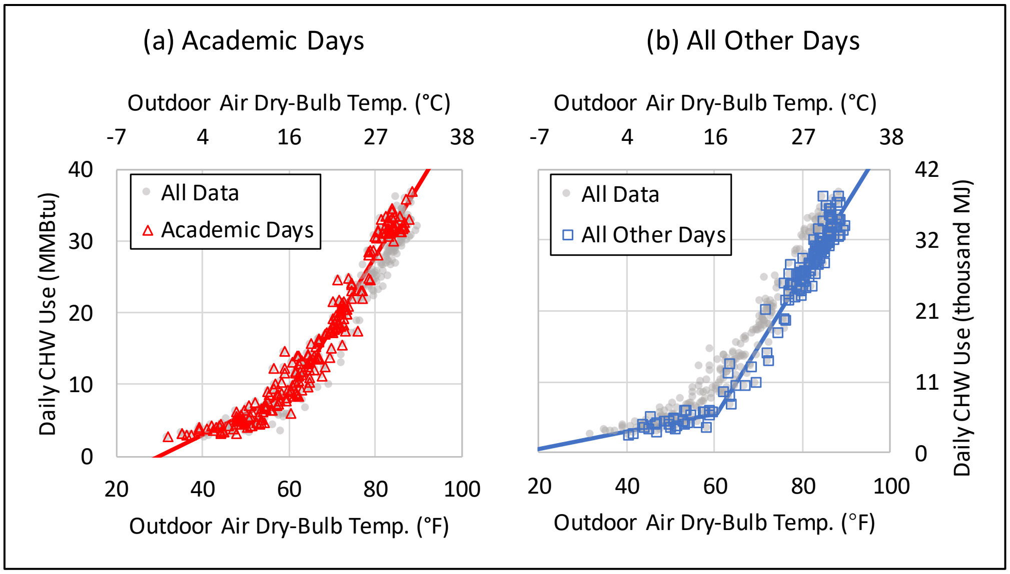

Figure 5.6 Building #60 Two-Component Separation Under the Academic Calendar Scheme. Data Subsets of the Corresponding Model Components are Superposed on the Whole Dataset. Both Sub-Plots Share the Same *y*-Axis.

Table 5.6 Building #60 Data Separation Favoured by Different Metrics Under the Academic Calendar Scheme. The Most Favourable Metric is Displayed in Bold. All Metric Values are Rounded. For the LOF Inferences, a "Preferred" Model Series is Also "Not Rejected".

Separation	g	p	SSE	R^2	CV-RMSE	R^2_{adj}	LOF Inference
No Separation	1	4	1671	0.960	11.7%	0.960	Rejected
AD & AE | Other	2	8	1181	0.972	9.9%	0.971	**Preferred**
SD & BD combined	6	22	1059	0.975	**9.6%**	**0.973**	Not Rejected
All Separated	7	24	**1057**	**0.975**	9.6%	0.973	Full

5.3.2.4. Complex Scheduling

There are also cases where the LOF test suggested that a more complex model series was necessary. Building #65 is an office building and its results are shown in Figure 5.7 and Table 5.7 . Upon observing the scatter plot, one could find it immediately evident that the weekday and weekend data of this building should be separated, as shown in Figure 5.7 (a). However, according to Table 5.7 this intuitive separation is rejected by the LOF test. The LOF test supported a four-component separation in which both Monday and Tuesday are both singled out, which improved CV-RMSE to 10.1% from the weekday-and-weekend separation with 14.1%. However, in this case the energy analyst may find the LOF's separation of Tuesday data oversensitive and favour the three-component option with only Monday singled-out, as the difference in CV-RMSE is only 0.3 percentage points and in R^2_{adj} only 0.002.

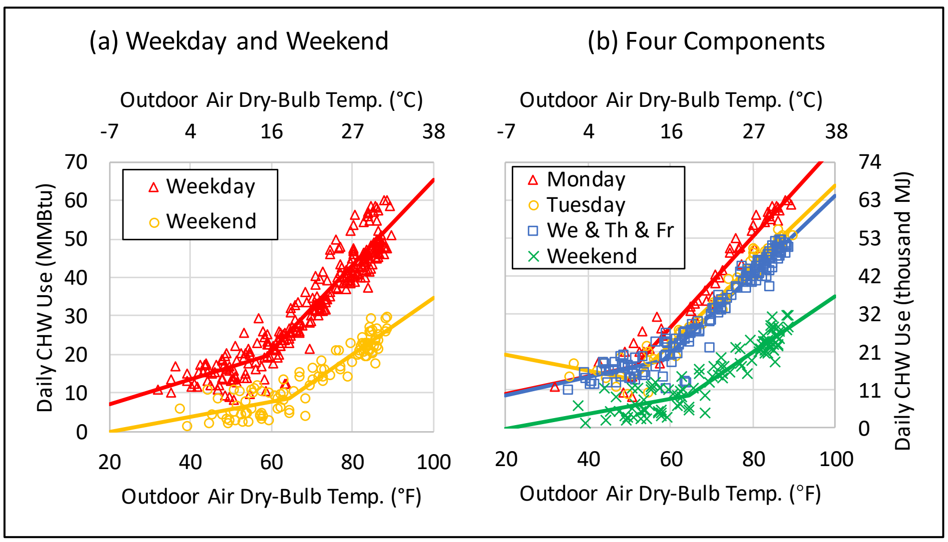

Figure 5.7 Building #65 Different Data Separation Results and Corresponding Model Series Under the Day-of-Week Scheme. Both Sub-Plots Share the Same *y*-Axis.

Table 5.7 Building #65 Data Separation Favoured by Different Metrics Under the Day-of-Week Scheme Together with the Weekday-and-Weekend Separation. The Most Favourable Metric is Displayed in Bold. All Metric Values are Rounded. For the LOF Inferences, a "Preferred" Model Series is also "Not Rejected".

	Separation	g	p	SSE	R^2	CV-RMSE	R^2_{adj}	LOF Inference
	No Separation	1	4	33026	0.594	34.5%	0.589	Rejected
(a)	Weekday/Weekend	2	8	5414	0.933	14.1%	0.932	Rejected
	Mo | WD | WE	3	12	2929	0.964	10.4%	0.963	Rejected
(b)	Mo | Tu | WD | WE	4	16	2745	0.966	10.1%	0.965	**Preferred**

Table 5.7 Continued

Separation	g	p	SSE	R^2	CV-RMSE	R_{adj}^2	LOF Inference
Th & Fr Combined	6	24	2605	0.968	**10.0%**	**0.966**	Not Rejected
All Separated	7	28	**2589**	**0.968**	10.0%	0.965	Full

5.4. Discussion

5.4.1. Comparing the Metrics

Among all cases in general, the proposed LOF method favoured model simplicity the most heavily, followed by CV-RMSE and R_{adj}^2, and finally SSE and R2. Using the number of parameters against the goodness-of-fit assessment, CV-RMSE and R_{adj}^2 never preferred a model series that is more complex than those suggested by SSE and R2. On the other hand, although SSE and R^2 are expected always to prefer the model series with the largest number of parameters according to their statistical properties, this is untrue for one case under the DOW scheme and for 11 cases under the ACal scheme, out of 76 cases each. The reason is thought to be that the regression process of the 4P-CP model is not strictly a single least-squares fitting. As the change-point temperature in Equation (1) and in Equation (2) is found through a grid search instead, this breaks the assumption underlying the statement that *SSE* and R^2 will always prefer a model with more parameters. This behaviour seems further amplified under the unbalanced ACal scheme.

For the two classic metrics that penalise number of model parameters, although they favoured the same model in all cases, CV-RMSE exhibited more obvious improvement than R_{adj}^2. Two reasons are thought to be behind this phenomenon. The first is that in the selected cases presented, R_{adj}^2 are already relatively close to the upper limit of 1, leaving little space for it to improve visibly. The second reason is the property that CHW change-point models tend to have one segment that is relatively flat, in which case CV-RMSE would be more sensitive to fit quality changes than R^2 or R_{adj}^2 [152].

The proposed LOF test always preferred the same or simpler model series compared to all other traditional metrics among all case study buildings. This validates its superiority should the analyst intend to weigh the model simplicity more heavily than done by CV-RMSE and R_{adj}^2.

5.4.2. Reliability

The proposed method is designed to aid the energy analyst's decision making quickly and automatically in data separation possibilities for a portfolio of buildings. However, for many buildings among the cases used in this study, it is evident that the dataset requires no treatment or needs but a simple weekday-weekend separation. The proposed method is therefore deemed reliable since it tends to agree with the analyst when there is a clearly favourable way to divide the data, as is demonstrated in Sections 5.3.2.1 and 5.3.2.2.

On the other hand, in the case demonstrated in Section 5.3.2.3, the two-component separation favoured by the LOF test significantly improved the model fit vis-à-vis a no-separation model. The CV-RMSE was improved from 11.7% to 9.9%. This is valuable because the difference between the two components is difficult to identify by simply visual examination. In the case demonstrated in Section 5.3.2.4, the LOF test successfully singled out Monday data which turned out to be visibly higher than all other data points, but notably it also singled out Tuesday data whose difference could be arguable. Table 5.7 showed that compared to having Tuesday combined with other weekdays, the favoured model's CV-RMSE improved from 10.4% to 10.1% and the three-component model series (without Tuesday singled out) was inferred by the LOF test to be significantly worse compared to the saturated model. It is ultimately the analyst's decision whether the LOF test is being too sensitive in this case.

5.4.3. The Data Separation Schemes

It is worth noting that this work purposely did not provide direct comparison between the two data separation schemes. The primary reason is that the LOF test deployed in this work is only suitable for situations where one model is a simplification of the other. To test non-nesting models across schemes, one may consider Vuong's closeness test [156]. However, this is a likelihood-ratio-based test that involves eigenvalue computation, which could be considered far too difficult for quick and wide implementation. A scheme such as ACal is not widely used but customarily made for a specific building or group of buildings. The rationale of including it in this work is merely to demonstrate how the proposed method behaves differently under unbalanced data treatment, and all seasonal data separation schemes may well be as unbalanced as the ACal scheme.

5.5. Summary

This work proposes a procedure based on the analysis of variance to provide data separation suggestions automatically and comprehensively. With a pre-defined set of EDT, the proposed method using a series of LOF tests is able to examine all data separation possibilities and balance between model accuracy and simplicity. The method was tested on 76 case study buildings in Texas, USA. Whole-year, whole-building daily measured CHW consumption data were used. The method was also tested under two data separation schemes, DOW and ACal, and compared to two classic goodness-of-fit metrics that penalise model complexity, CV-RMSE and R_{adj}^2.

For all cases, the LOF test always favoured models simpler than or the same as those favoured by the other metrics. This indicates that the proposed method favours model simplicity more heavily than the classic metrics. The proposed method suggested that no data separation was needed in 61.8% of the cases when the data separation scheme was balanced. This is seen as an indication that the method is reliable and not overly sensitive. However, it was also able to identify subtle differences between consumption levels and point out data separation plans which led to substantial model fit improvement. Across the cases where the proposed method did favour a data separation, the CV-RMSE was improved by 7.6 percentage points on average compared to their no-separation counterparts.

Two different types of data separation schemes, both with seven elements, have been discussed. Under the balanced data separation scheme DOW, the proposed procedure favoured no separation in most cases and CV-RMSE and R_{adj}^2 favoured a two-component separation in a plurality of cases (44.7%). This indicates that these two traditionally popular parameter-penalising metrics are somewhat safe to use when the data separation is balanced. On the other hand, when an unbalanced scheme such as ACal was applied to the same dataset, all complexity-penalising metrics tended to favour models with more components. The distance between the LOF test and the two other metrics also widened, with LOF favouring two-component separations in a plurality of cases (38.2%) and the other two metrics favouring five components (34.2%). This suggests that the proposed method is a more reliable option should simplicity be preferred for buildings with unbalanced seasonal patterns.

6. CONCLUSIONS AND FUTURE WORK

6.1. Conclusions

The aim of this work is to expand and update the statistical modelling framework for whole-building energy consumption data for a commercial building. It has applied measured building energy data on the Texas A&M University campus in College Station to test a specialised formulation, two occupancy proxies, and automated scheduling techniques.

Firstly, a break-point formulation is proposed in Section 3. It was designed to address an unexpected load shape of CVRH systems' HHW consumption specifically and it managed to improve the HHW consumption models significantly, producing more favourable goodness-of-fit metrics and residual behaviour. Although the average change in CV-RMSE of 1.2 percentage points appears marginal, the improvement of each sample was statistically significant with lack-of-fit tests' *p*-values far smaller than 0.05. This improvement is not only related to regression metrics, but also in its implications for enhanced reliability in diagnostics and energy consumption prediction due to the improvement residual behaviour. Buildings with constant volume reheat systems often have simultaneous heating and cooling during their operation. There is appreciable energy savings potential and they can be good candidates for commissioning. However, such buildings are usually old and lack modern control and metering systems. Their energy data, even when recorded, tend to be lower quality and their profile can be difficult to recognize. The proposed modelling method in this work is valuable in these applications.

Secondly, this work explored the use of domestic cold water (DCW) or ELE as an occupancy proxy to improve whole-building statistical models in Section 4. It found that a combination of DCW consumption data and the DBSCAN clustering technology could better capture the occupancy fluctuation and thus produce better models than attempting to divide data into groups by calendar only. For the typical years, even when the weekday-and-weekend method on average had favourable models with CV-RMSE values averaging 14.2%, the DCW clustering made further improvements of 2.9 percentage points on average. These results are valuable in a number of aspects. Firstly, as occupancy plays an important role in building energy consumption, this work compared several methods and managed to recommend one of them to use this information effectively in whole-building statistical models. Secondly, since occupancy sensors are still not widely adopted in buildings, especially not in buildings typically targeted in energy efficiency projects, the proposed method is valuable since it provides energy calculation methods of higher quality for M&V applications.

Last but not least, an automatic and comprehensive day-type grouping technique was proposed in Section 5. The proposed method simplifies the energy analyst's task of finding a suitable way to separate energy data to accommodate building operation or occupancy schedule, when such information is incomplete or not easily available. Its comprehensive search approach also helps the energy analyst find unintuitive data separation options that could lead to substantial model-fit improvement. Across the cases where the proposed method did favour a data separation, the CV-RMSE was improved by 7.6 percentage points on average compared to their no-separation counterparts.

Compared to traditional complexity-penalising metrics of R^2_{adj} and CV-RMSE, the proposed method favoured even simpler models. The two traditional metrics most often favoured models with two components under the balanced data separation scheme (44.7%) and five under the unbalanced scheme (34.2%), while the proposed method favoured one (61.8%) and two (38.2%), respectively. Although this work introduced this method in the context of statistical modelling using daily CHW data, its procedure and results may be valuable in other energy baseline modelling scenarios such as hourly data modelling, other energy types, or calibrated simulation.

A good model-fit is the cornerstone for important applications such as M&V, some control strategies, and anomaly detection. The techniques proposed in this work took advantage of the improved quantity and quality of energy consumption data and built upon the classic whole-building statistical methods. It aims to expand the standard practice beyond a single change-point, beyond a simple weekday-and-weekend data separation. Ultimately, the goal of this work is to update the framework of whole-building energy consumption statistical models in a more contemporary mindset.

6.2. Future Work

This work has potential and value for expansion.

The results of coupling a clustering technique with DCW consumption data discussed in Section 4 showed significant potential for producing physically meaningful results using a machine learning method. This finding is valuable because machine learning methods such as ANN are purely black-box and their parameters cannot be interpreted. This part of the study was based on a special on-campus classroom building that did not have laboratory spaces. This helped avoid the situation where the effects of a high outdoor air intake ratio overwhelms the influence of occupancy change but largely limited the type of building on which this study can be conducted. Because of that, this study was only able to use one case study building in this part. In the future, if more suitable buildings and data can be found, the conclusions found in this part can be strengthened.

The day-type grouping technique proposed in Section 5 was tested on CHW consumption data of 76 buildings. While the size of the sample of buildings is large, the study was not able to determine whether the grouping results of each building could be connected to the building type, system type, or other building or system characteristics. This is because such information about these buildings could not be consistently collected due to constraints of time and resources. However, it is nonetheless potentially valuable if buildings whose energy consumption models could benefit from careful day-type grouping could be pre-identified based on their characteristics.

REFERENCES

[1] USEIA. U.S. energy facts explained: U.S. Energy Information Administration [updated 7 May 2020. https://www.eia.gov/energyexplained/us-energy-facts/; 2020 [accessed 7 April 2021].

[2] LLNL. Estimated U.S. energy consumption in 2019: 100.2 Quads: Lawrence Livermore National Laboratory. https://flowcharts.llnl.gov/content/assets/images/energy/us/Energy_US_2019.png; 2020 [accessed 20 October 2020].

[3] ASHRAE. Energy Estimating and Modeling Methods. 2013 ASHRAE Handbook - Fundamentals: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE); 2013. p. 19.1 - .42.

[4] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews. 2012;16(6):3586-92.

[5] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews. 2017;74:902-24.

[6] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews. 2018;81:1192-205.

[7] Bourdeau M, Zhai X-Q, Nefzaoui E, Guo X, Chatellier P. Modelling and forecasting building energy consumption: a review of data-driven techniques. Sustainable Cities and Society. 2019.

[8] Fumo N. A review on the basics of building energy estimation. Renewable and Sustainable Energy Reviews. 2014;31:53-60.

[9] EVO. Core Concepts: International Performance Measurement and Verification Protocol (IPMVP). Energy Valuation Organization; 2016.

[10] ASHRAE. ASHRAE Guideline 14: Measurement of energy, demand, and water savings. Atlanta: ASHRAE; 2014.

[11] Haberl JS, Claridge DE, Culp C. ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine what was really saved by the retrofit. Proceedings of the Fifth International Conference for Enhanced Building Operations; 2005 October 11-13, 2005; Pittsburgh, Pennsylvania.

[12] Haberl JS, Culp CC. Measurement and verification of energy savings. In: Turner W, editor. Energy Management Handbook. 6th ed. ed: The Fairmont Press, Inc.; 2009. p. 707-54.

[13] Levermore G. Performance lines and energy signatures: analysis with reference to the CIBSE Building Energy Code. Building Services Engineering Research and Technology. 1995;16(1):47-50.

[14] Fu H, Baltazar J-C, Claridge DE. Review of developments in whole-building statistical energy consumption models for commercial buildings. Renewable and Sustainable Energy Reviews. 2021;147:111248.

[15] Fu H, Baltazar J-C, Claridge DE. Break-Point Statistical Formulations for Whole Building Heating Hot Water Consumption Modeling. ASHRAE Annual Conference; June 22-26; Kansas City, MO2019.

[16] Fu H, Baltazar J-C, Claridge DE. Break-point statistical model for building heating hot water consumption with constant volume reheat systems. Energy and Buildings. 2020:109575.

[17] Fu H, Lee S, Baltazar J-C, Claridge DE. Use of domestic water consumption as proxy for occupancy level in building cooling energy model. ASHRAE Virtual Winter Conference; February 9-11, 20212021.

[18] Ruch DK, Claridge DE. A four-parameter change-point model for predicting energy consumption in commercial buildings. Journal of Solar Energy Engineering. 1992;114(2):77-83.

[19] Hudson DJ. Fitting segmented curves whose join points have to be estimated. Journal of the American Statistical Association. 1966;61(316):1097-129.

[20] KDD. 2014 SIGKDD TEST OF TIME AWARD. https://www.kdd.org/News/view/2014-sigkdd-test-of-time-award; 2014 [accessed March 14, 2021].

[21] Ester M, Kriegel H-P, Sander J, Xu X. Density-based clustering algorithms for discovering clusters. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining; 1996.

[22] Chire. DBSCAN-Illustration [updated 20 October 2011. https://en.wikipedia.org/wiki/File:DBSCAN-Illustration.svg; 2011 [accessed 17 June 2021].

[23] Hahsler M, Piekenbrock M, Doran D. DBSCAN: Fast density-based clustering with R. Journal of Statistical Software. 2019;25:409-16.

[24] Hirst E. Household energy conservation: A review of the federal residential conservation service. Public Administration Review. 1984:421-30.

[25] Hammarsten S. A critical appraisal of energy-signature models. Applied Energy. 1987;26(2):97-110.

[26] MacDonald J, Wasserman D. Investigation of metered data analysis methods for commercial and related buildings. Oak Ridge National Laboratory Oak Ridge, TN; 1989.

[27] Abushakra B, Haberl JS, Claridge DE. Overview of existing literature on diversity factors and schedules for energy and cooling load calculations. ASHRAE Transactions. 2004;110(1).

[28] Reddy TA. Literature review on calibration of building energy simulation programs: uses, problems, procedures, uncertainty, and tools. ASHRAE Transactions. 2006;112:226.

[29] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renewable and sustainable energy reviews. 2014;37:123-41.

[30] Abushakra B, Paulus MT. An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part I: Background (1404-RP). Science and Technology for the Built Environment. 2016;22(7):976-83.

[31] Harish V, Kumar A. A review on modeling and simulation of building energy systems. Renewable and sustainable energy reviews. 2016;56:1272-92.

[32] Verhelst J, Van Ham G, Saelens D, Helsen L. Model selection for continuous commissioning of HVAC-systems in office buildings: A review. Renewable and Sustainable Energy Reviews. 2017;76:673-86.

[33] Herrera M, Natarajan S, Coley DA, Kershaw T, Ramallo-González AP, Eames M, et al. A review of current and future weather data for building simulation. Building Services Engineering Research and Technology. 2017;38(5):602-27.

[34] Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martín-Bautista MJ. Data science for building energy management: A review. Renewable and Sustainable Energy Reviews. 2017;70:598-609.

[35] Claridge DE. A perspective on methods for analysis of measured energy data from commercial buildings. ASHRAE Transactions. 1998;120:150-5.

[36] Fels MF. PRISM: an introduction. Energy and Buildings. 1986;9(1-2):5-18.

[37] Reddy T. Application of Dynamic Building Inverse Models to Three Occupied Residences. Proceedings of the ASHRAE/DOE/BTECC/CIBSE Conference, Thermal Performance of the Exterior Envelopes of Buildings IV; 1989 December 4-7, 1989; Orlando, Florida.

[38] Reddy TA, Kissock JK, Ruch DK. Uncertainty in baseline regression modeling and in determination of retrofit savings. Journal of solar energy engineering. 1998;120(3):185-92.

[39] Haberl JS, Abbas M. Development of graphical indices for viewing building energy data: Part I. Journal of solar energy engineering. 1998;120(3):156-61.

[40] Haberl JS, Abbas M. Development of graphical indices for viewing building energy data: Part II. Journal of solar energy engineering. 1998;120(3):162-7.

[41] Christensen C. Digital and color energy maps for graphic display of hourly data. Solar Energy Research Inst., Golden, CO (USA); 1984.

[42] Haberl JS, Vajda E. Use of metered data analysis to improve building operation and maintenance: Early results from two federal complexes. Management Information Support, Lakewood, CO (USA); 1988.

[43] Reddy TA, Katipamula S, Kissock JK, Claridge DE. The Functional Basis of Steady-State Thermal Energy Use in Air-Side HVAC Equipment. Journal of Solar Energy Engineering. 1995;117:31-9.

[44] Claridge DE, Haberl JS, Liu M, Houcek J, Athar A. Can you achieve 150% of predicted retrofit savings? Is it time for recommissioning? Proceedings of the 1994 ACEEE Summer Study; 1994; Washington, DC: ACEEE.

[45] Kissock JK, Reddy TA, Claridge DE. Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. Journal of Solar Energy Engineering. 1998;120(3):168-76.

[46] Schrock DW, Claridge DE. Predicting energy usage in a supermarket. 1989.

[47] Lerman PM. Fitting segmented regression models by grid search. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1980;29(1):77-84.

[48] Kissock JK, Haberl JS, Claridge DE. Inverse modeling toolkit: numerical algorithms. ASHRAE Transactions. 2003;109:425.

[49] Haberl JS, Sreshthaputra A, Claridge DE, Kissock JK. Inverse model toolkit: Application and testing. ASHRAE Transactions. 2003;109:435.

[50] Kissock JK, Haberl JS, Claridge DE. Development of a Toolkit for Calculating Linear, Change-Point Linear and Multiple-Linear Inverse Building Energy Analysis Models, ASHRAE Research Project 1050-RP, Final Report. Energy Systems Laboratory, Texas A&M University; 2002.

[51] Paulus MT. Algorithm for explicit solution to the three parameter linear change-point regression model. Science and Technology for the Built Environment. 2017;23(6):1026-35.

[52] Paulus MT, Claridge DE, Culp C. Algorithm for automating the selection of a temperature dependent change point model. Energy and Buildings. 2015;87:95-104.

[53] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Transactions on Smart Grid. 2011;2(3):507-18.

[54] Crawford R, Dykowski R, Czajkowski S. A segmented linear least-squares modeling procedure for nonlinear HVAC components. ASHRAE Transactions. 1991;97(2):11-8.

[55] Kim H, Haberl JS. Development and application of weather-normalized monthly building water use model. Energy and Buildings. 2014;69:267-77.

[56] Tereshchenko T, Ivanko D, Nord N, Sartori I. Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway. CLIMA 2019; 26 - 29 May 2019; Bucharest, Romania2019.

[57] Reddy TA, Kissock JK, Katipamula S, Claridge DE. An energy delivery efficiency index to evaluate simultaneous heating and cooling effects in large commercial buildings. Journal of solar energy engineering. 1994;116(2):79-87.

[58] Kissock JK, Eger C. Measuring industrial energy savings. Applied Energy. 2008;85(5):347-61.

[59] Granderson J, Price PN. Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models. Energy. 2014;66:981-90.

[60] Heo Y, Zavala VM. Gaussian process modeling for measurement and verification of building energy savings. Energy and Buildings. 2012;53:7-18.

[61] Kissock JK, Seryak J. Lean energy analysis: identifying, discovering and tracking energy savings potential. Proceedings of Society of Manufacturing Engineers: Advanced Energy and Fuel Cell Technologies
Conference; 2004 Oct 11-13; Livonia, MI.

[62] Kissock K, Eger C. Measuring Plant-Wide Energy Savings. SAE Transactions. 2006:290-302.

[63] Abels B, Sever F, Kissock K, Ayele D. Understanding industrial energy use through lean energy analysis. SAE International Journal of Materials and Manufacturing. 2011;4(1):495-504.

[64] Lammers N, Kissock K, Abels B, Sever F. Measuring progress with normalized energy intensity. SAE International Journal of Materials and Manufacturing. 2011;4(1):460-7.

[65] Ruch DK, Kissock JK, Reddy TA. Model identification and prediction uncertainty of linear building energy use models with autocorrelated residuals. ASME International Solar Energy Conference; 1993: Publ by ASME.

[66] Djuric N, Novakovic V. Identifying important variables of energy use in low energy office building by using multivariate analysis. Energy and Buildings. 2012;45:91-8.

[67] Ruch DK, Chen L, Haberl JS, Claridge DE. A change-point principal component analysis (CP/PCA) method for predicting energy usage in commercial buildings: the PCA model. Journal of Solar Energy Engineering. 1993;115(2):77-84.

[68] Masuda H, Ji J, Baltazar JC, Claridge DE. Use of first law energy balance as a screening tool for building energy use data: Experiences on the inclusion of outside air enthalpy variable. Proceedings of the Ninth International Conference for Enhanced Building Operations; 2009 November 17 - 19, 2009; Austin, TX.

[69] Shin M, Do SL. Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate. Energy and Buildings. 2016;110:57-70.

[70] Li X, Baltazar JC. Analysis of cooling regression models for hot and humid climates based on" operational effective enthalpy. The 10th International Conference for Enhanced Building Operations; 2013.

[71] Li X, Baltazar JC, Wang L. Determination of the Influence of Outdoor Air Intake Fraction on Choosing Independent Variable for Cooling Regression Modeling in Hot and Humid Climates. ASHRAE Transactions. 2016;122(1).

[72] Kim H, Haberl J. Field-test of the ASHRAE/CIBSE/USGBC performance measurement protocols: Part I intermediate level energy protocols. Science and Technology for the Built Environment. 2018;24(3):281-97.

[73] Shonder JA, Hughes PJ. Estimating Energy, Demand, and Cost Savings from a Geothermal Heat Pump ESPC Project Through Utility Bill Analysis. ASHRAE Transactions. 2006;112(2).

[74] ASHRAE. SECTION MTG -- MULTIDISCIPLINARY TASK GROUPS. https://www.ashrae.org/technical-resources/technical-committees/section-mtg-multidisciplinary-task-groups; 2017 [accessed 15 December 2020].

[75] Yan D, Hong T, Dong B, Mahdavi A, D'Oca S, Gaetani I, et al. IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. Energy and Buildings. 2017;156:258-70.

[76] Rouleau J, Ramallo-González A, Gosselin L. Towards a comprehensive tool to model occupant behaviour for dwellings that combines domestic hot water use with active occupancy. Proceedings of the 15th IBPSA Conference, San Francisco, CA, USA; 2017.

[77] Marszal-Pomianowska A, Heiselberg P, Larsen OK. Household electricity demand profiles–A high-resolution load model to facilitate modelling of energy flexible buildings. Energy. 2016;103:487-501.

[78] Jalori S, Reddy TA. A New Clustering Method to Identify Outliers and Diurnal Schedules from Building Energy Interval Data. ASHRAE Transactions. 2015;121(2).

[79] Abushakra B, Claridge DE. Modeling Office Building Occupancy in Hourly Data-Driven and Detailed Energy Simulation Programs. ASHRAE Transactions. 2008;114(2):472-81.

[80] Abushakra B, Paulus MT. An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part II: Methodology (1404-RP). Science and Technology for the Built Environment. 2016;22(7):984-95.

[81] Brown N, Wright A, Shukla A, Stuart G. Longitudinal analysis of energy metering data from non-domestic buildings. Building Research & Information. 2010;38(1):80-91.

[82] Katipamula S, Reddy TA, Claridge DE. Effect of time resolution on statistical modeling of cooling energy use in large commercial buildings. Proceedings of the 1995 ASHRAE Annual Meeting; 1995; Atlanta, GA: ASHRAE.

[83] Reddy TA, Kissock K, Claridge DE. Uncertainty analysis in estimating building energy retrofit savings in the LoanSTAR Program. Proceedings of the ACEEE 1992 Summer Study. 1992.

[84] Dhar A, Reddy TA, Claridge DE. Modeling hourly energy use in commercial buildings with Fourier series functional forms. Journal of Solar Energy Engineering. 1998;120(3):217-23.

[85] Ruch DK, Kissock JK, Reddy TA. Prediction uncertainty of linear building energy use models with autocorrelated residuals. Journal of solar energy engineering. 1999;121(1):63-8.

[86] Reiter P. Early results from commercial ELCAP buildings: schedules as a primary determinant of load shapes in the commercial sector. Pacific Northwest Lab., Richland, WA (USA); 1986.

[87] Claridge DE, Abushakra B, Haberl JS, Sreshthaputra A. Electricity diversity profiles for energy simulation of office buildings. ASHRAE Transactions. 2004;110(1):365-77.

[88] Katipamula S, Haberl JS. Methodology to identify diurnal load shapes for non-weather dependent electric end-uses. ASME JSES JSME INT SOL ENERG CONF 1991; 1991.

[89] Thamilseran S, Haberl JS. A bin method for calculating energy conservation retrofit savings in commercial buildings. 1994.

[90] Eto JH. On using degree-days to account for the effects of weather on annual energy use in office buildings. Energy and Buildings. 1988;12(2):113-27.

[91] Rabl A, Rialhe A. Energy signature models for commercial buildings: test with measured data and interpretation. Energy and Buildings. 1992;19(2):143-54.

[92] Katipamula S, Reddy TA, Claridge DE. Multivariate regression modeling. Journal of Solar Energy Engineering. 1998;120(3):177-84.

[93] Wang J, Claridge DE. Impact of nighttime shut down on the prediction accuracy of monthly regression models for energy consumption in commercial buildings. Proceedings of the Eleventh Symposium on Improving Building Systems in Hot and Humid Climates; 1998 June 1-2, 1998; Fort Worth, TX.

[94] Katipamula S. Great energy predictor shootout II: modeling energy use in large commercial buildings. ASHRAE Transactions. 1996;102(2).

[95] Thamilseran S. An inverse bin methodology to measure the savings from energy conservation retrofits in commercial buildings: Texas A&M University; 1999.

[96] Landman DS, Haberl JS. Development of pre-screening methods for analyzing energy use in K-12 public schools. The Eleventh Symposium on Improving Building Systems in Hot and Humid Climates; June 1-2, 1998; Fort Worth, TX1998.

[97] Landman DS. Development of pre-screening indices to improve energy analysis of public K-12 schools. College Station, TX: Texas A&M University; 1998.

[98] Kaplan M, McFerran J, Jansen J, Pratt R. Reconciliation of a DOE2. 1C model with monitored end-use data for a small office building. ASHRAE Transactions. 1990;96(1):981-93.

[99] Bou-Saada TE, Haberl JS. A weather-daytyping procedure for disaggregating hourly end-use loads in an electrically heated and cooled building from whole-building hourly data. American Society of Mechanical Engineers, New York, NY (United States); 1995.

[100] Seem JE. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. Energy and Buildings. 2005;37(2):127-39.

[101] Li X, Bowers CP, Schnier T. Classification of energy consumption in buildings with outlier detection. IEEE Transactions on Industrial Electronics. 2010;57(11):3639-44.

[102] Reddy TA, Claridge DE. Uncertainty of “measured” energy savings from statistical baseline models. HVAC&R Research. 2000;6(1):3-20.

[103] Reddy TA, Saman NF, Claridge DE, Haberl JS, Turner WD, Chalifoux AT. Baselining methodology for facility-level monthly energy use-part 1: Theoretical aspects. ASHRAE Transactions. 1997:336-47.

[104] Reddy TA, Saman NF, Claridge DE, Haberl JS. Baselining Methodology for Facility-Level Monthly Energy Use--Part 2: Application to Eight Army Installations. ASHRAE Transactions. 1997;103:348.

[105] Shonder JA, Im P. Bayesian analysis of savings from retrofit projects. ASHRAE Transactions. 2012;118:367.

[106] Walter T, Price PN, Sohn MD. Uncertainty estimation improves energy measurement and verification procedures. Applied Energy. 2014;130:230-6.

[107] Granderson J, Price PN, Jump D, Addy N, Sohn MD. Automated measurement and verification: Performance of public domain whole-building electric baseline models. Applied Energy. 2015;144:106-13.

[108] Granderson J, Touzani S, Custodio C, Sohn MD, Jump D, Fernandes S. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. Applied Energy. 2016;173:296-308.

[109] Granderson J, Touzani S, Fernandes S, Taylor C. Application of automated measurement and verification to utility energy efficiency program data. Energy and Buildings. 2017;142:191-9.

[110] Effinger M, Anthony J, Webster L. Case Studies in Using Whole Building Interval Data to Determine Annualized Electrical Savings. Proceedings of the Ninth International Conference for Enhanced Building Operations; November 17 - 19, 2009; Austin, Texas2009.

[111] Abushakra B. An inverse model to predict and evaluate the energy performance of large commercial and institutional buildings. Building simulation; 1997.

[112] Reddy TA, Elleson JS, Haberl JS. Methodology development for determining long-term performance of cool storage systems from short-term tests. ASHRAE Transactions. 2002;108(1):1085-103.

[113] Singh V, Reddy TA, Abushakra B. Predicting Annual Energy Use in Buildings Using Short-Term Monitoring and Utility Bills: The Hybrid Inverse Model Using Daily Data (HIM-D). ASHRAE Transactions. 2013;119(2).

[114] Singh V, Reddy TA, Abushakra B. Predicting Annual Energy Use in Buildings Using Short-Term Monitoring: The Dry-Bulb Temperature Analysis (DBTA) Method. ASHRAE Transactions. 2014;120(1).

[115] Katipamula S, Reddy TA, Claridge DE. Bias in predicting annual energy use in commercial buildings with regression models developed from short data sets. Solar Engineering. 1994;1.

[116] Perez KX, Cetin K, Baldea M, Edgar TF. Development and analysis of residential change-point models from smart meter data. Energy and Buildings. 2017;139:351-9.

[117] Burak Gunay H, Shen W, Newsham G, Ashouri A. Detection and interpretation of anomalies in building energy use through inverse modeling. Science and Technology for the Built Environment. 2019;25(4):488-503.

[118] Claridge DE, Haberl JS, Sparks R, Lopez R, Kissock K. Monitored commercial building energy data: reporting the results. ASHRAE Winter Meeting, Anaheim, CA, USA, 01/25-29/92; 1992.

[119] Turner WD, Claridge DE, O'Neal D, Haberl JS, Heffington W, Taylor D, et al. Program Overview: The Texas LoanSTAR Program: 1989–1999 A 10-year Experience. Proceedings of the 2000 ACEEE Summery Study; 2000 August.

[120] USDOE. North American Energy Measurement and Verification Protocol (NEMVP). DOE/EE-0081. Washington, DC: United States Department of Energy; 1996.

[121] Liu M, Houcek J, Athar A, Reddy TA, Claridge DE, Haberl JS. Identifying and implementing improved operation and maintenance measures in Texas LoanSTAR buildings. ACEEE 1994 Summer Study on Energy Efficiency in Buildings Proceedings: Commissioning, Operation and Maintenance; 1994.

[122] Claridge DE, Liu M, Turner WD. Commissioning for Energy Management. Energy Management Handbook. 9th Edition ed. New York: Taylor & Francis Group; 2018. p. pp. 695-731.

[123] Fels MF, Goldberg ML. Using the scorekeeping approach to monitor aggregate energy conservation. Energy and Buildings. 1986;9(1-2):161-8.

[124] Rachlin J, Fels MF, Socolow RH. The stability of PRISM estimates. Energy and Buildings. 1986;9(1-2):149-57.

[125] Stram DO, Fels MF. The applicability of PRISM to electric heating and cooling. Energy and Buildings. 1986;9(1-2):101-10.

[126] Ruch DK, Claridge DE. A development and comparison of NAC estimates for linear and change-point energy models for commercial buildings. Energy and Buildings. 1993;20(1):87-95.

[127] Kissock K, Joseph HB, McBride JR. The effects of varying indoor air temperature and heat gain on the measurement of retrofit savings. ASHRAE Transactions. 1998;104:895.

[128] Shin M, Baltazar J-C, Haberl JS, Frazier E, Lynn B. Evaluation of the energy performance of a net zero energy building in a hot and humid climate. Energy and Buildings. 2019;204:109531.

[129] ASHRAE/CIBSE/USGBC. Performance Measurement Protocols for Commercial Buildings. American Society of Heating, Refrigeration, and Air Conditioning Engineers …; 2010.

[130] Kim H. Field-test of the new ASHRAE/CIBSE/USGBC performance measurement protocols for commercial buildings: basic Level. ASHRAE Transactions. 2012;118:135.

[131] Kim H, Haberl J. Field-test of the ASHRAE/CIBSE/USGBC performance measurement protocols: Part II advanced level energy protocols. Science and Technology for the Built Environment. 2018;24(3):298-315.

[132] Shao X, Claridge DE. Use of first law energy balance as a screening tool for building energy data, part I--methodology. ASHRAE Transactions. 2006;112(2):717-32.

[133] Ji J, Baltazar JC, Claridge DE. Study of the outside air enthalpy effects in the screening of metered building energy data. Proceedings of the Sixteenth Symposium on Improving Building Systems in Hot and Humid Climates; 2008 December 15-17; Plano, TX.

[134] Masuda H, Baltazar JC, Ji J, Claridge DE. Development of Data Quality Control Limits for Data Screening Through the" Energy Balance" Method. Proceedings of the Sixteenth Symposium on Improving Building Systems in Hot and Humid Climates; 2008 December 15-17; Plano, TX.

[135] Seem JE. Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy and Buildings. 2007;39(1):52-8.

[136] Masuda H, Claridge DE. Estimation of building parameters using simplified energy balance model and metered whole building energy use. Proceedings of the Twelfth International Conference for Enhanced Building Operations; 2012 October 23-26; Manchester, UK.

[137] Reddy TA, Deng S, Claridge DE. Development of an inverse method to estimate overall building and ventilation parameters of large commercial buildings. Journal of solar energy engineering. 1999;121(1):40-6.

[138] MacKay DJ. Bayesian nonlinear modeling for the prediction competition. ASHRAE Transactions. 1994;100(2):1053-62.

[139] Bou-Saada T, Haberl J, Vajda EJ, Shincovich M, D'Angelo L, Harris L. Total utility savings from the 37,000 fixture lighting retrofit to the US DOE Forrestal Building. Proceedings of the 1996 ACEEE Summery Study. 1996.

[140] Chonan Y. A Bayesian nonlinear regression with multiple hyperparameters on the ASHRAE II time-series data. ASHRAE Transactions 1996. 1996;102:405-11.

[141] Haberl JS, Thamilseran S, Reddy TA, Claridge DE. Baseline calculations for measurement and verification of energy and demand savings in a revolving loan program in Texas. ASHRAE Transactions. 1998;104:841.

[142] Shin M, Baltazar J-C, Haberl JS, Frazier E, Lynn B. Side-by-side tests of a net-zero energy building. Building Simulation 2017; 2017 Aug. 7-9; San Fracisco, CA, USA: International Building Performance Simulation Association (IBPSA) San ….

[143] Server F, Kissock JK, Brown D, Mulqueen S. Estimating industrial building energy savings using inverse simulation. 2011 ASHRAE Winter Conference; 2011 January 29-February 2; Las Vegas, NV.

[144] Katipamula S, Reddy TA, Claridge DE. Development and application of regression models to predict cooling energy consumption in large commercial buildings. Proceedings of the 1994 ASME/JSME/JSES International Solar Energy Conference; 1994: ASME.

[145] Kissock JK, Reddy TA, Fletcher D, Claridge DE. Effect of short data periods on the annual prediction accuracy of temperature-dependent regression models of commercial building energy use. ASME International Solar Energy Conference; 1993: ASME.

[146] Knebel DE. Simpilified energy analysis using the modified bin method: ASHRAE; 1983.

[147] Montgomery DC. Design and analysis of experiments: John wiley & sons; 2017.

[148] ASHRAE. Load and Energy Calculations. 2013 ASHRAE Handbook - Fundamentals2013.

[149] Hoes P, Hensen JL, Loomans MG, de Vries B, Bourgeois D. User behavior in whole building simulation. Energy and Buildings. 2009;41(3):295-302.

[150] O'Brien W, Tahmasebi F, Andersen RK, Azar E, Barthelmes V, Belafi ZD, et al. An international review of occupant-related aspects of building energy codes and standards. Building and Environment. 2020;179:106906.

[151] Feng X, Yan D, Hong T. Simulation of occupancy in buildings. Energy and Buildings. 2015;87:348-59.

[152] Draper NR, Smith H. Applied regression analysis: John Wiley & Sons; 1998.

[153] Fu H, Baltazar J-C, Claridge DE. Identifying Peak and Base Energy Consumption Hour Ranges for Commercial Buildings Using a Non-Parametric Method. ASHRAE Winter Conference; January 20-24; Chicago, Illinois2018.

[154] Dunn OJ. Multiple comparisons among means. Journal of the American Statistical Association. 1961;56(293):52-64.

[155] Bonferroni C. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze. 1936;8:3-62.

[156] Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society. 1989:307-33.

APPENDIX A

DEFINITIONS AND FORMULAE OF STATISTICAL METRICS

Sum of square error (SSE), or the residual sum of squares, defined in Equation A.1, is a measure of the discrepancy between the data and the model estimation and is used as an optimality criterion in model selection. A small SSE indicates a tight model fit.

$$SSE = \sum_i (y_i - \hat{y}_i)^2 \tag{A.1}$$

where y_i is the measured value and $\hat{y}_i$ is the estimated value.

Coefficient of determination (R^2), defined in Equation A.2, is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). A large R^2 indicates a good model fit.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{A.2}$$

where $\bar{y}$ is the sample average.

Adjusted R^2 (R^2_{adj}), defined in Equation A.3, accounts for the phenomenon of the R^2 automatically increasing when extra regression parameters are added to the model. A large R^2_{adj} indicates a good model fit.

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{A.3}$$

where n is the sample size and p is the number of model parameters.

Coefficient of variance of the root mean square error (CV-RMSE), defined in Equation A.4, is a normalisation of the SSE by the sample average, which also accounts for the number of regression parameters.

$$CV(RMSE) = \frac{\sqrt{\frac{1}{n-p}\sum_i (y_i - \hat{y})^2}}{\bar{y}} \qquad \text{(A.4)}$$

where all symbols have the same meaning as when they were used in the previous equations.

APPENDIX B

DAY-TYPE GROUPING RESULTS FOR ALL CASE STUDY BUILDINGS

Table B.1. Results for all case study buildings. Number of model components (g) and number of model parameters (p) are listed for models favoured by each metric.

Bldg.	Sample Size (day)	Day-of-Week (DOW) Scheme						Academic Calendar (ACal) Scheme					
		SSE or R^2		CV-RMSE or R^2_{adj}		LOF		SSE or R^2		CV-RMSE or R^2_{adj}		LOF	
		g	p	g	p	g	p	g	p	g	p	g	p
#01	365	7	28	4	16	2	8	7	24	6	20	4	14
#02	356	7	28	2	8	1	4	7	24	4	16	3	12
#03	354	7	28	4	16	2	8	7	24	5	16	2	8
#04	312	7	28	3	12	2	8	6	20	3	12	3	10
#05	335	7	28	3	12	1	4	7	24	5	16	2	8
#06	364	7	28	2	8	1	4	7	24	5	18	2	8
#07	334	7	28	1	4	1	4	7	24	5	18	1	4
#08	365	7	28	4	16	2	8	7	24	6	20	4	14
#09	365	7	28	5	20	2	8	7	24	5	16	2	8
#10	365	7	28	2	8	1	4	7	24	6	20	2	6
#11	365	7	28	2	8	1	4	7	24	6	22	3	12
#12	365	7	28	1	4	1	4	6	20	5	18	3	12
#13	365	7	28	2	8	1	4	6	20	6	20	1	4
#14	364	7	28	1	4	1	4	7	24	7	24	3	12
#15	365	7	28	2	8	2	8	6	20	5	18	3	12
#16	364	7	28	2	8	1	4	7	24	5	18	3	12
#17	360	7	28	2	8	1	4	7	24	6	22	3	12
#18	335	7	28	2	8	1	4	7	24	7	24	4	14
#19	365	7	28	2	8	1	4	7	24	6	22	4	16
#20	365	7	28	3	12	2	8	7	24	7	24	3	12
#21	365	7	28	2	8	2	8	7	24	5	16	2	8

Bldg.	Sample Size (day)	Day-of-Week (DOW) Scheme						Academic Calendar (ACal) Scheme					
		SSE **or** $\boldsymbol{R^2}$		**CV-RMSE or** $\boldsymbol{R^2_{adj}}$		**LOF**		***SSE*** **or** $\boldsymbol{R^2}$		**CV-RMSE or** $\boldsymbol{R^2_{adj}}$		**LOF**	
		g	***p***	***g***	***p***	***g***	***p***	***g***	***p***	***g***	***p***	***g***	***p***
#22	365	7	28	2	8	1	4	6	20	3	12	1	4
#23	365	7	28	2	8	1	4	7	24	5	18	2	6
#24	365	7	28	6	24	4	16	6	22	6	22	4	14
#25	365	7	28	2	8	1	4	7	24	5	16	2	8
#26	360	7	28	5	20	2	8	7	24	5	18	4	14
#27	364	7	28	4	16	2	8	7	24	6	22	3	10
#28	365	7	28	3	12	2	8	7	24	6	20	3	12
#29	365	7	28	3	12	2	8	7	24	5	18	3	10
#30	313	7	28	4	16	1	4	6	22	3	12	2	8
#31	365	7	28	3	12	1	4	7	24	5	20	3	12
#32	365	7	28	7	28	4	16	7	24	6	22	4	14
#33	365	7	28	2	8	1	4	6	20	6	20	3	10
#34	365	7	28	4	16	1	4	7	24	5	18	1	4
#35	365	7	28	5	20	3	12	7	24	7	24	5	18
#36	365	7	28	2	8	1	4	7	24	5	16	2	8
#37	365	7	28	2	8	2	8	7	24	5	18	3	12
#38	364	7	28	4	16	2	8	7	24	5	18	2	8
#39	364	7	28	2	8	2	8	7	24	7	24	3	12
#40	354	7	28	4	16	1	4	7	24	7	24	2	8
#41	362	7	28	2	8	1	4	6	20	5	18	3	10
#42	365	7	28	3	12	2	8	7	24	5	16	2	8
#43	364	7	28	3	12	2	8	7	24	6	20	2	8
#44	360	7	28	3	12	1	4	7	24	3	10	1	4
#45	325	7	28	2	8	1	4	7	24	5	18	2	8
#46	365	7	28	2	8	1	4	7	24	3	10	1	4
#47	365	7	28	3	12	1	4	7	24	3	12	2	8
#48	365	7	28	3	12	2	8	7	24	6	20	4	14
#49	365	7	28	2	8	1	4	7	24	2	8	1	4
#50	365	7	28	6	24	3	12	7	24	7	24	4	14
#51	365	7	28	1	4	1	4	7	24	4	14	2	8
#52	365	7	28	1	4	1	4	7	24	4	12	1	4
#53	354	7	28	6	24	3	12	7	24	5	18	2	8
#54	282	7	28	6	24	4	16	7	22	5	16	2	8

Bldg.	Sample Size (day)	Day-of-Week (DOW) Scheme						Academic Calendar (ACal) Scheme					
		SSE or R^2		CV-RMSE or R^2_{adj}		LOF		SSE or R^2		CV-RMSE or R^2_{adj}		LOF	
		g	p	g	p	g	p	g	p	g	p	g	p
#55	355	6	24	3	12	2	8	7	24	6	20	2	8
#56	365	7	28	2	8	1	4	7	24	6	20	2	6
#57	344	7	28	2	8	1	4	6	20	4	16	2	8
#58	365	7	28	1	4	1	4	7	24	6	22	2	8
#59	365	7	28	2	8	1	4	7	24	7	24	4	14
#60	365	7	28	2	8	1	4	7	24	6	22	2	8
#61	365	7	28	2	8	1	4	7	24	5	18	3	10
#62	365	7	28	5	20	3	12	7	24	6	20	4	14
#63	365	7	28	2	8	1	4	7	24	4	14	2	8
#64	354	7	28	2	8	1	4	7	24	5	18	1	4
#65	365	7	28	6	24	4	16	7	24	5	18	3	12
#66	365	7	28	2	8	1	4	7	24	3	10	1	4
#67	365	7	28	2	8	1	4	7	24	4	12	2	6
#68	365	7	28	2	8	1	4	7	24	4	14	2	6
#69	362	7	28	1	4	1	4	7	24	3	10	1	4
#70	365	7	28	3	12	1	4	7	24	3	10	1	4
#71	365	7	28	1	4	1	4	7	24	3	10	1	4
#72	365	7	28	2	8	1	4	7	24	3	10	1	4
#73	365	7	28	2	8	1	4	7	24	3	10	2	6
#74	299	7	28	5	20	3	12	6	20	4	14	3	12
#75	313	7	28	6	24	2	8	7	24	6	20	2	8
#76	312	7	28	3	12	1	4	7	24	4	14	1	4