IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 20, 2018, accepted June 20, 2018, date of publication July 2, 2018, date of current version August 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2852478

Going Beyond the Mean: Distributional Degree-Day Base Temperatures for Building Energy Analytics Using Change Point Quantile Regression

QINGLONG MENG[1,2], MONJUR MOURSHED[2], AND SHEN WEI[3]
[1]School of Civil Engineering, Chang'an University, Xi'an 710061, China
[2]School of Engineering, Cardiff University, Cardiff CF24 3AA, U.K.
[3]The Bartlett School of Construction and Project Management, University College London, London WC1E 7HB, U.K.

Corresponding author: Qinglong Meng (mengqinglong@chd.edu.cn)

This work was supported in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2018JM5084, the School of Engineering, Cardiff University, and in part by the European Commission via the Horizon 2020 Project, TABEDE, under Grant 766733.

ABSTRACT Building energy consumption patterns are primarily affected by building function, operation, occupancy and thermal characteristics. A robust method of energy use pattern recognition is, therefore, essential. Heating degree-days (HDD) are routinely used for heating energy consumption prediction and analytics, the accuracy of which depends on how well the base temperature corresponds with the patterns of energy use. A change-point quantile regression (CPQR) technique is proposed for better identification of the base temperature, which is then applied in three buildings with distinct operational energy use patterns: weekday only, weekday plus occasional weekend, and all-year operation. Compared with the conventional regression and change-point least square (CPLS) methods, our CPQR approach determines a range of base temperatures of corresponding energy use patterns across quantiles from 0.05 to 0.95, at an interval of 0.05. Consequently, daily HDDs computed using the range of base temperatures of corresponding quantiles result in more accurate predictions of heating energy consumption. CPQR improves estimation accuracy and is more robust than CPLS because (a) it considers the whole distribution of energy consumption not just the mean, (b) pre-processing of raw data other than the removal of anomalies is not needed, and (c) it can better characterize the data with abnormal energy distribution. Also, CPQR-based method can better characterize the weather dependence of energy consumption than the conventional CPLS regression.

INDEX TERMS Base temperature, building energy use pattern, change-point quantile regression, gas consumption, heating degree-day.

I. INTRODUCTION

Buildings account for 35-40% global energy consumption and are responsible for over 33% of global greenhouse gas (GHG) emissions [1], [2]. Consequently, buildings are crucial for reducing energy consumption and corresponding emissions. To reduce the use of non-renewable fossil fuel (e.g., oil, natural gas) within buildings, two measures can mainly be applied, which are (a) broadening the sources of energy and (b) minimizing the demand for energy. The former measure can be achieved by utilizing of renewable energy sources such as solar, wind and geothermal energy. However, the latter involves many aspects in which active energy technologies are recommended to tackle the projected increase in building energy demand, which is essential to the sustainability of our environment.

The primary aim of sustainability related policies and regulations related to buildings is to minimize energy consumption from buildings and improve their performance to reduce carbon footprint, whilst not compromising the standard requirements for usage, such as occupant thermal comfort. However, energy consumption in buildings is highly linked to weather conditions. For example, without policy interventions, building energy use could increase by as much as 41-87% between 2020 and 2060 in Los Angeles, California, due to the temperature increase caused by climate change [3]. Therefore, a good understanding of the response of building energy consumption to local climate is of benefit to energy-efficient operation and management of buildings. In the past several decades, researchers have investigated ways to develop methods to reliably predict building energy consumption, namely engineering, statistical, grey-box modeling, machine learning and artificial intelligence methods [4]. According to the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) [5] the degree-day method is one of the simplest for energy analytics in buildings, and can be adopted for day-to-day energy monitoring and management.

The degree-day method can be used for both building energy prediction [6], [7] and energy management. Heating degree-day (HDD), is a versatile measure of the impact of the severity and duration of cold weather. It enables weather-related analysis on the consumption of fuels such as natural gas. HDDs are a summation of the difference between actual outdoor temperature and a 'reference' or 'base' temperature.

Appropriately determined base temperatures can help to derive a realistic representation of building energy consumption and efficiency, while inappropriate one would lead to energy waste or uncomfortable indoor environment [8]. Therefore, the determination of appropriate base temperatures is the premise of using the HDD method. Although there are two widely-used official base temperatures; i.e., 15.5°C in the UK [9] and 18.3°C in the USA [5], both temperatures cannot address individual buildings well. This is because the actual value of the base temperatures may vary widely from one building to another, due to building characteristics, occupant density, occupant behavior and other factors.

Currently, there are two classical methods, i.e. energy signature method and performance line method, which can be used to determine the base temperatures for buildings [10]–[12]. Several research have further improved these methods [13]–[15]. The use of the energy signature method is growing due to the increased availability of detailed utility bills, historical weather data and high-resolution smart meter data [16]. Most previous studies used data from simulation, although some have adopted data from monitoring. The data used in the past studies are often of low resolution (e.g., daily), spanning short periods (e.g., no longer than a year). For a robust determination of the base temperature, rich field-measured data is still necessary to reflect what is really happening in the building.

There are some difficulties that need to be addressed before using rich raw data for the determination of base temperatures of a building. First, rich raw measurement data is generally not as tidy as simulated data or measurements over short periods, and may contain anomalies [17]. Second, the disaggregated data of occupied and unoccupied period has a significant impact on building energy use analysis. Common methods directly disaggregate occupied and unoccupied data according to the day of a week or public bank holidays, and this may lose important information. Finally, an important assumption behind the OLS (Ordinary Least Squares) regression is that all independent variables are normally distributed. However, building energy use data is often not normally distributed, hence not suitable for the ordinary least square method [18]. On the other hand, quantile regression (QR) is a good candidate for solving the above challenges.

The definition of base temperature suggests that linear quantile regression is not robust enough to characterize the relationship between energy use and ambient temperature. We, therefore, combined the change-point model with quantile regression, and the resulting method is called change-point quantile regression (CPQR). The evaluation of the CPQR method was based on sub-hourly gas consumption measured from three case-study buildings located in Cardiff, UK. The work can be considered as one of the most comprehensive studies to date on the determination of base temperatures for specific buildings - not only because of a full distribution picture of base temperatures but also because of automatic identification of building energy use patterns. This work provides a demand-side lever for energy users to explore the potential impact of ambient temperature on building energy consumption.

The rest of the paper is structured as follows. Section II describes the theory and related work; Section III briefly introduces the data used to evaluate the proposed method; Section IV details the proposed method from this study, and Section V discusses the results from the quantile regression. A brief conclusion about this study is given in Section VI.

II. THEORY AND RELATED WORK

Theories related to the methodology used in the paper is discussed in this section. The theories include one of energy signature techniques-three-parameter change-point model for heating, daily heating degree-days (HDD) calculation and statistical indices for regression model evaluation.

A. CHANGE-POINT REGRESSION

The change-point (CP) [19] model has been further developed to estimate the change of patterns in a CP regression model, which has been widely used as baseline energy signature models for assessing energy efficiency. Furthermore, due to its simplicity, the CP model is the most appropriate model in terms of accuracy vs. effort spent for verification of whole building energy consumption and estimating building parameters [15]. In this study, therefore, the CP model has been selected to capture the characteristics of the relationship between dependent (e.g. building energy consumption) and independent variables (e.g., ambient air temperature, HDD). The best-fit change-point model from ASHRAE Inverse Modelling Toolkit (IMT)) [20] was used in this research to derive regression models of building energy use. The functional form for best-fit three-parameter change-point models for heating (3PH), is given in (1).

$$Y_h = \alpha_1 + \beta_1 (t_b - X_1)^+ \tag{1}$$

where Y_h is energy use (here, gas consumption in kWh), X_1 is ambient dry bulb temperature (°C), α_1 is baseline energy consumption or base load, the part of energy use which is not dependent of weather conditions, β_1 represents the slope with physical meaning of total heat loss coefficient, and t_b is *reference* or base temperature (°C). The $(^+)$ notation indicates that values of the parenthetic term shall be set to zero when it is negative. Using the three parameters heating regression analysis can identify the breakpoint, i.e., change-point (e.g., the red point P shown in Fig. 1a). The 3PH model (Fig. 1a) is appropriate for modelling building energy use that varies linearly with an independent variable over part of its range and remains constant over the remainder.

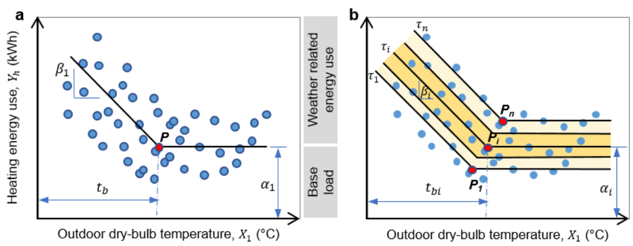

FIGURE 1. 3PH regression models. (a). Least-squares, (b). Quantile regression.

B. STATISTICAL INDICES FOR MODEL EVALUATION

The statistical indices: coefficient of determination; i.e. R-squared (R^2), coefficient of variation (CV), coefficient of variation of root mean square error (CV-RMSE) and normanormalized bias error (NMBE), are commonly used to evaluate 3PH model's performance, i.e. its goodness-of-fit of the predicted values against the measured value, and to describe the statistical characteristics of the model. The four indices used in this study were defined by (4)-(7) [16] and the CV was calculated using (2).

$$CV = \frac{SD(Y_i)}{\bar{Y}_i} \tag{2}$$

where Y_i is the i-th measured heating energy use (kWh), $SD(Y_i)$ denotes the standard deviation of all Y_i. $\hat{Y}_i$ is the corresponding i-th heating energy use predicted by the model (kWh), n is the total number of data points, and $\bar{Y}$ is the mean of the measured heating energy use over the analysis period (kWh). The greater the R^2, the smaller the CV-RMSE and NMBE, the closer the predicted values are to the actual values. Generally, $R^2 > 0.7$ is considered acceptable, indicating confidence in the relationship. The requirement of CV-RMSE (22.5%) and NMBE (±7.5%) for evaluating daily model are interpolated from ASHRAE Guideline 14 [21], which provides recommended values for evaluating monthly and hourly base line models.

III. DATA

A. WEATHER DATA

Three candidate stations are available for South Glamorgan, i.e. Bute Park, Rhoose and St Athan – located in CF1, CF6 and CF64 postcode areas in Cardiff, respectively. Bute Park (Latitude: 51.4878, Longitude: −3.18728, with World Meteorological Organization (WMO) [22] number 037170 was chosen as the source station for this study as it is located within 1.0 km of Cardiff City center. The average distance between the selected buildings and the Bute Park weather station is around 1.83 km. The Bute Park is an AWSHRLY (Automatic Weather Station HouRLY) station that automatically logs weather parameters and reports hourly [23].

B. BUILDINGS' MEASURED ENERGY USE

Three non-domestic buildings located in Cardiff, UK, have been selected to evaluate the proposed method. The selected buildings represent diverse building type, such as community, arts and leisure (CL) – clubs and community centers; education (ED) – primary school; and health (HL) – nursing and care homes. The maximum distance between the weather station (Bute Park, Cardiff) and the selected buildings is less than 3.0km. The detailed characteristics of all case-study buildings are listed in Table 1.

TABLE 1. Case-study buildings characteristics.

Building code	CL	ED	HL
Photos			
Building sectors[1]	Community, arts and leisure	Education	Health
Sub-sectors[1]	Clubs and community centers	Primary school	Nursing and care homes
Floor area (m^2)	772	2437	317
Occupancy number (-)	N/A[2]	387	33
Floor number (-)	1	2	2
Year built (-)	Post 1976	Pre 1976	Post 1976
Occupancy schedule	09:30-21:30 (weekdays) 10:30-17:30 (Saturday) 11:00-15:30 (Sunday)	09:00-17:00 (weekdays	24h (weekdays and weekend)

Note: [1] Buildings are categorized according to the Building Energy Efficiency Survey (BEES) [24]
[2] N/A means that the variable is not available

According to the energy use patterns of the three buildings above, i.e. CL, ED and HL, three types of building use patterns could be defined, i.e. Pattern A, Pattern B and Pattern C. The characteristics of three building use patterns is given in Table 2.

TABLE 2. Characteristics of building energy use patterns.

Energy use pattern	Main characteristics
Pattern A	The building (e.g., clubs and community centers, leisure & sports Buildings) is not occupied for 24 hours in both weekdays and weekends/holidays. There are different durations of energy use between weekdays and weekends/holidays. There is a little gap in energy use between weekdays and weekends/holidays.
Pattern B	The building (e.g., primary school, public office) is not occupied for 24 hours in weekdays and wholly unoccupied in weekends/holidays. Energy use in most weekends/holidays, except for occasional activity days, is for basic requirements, e.g., protecting pipes from freezing in winter. There is a clear gap in energy use between weekdays and weekends/holidays.
Pattern C	The building (e.g., hospital, nursing and care homes) is occupied for 24 hours in both weekdays and weekends/holidays. There are similar durations of energy use between weekdays and weekends/holidays. The energy use in both periods is almost the same.

IV. METHODOLOGY

A. ADVANTAGES OF CPQR

The quantile regression (QR), introduced by Koenker and Bassett [25], is different from the OLS regression on the conditional mean. The QR offers, on a systematic level, a more comprehensive distribution picture of the relationship between variables than OLS [26]. In terms of HDD-based building energy predictions, the advantages of QR are mainly from the following aspects:

First, the QR approach can provide rich information regarding both buildings' energy use and weather. Hence, it can identify energy use anomalies in system behavior (e.g., due to control failure or unrelated-weather occupant behavior) and poor-quality data.

Second, QR offers a natural approach without unnecessary disaggregation of unoccupied and occupied data. It uses different values for the regression coefficients, discussed further (to be shown in Section C).

Third, QR does not specify any distribution for the residuals; therefore, it is distribution free. Accordingly, its flexibility focuses on detecting more subtle relationships between energy use and ambient temperature, dealing with non-normally distributed errors, robustness against outliers, and the ability to detect heterogeneity.

Finally, the studies on base temperatures mostly focused on the average using the conditional mean function. Unlike classical statistical regression methods, the outcomes of QR are not only point estimations, but a full picture of base temperatures showing more informative knowledge under different quantile levels.

B. CHANGE-POINT QUANTILE REGRESSION

QR is a crucial expansion of the empirical mean regression model. It provides a more accurate understanding of the distribution between independent and dependent variables, unlike OLS that only offers mean conditional point relationship. The detailed theory about linear quantile regression could be found in relevant literatures [26]. From the energy signature model, we could know that linear quantile regression cannot well describe the relationship between building energy use and ambient temperature to infer base temperatures, and therefore segmented linear quantile regression may be preferable. Using the piecewise linear quantile regression method a more accurate range of base temperatures for specific buildings could be obtained, while not like one estimated point from the OLS method. Following previous studies on bent line quantile regression [27]–[29], we extensively developed the change-point quantile regression (CPQR) method in this study.

Given a probability τ strictly between 0 and 1, we considered a three-parameter change-point quantile regression model for heating (shown in Fig. 1b), which was derived from (1).

$$Y_i = \alpha_1 + \beta_1 (t_b - X_i)^+ + \mathbf{z}_i^T \boldsymbol{\gamma} + e_i \quad i = 1, \cdots, n \quad (3)$$

where Y_i is the i-th response, X_i is the scalar covariate whose slope changes at the change-point t_b, $\mathbf{z_i}$ is a q-dimensional vector of linear covariates with constant slopes and e_i is the error term whose τ-th quantile is zero conditional on $(X_i, \mathbf{z}_i)$. $\boldsymbol{\gamma}$ is the effect of $\mathbf{z}_i$. Here, t_b is an unknown variable to be estimated. Let $\boldsymbol{\eta} = (\alpha_1, \beta_1, \boldsymbol{\gamma})$. Given t_b, the τ -th quantile of Y_i given X_i and $\mathbf{z_i}$, $Q(X_i, \mathbf{z}_i; \boldsymbol{\eta})$ is

$$Q(X_i, z_i; \boldsymbol{\eta}) = \alpha_1 + \beta_1 (t_b - X_i)^+ + \mathbf{z}_i^T \boldsymbol{\gamma} i = 1, \cdots, n \quad (4)$$

As t_b is known in (4), then, conditional on this known t_b, the best estimate $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta}$ is

$$\hat{\boldsymbol{\eta}}(\tau) = \mathrm{argmin}_{\eta \in \mathbb{R}^{2+q}} \sum_{i=1}^{n} \rho_\tau [Y_i - Q(X_i, \mathbf{z}_i; \boldsymbol{\eta})] \quad (5)$$

where $\rho_\tau(v) = v(\tau - I(v < 0)), 0 < \tau < 1$, is the quantile regression loss function [26]. Here $I(\cdot)$ denotes the indicator function.

Letting $u_i = \max(0, t_b - X_i)$, the quantile conditional on t_b is linear in form, that is,

$$Q_i = Q(X_i, \mathbf{z}_i; \boldsymbol{\eta}) = \alpha_1 + \beta_1 u_i + \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6)$$

As (6) is equivalent to (4), (6) can be

$$\hat{\boldsymbol{\eta}} = \mathrm{argmin}_{\eta \in \mathbb{R}^{2+q}} \sum_{i=1}^{n} \rho_\tau (Y_i - \mathbf{w}_{i,t_b}^T \boldsymbol{\eta}) \quad (7)$$

where $\mathbf{w}_{i,t_\mathrm{b}} = (1, u_i, \mathbf{z}_i)$. Now we can estimate $\boldsymbol{\eta}$ conditional on t_b. As in [27], we adopted $\mathrm{S}(\hat{\boldsymbol{\eta}}|t_\mathrm{b})$ (a function of t_b) as the following equation:

$$S(\hat{\boldsymbol{\eta}} \mid t_b) = \sum_{i=1}^{n} \rho_\tau \left(Y_i - \mathbf{w}_{\mathbf{i},t_b}^T \hat{\boldsymbol{\eta}}\right). \quad (8)$$

Finally, we obtained all values of $\mathrm{S}(\hat{\boldsymbol{\eta}}|t_b)$ corresponding to all values of t_b in the range of X_i's. The estimated change-point value $\hat{t}_b$, the t_b value at which the minimum value of $S(\hat{\boldsymbol{\eta}}|t_b)$ is realized, is calculated through the following equation:

$$\hat{t}_b = \underset{t_b}{\operatorname{argmin}} \, S(\hat{\boldsymbol{\eta}}|t_b) \tag{9}$$

Both asymptotic and bootstrapping methods provide robust results for standard errors and confidence limits for regression coefficient estimates [30]. We used the bootstrap technique for deriving confidence intervals (CIs) of the derivative of a QR model as it was more practical [31]. The n triplets of variables (α_1, t_b, and β_1) with replacement from $\{(Y_i, X_i, \mathbf{z_i})|i = 1, \cdots, n|\}$ were resampled, and these bootstrap samples were then used to re-estimate α_1, t_b, and β_1. We calculated the 95% CIs for parameters α_1, t_b, and β_1 using the bootstrap method [32].

C. METHODOLOGY

The schematic diagram of our methodology for calculating base temperatures and running HDD-based energy use prediction is illustrated in Fig. 2a. Using the method based on CPQR, we can automatically identify the building use patterns. In the step, data pair collection', CPQR is robust enough that it is not affected by outliers. Consequently, it is free of preprocessing in spite of some abnormal points in raw data pair. CPQR at quantiles from 0.05 to 0.95 with 0.05 quantile interval could grasp most characteristics of the distribution variations in relationships between building energy use and weather. In all coefficients, the slope β_1, the coefficient of total thermal loss, is the most vital since rates of change across quantiles in the slope parameter estimates can be used to provide additional information. First, the slope β_1 can reflect directly the ratio relationship between building energy use and weather parameters. Second, the change in the slope β_1 could help to reveal building use patterns (STEP A in Fig. 2a). There are three kinds of variations of slope, leading to three curves for the selected case-study buildings, i.e. linear monotonically decreasing curve, S-curve and constant line (Fig. 2b). The coefficient of variation of slope β_1, CV(β_1) with critical quantile level τ_c (corresponding to point of inflection of fitting line of scatter slope) could be used as criteria index to identify the building use patterns (e.g. Fig. 2b-2). In the next step the base temperatures of buildings with different use patterns would be determined (STEP B in Fig. 2a).

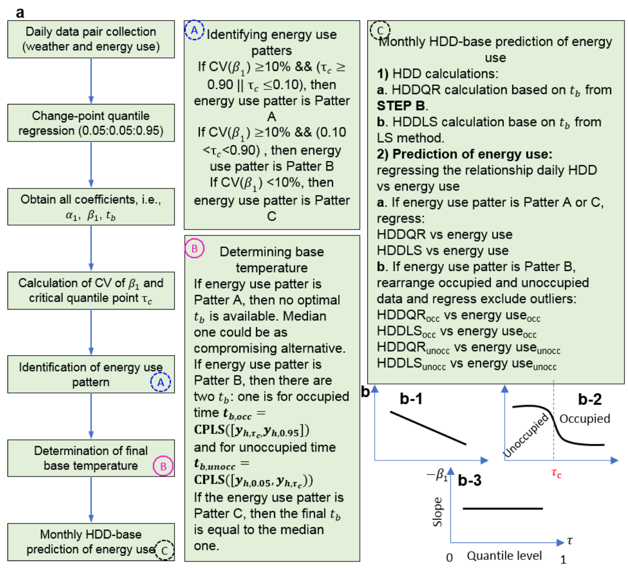

FIGURE 2. **Flowchart of the methodology (a) and the slopes of CPQR of three typical buildings with corresponding energy use patterns (b-1. Pattern A. b-2. Pattern B. b-3. Pattern C.). Note: τ_c is critical quantile point dividing the period of occupied and unoccupied, namely, point of inflection of coefficient *slope* line, which could be solved with mathematical method. In the step A, CV(x) means the coefficient of variations of series x. If τ_c obtained is outside of the range of [0, 1], it means that there is no τ_c. Meanwhile, if the point of inflection τ_c obtained is located at near lowest or highest quantile, it maybe is fake one due to lowest or highest quantile's abnormal impact (Generally sampling variation will increase as the value of approaches 0 or 1 [26].). In the step B, $y_{h,x}$ denotes gas consumption y_h at quantile x by change-point quantile regression (CPQR) approach. CPLS(x_1, x_2) represents the regression results of regressing gas consumption between x_1 and x_2 by change-point least square (CPLS) method.**

For the building with building use Pattern A, the variation of the building energy use regarding to climatic weather looks to be uniform and continuous monotone from lower to higher quantiles, and followed a skewed response distribution. If one or two base temperatures are used for buildings with either energy use Pattern A or B, it would fail to comprehensively describe the relationship between energy use and weather. Consequently, it is always worthy of estimating a range of base temperatures rather than a single one when the building follows Pattern A. However, if only one base temperature is needed, the one from the median quantile regression may be a compromising alternative t_b. In contrast, buildings with Pattern B have a jump variation of quantile slope. There is a critical quantile level τ_c dividing the energy use into real occupied and unoccupied parts which are different between weekdays and weekends/holidays. The real occupied and unoccupied parts are the results of the rearrangement of the full original period. Therefore, there were two base temperatures, $t_{b,occ}$ and $t_{b,\text{unocc}}$, corresponding to base temperatures for occupied period and unoccupied period, respectively. To reduce the risk of underestimating the base temperature, it may be necessary to use the data exclusive of outliers. The gas consumption of extreme quantiles, i.e. above 0.95 quantile and below 0.05 quantile of CPQR fit line, is seen as outliers. Different from the former two building types as mentioned above, buildings with Pattern C indicated homogeneous model characteristics. In this case, a homogeneous variance regression model associated with ordinary least squares regression is sufficient for the building except impact from some outliers on the regression results. Under this condition, we focused on the median rather all quantiles (percentiles) due to the indication of some form of homogeneity. Median relationship results could potentially represent a whole general phenomenon for the relationship of gas consumption and temperature. After STEPs A and B, STEP C is for HDD based energy use prediction. Both building energy predicted and actual results of CPLS (change-point least square method) based t_b and CPQR (change-point quantile regression method) based t_b were provided for comparison.

V. EMPIRICAL RESULTS AND DISCUSSIONS BASED ON CASE-STUDY BUILDINGS

According to the methodological steps defined in Fig. 2 and prior to the degree day analysis, the base temperature was determined by three parameter heating (3PH) based on daily gas consumption and ambient temperature, followed by a comparison of actual and predicted monthly energy use. For the evaluation of final results, we implemented the above steps with CPLS-based as well as CPQR-based approaches. What follows in the following subsections are detailed discussions and analysis of our results.

A. CPQR ANALYSIS ON CASE-STUDY BUILDINGS

To test and validate the proposed methodology based on Equation (3), the change-point quantile regression models were fitted to the above three sets of data collected from three different buildings under a full series of 19 specific quantiles of τ, ranging from 0.05 to 0.95 with a 0.05 increment. The piecewise line patterns for the three buildings are illustrated in Fig. 3a-c, with corresponding coefficients presented in Fig. 3d-f, respectively. From Fig. 3 demonstrates that the response distribution patterns vary according to the quantile, and this reveals that different building types have different energy use patterns during different periods.

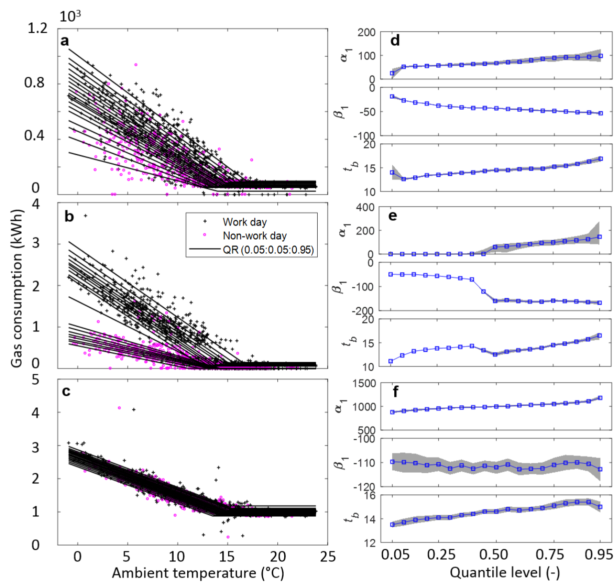

FIGURE 3. Quantile regression (left panels a-c) and coefficients (right panel d-f) of four-year's data pair: gas consumption vs ambient temperature. a and d: CL building, b and e: ED building, c and f: HL building. Note: The effects of ambient temperature on the gas consumption across quantiles τ covering from 0.05 to 0.95 with a 0.05 quantile interval. In the right panels, the squared blue lines show the varying effects across quantiles, with their respective bootstrapped confidence intervals (95%) displayed as shadowed areas. The coefficients in subplots from upper to lower are constant (α_1, base load), slope (β_1, total thermal loss coefficient) and change point (t_b, base temperature), respectively.

Particularly, the changes of the slopes at different quantile levels reveal the heterogeneity and homogeneity of gas consumption dependency on ambient temperature. Since buildings have different uses, a detailed analysis is conducted on the quantile regression results of the three buildings.

For the CL building (Fig. 3a and d), the slope with respect to daily ambient temperature was decreasing with a constant monotonic speed in the change-point quantile regression model. The ratio between the maximum and the minimum of the slope was 2.86, with a CV of 0.22 (Table 3). It reflects the wide gap between different quantile levels. Since the estimated slope parameters increase with the quantile, these estimations reflect the same increasing dispersion, or heteroscedasticity. This implies that the gas consumption for upper quantiles increased more rapidly than that for lower quantiles, due to the decrease in ambient temperature. However, there is no clear boundary between weekdays, weekends and holidays. From Table 2 it could be found that the building was being used not only on weekdays, but also on weekends and holidays. This is likely to be the reason why no significant boundary has been observed between weekdays and weekends/ holidays for this building type. A plausible fact is that different day of the week has different slopes reflecting different actions (such as thermostat settings). This may partly reflect the heterogeneity characteristics between gas consumption and ambient temperature. Because of this, we did not use the OLS method to regress the relationship between independent and dependent variables, as OLS regressions are not suitable for dependent variables that have non-normal distributions. A range of base temperature should be recommended for the building type.

For the ED building (Fig. 3b and e), the slope with respect to daily ambient temperature was also changing in the change-point quantile regression model, but with a variable speed. This is different from the CL building discussed above. The ratio between the maximum and the minimum of the slope was bigger than that of the CL building, and it was 3.38, with a CV of 0.45 (Table 3). Similar to the CL building, it also shows wide gaps among different quantile levels. Furthermore, there was a quick change from a certain quantile level (i.e. 0.40). After comparing under the same quantile level with Fig. 3a, it was found that the quantile level of 0.45 is likely to be the critical point that separates weekdays and weekends/holidays. Before and after the critical quantile level, there were two different nearly constant slopes, reflecting some form of homogeneity. Due to its function as a primary school, the ED building consumed more gas in occupied time (weekdays) and less in unoccupied time (weekends/holidays). Another phenomenon can be found that in some weekends and holidays the energy use still maintained at a high level, like weekdays. The reasons may come from two aspects. First, on weekends or holidays there may have no person working, but the heating system should continue to work to remain the minimum temperature indoors. This is common for unoccupied conditions. Second, some activities may be held during weekends and holidays when the heating system was running at normal set-points. This is not a normal unoccupied holiday pattern, but an occupied pattern. Therefore, for this type of building, two pattern regressions should be done separately for weekdays and weekends/holidays. Before we do that, we should identify which energy use was for occupied period and which was for unoccupied period. Using the traditional method (e.g. date) to differentiate the occupied and unoccupied energy use is no longer reliable. The quantile regression is a suitable one.

For the HL building (Fig. 3c and f), although the slope was variable, the amplitude of the variation was very small, with a CV less than 0.01 (Table 3). This means the effect of ambient temperature on the gas consumption across the whole quantile level τ from 0.05 to 0.95 was not significantly different. It is also well-known that the median is more robust to outliers than the mean. Therefore, for this situation, the result under median quantile regression could be used.

TABLE 3. Characteristics of β_1 and determination of critical quantile τ_c.

Case building	CV of β_1	Ratio max to min of β_1	Critical quantile
CL	0.22	2.86	None[1]
'ED	0.45	3.38	0.45
HL	0.01	1.03	None

[1] Note: no critical quantile τ_c is available

For both CL and ED buildings, the confidence interval (CI) of all parameters under upper quantile levels were wider than those under lower quantile levels. This implies that there may be other variables that have not been considered under upper quantile levels. This also means that lower quantiles of energy use is very dependent on ambient temperature. In contrast, the CI of β_1 of the HL building was wider than those of the CL and ED buildings. The implications are that multiple-variables rather than ambient temperature only should be used, though the energy use data of the HL building appeared tidier.

B. DETERMINATION OF BASE TEMPERATURES

Due to different energy use patterns identified for different case-study buildings, base temperatures were determined using different methods (STEP B in Fig. 2). Prior to the determination of base temperatures, the building energy use patterns were identified according to statistical information of β_1 and critical quantile τ_c (Table 3). The determined base temperatures with fitting equation and R^2 for the three case-study buildings have been illustrated in Fig. 4. As mentioned in Section IV, building energy use patterns should be identified before the determination of base temperatures. As an example, the CL building was identified as building use Pattern A, for which there was no clear boundary between weekdays and weekends/holidays. The base temperatures (Fig. 4a and Fig. 4d) were determined as 14.9°C and 14.5°C using CPLS-based ($R^2 = 0.69$) and CPQR-based ($R^2 = 0.68$) methods, respectively. Both R^2 values were lower than 0.70. Both mean (LS) and median (QR) regressions provided measures of central tendency for the relationship. However, the median regression has the additional advantage of not being sensitive to outliers in raw data distribution. The distribution was scattered because building gas consumption was affected not only by ambient temperature, but also other environmental parameters, such as solar radiation, the operation of devices and occupant behavior. Therefore, the univariate regression is not sufficient here and multi-variable regressions are needed for the CL building. This is outside our topic and will be addressed in future work. For the ED building, there were two base temperatures for different periods due to different energy use patterns. From Fig. 4a weekends and holidays were with high gas consumption while it was low for weekdays. This resulted in the fact that the base temperature of weekdays (13.9°C) was less than that for weekends/holidays (15.1°C). The regression result from the energy use data during weekends/holidays was biased towards the top right, which may be because of the presence of some occupied data. Accordingly, real energy use periods for occupied and unoccupied periods were identified to obtain the critical quantile by the CPQR-based method. A critical quantile τ_c (0.45) was identified, as shown in Fig. 5. The results of base temperature were found to be 14.3°C and 13.8°C for occupied and unoccupied periods(Fig. 4e), respectively. The $R^2$from the CPQR-based method (Fig. 4e) was higher than that from the CPLS-based approach (Fig. 4b). For the HL building, similar results (14.5°C and 14.6°C, Fig. 4c and f) were obtained, which implied that both CPLS-based method and CPQR-based method were suitable for buildings with energy use Pattern C. We opted for CPQR as the method is more robust to outliers in the raw data.

C. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORK

The CPQR-based method proposed in the paper has its own advantages and disadvantages. The main disadvantage of this approach, and the reason that it was not widely used in the past, is that it requires higher computing resources with respect to both time and hardware. However, most current computers can only run programs using alternative methods in a reasonable amount of time, but not the CPQR method. Indeed, more studies could be performed to gain insights into the capabilities and limitations of the proposed method. First, the above work is based on case-study buildings, hence case specific. The universality of the method proposed should be investigated further. Second, we should answer the question: which parameter(s) is statistically more important for predicting building energy consumption? Energy use is known to be associated with a wide range of subsequent weather variations, which are usually not controllable, and is also dependent on non-climate variables, which are often controllable. The multi-regression method based on CPQR energy analysis could examine some key consumption drivers such as solar radiation and occupant behavior for the building with Pattern A, and be further helpful to the development

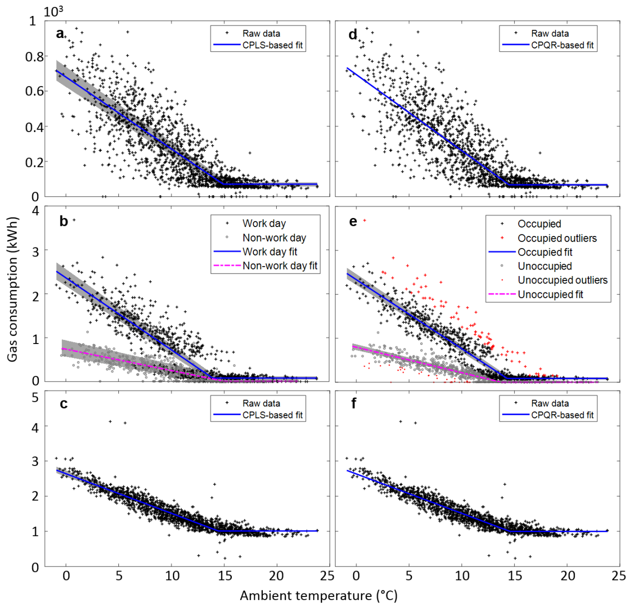

FIGURE 4. **Comparisons of base temperature determined using CPLS-based (left panels) and CPQR-based (right panels) method for CL building (a and d), ED building (b and e), and HL building (c and f). The fitting model with R^2 is also present in each subplot. Note: The gray shaded area around the regression lines is the confidence interval (95%) of the regression. *Non-work days* represents weekend and holidays.**

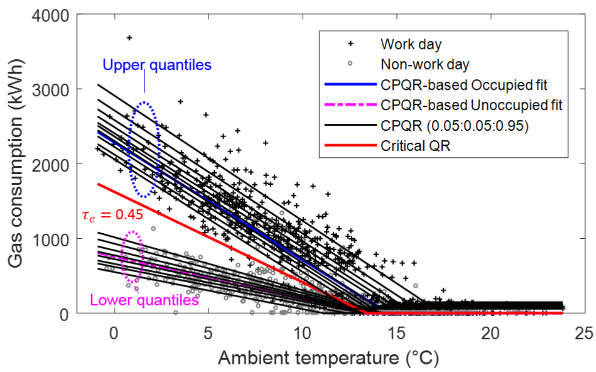

FIGURE 5. **Determination of critical quantile (τ_c) for ED building using CPQR method.**

of predictive algorithms to estimate energy consumption and savings.

VI. CONCLUSION

This research conducted a comprehensive investigation on the energy use patterns in buildings, and proposed a novel change-point quantile regression (CPQR) based method to automatically identify building energy use patterns. Our method offers a more comprehensive picture of suitable base temperatures across quantiles compared to the single base temperature obtained via the conventional method, change-point least square (CPLS). The contribution of this paper can be considered as threefold. First, the automated robust identification of building energy use patterns. Second, the ability to disaggregate energy use between occupied and unoccupied periods using CPQR can benefit building energy management and optimization strategies of supervisory control, ultimately leading to improved building performance. Third, a comprehensive understanding of the distribution of the relationship between building energy use and weather could be obtained through CPQR.

REFERENCES

[1] *International Energy Outlook 2014*, U.S Energy Inf. Admin., Washington, DC, USA, 2014.

[2] *Key World Energy Statistics*, Int. Energy Agency, Paris, France, 2014.

[3] J. L. Reyna and M. V. Chester, "Energy efficiency to reduce residential electricity and natural gas use under climate change," *Nature Commun.*, vol. 8, May 2017, Art. no. 14916.

[4] H.-X. Zhao and F. Magouls, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012.

[5] *ASHRAE Handbook: Fundamentals*, Amer. Soc. Heating, Refrigerating Air-Conditioning Eng., Atlanta, GA, USA, 2013.

[6] *Degree-Days: Theory and Application*, Chartered Inst. Building Services Eng., London, U.K., 2006.

[7] R. C. Sonderegger, "A Baseline model for utility bill analysis using both weather and non-weather-related variables," *Trans.-Amer. Soc. Heating Refrigerating Air Conditioning Eng.*, vol. 104, pp. 859–870, Jun. 998.

[8] Z. Verbai, Á. Lakatos, and F. Kalmár, "Prediction of energy demand for heating of residential buildings using variable degree day," *Energy*, vol. 76, pp. 780–787, Nov. 2014.

[9] *TM41: Degree Days: Theory & Application*, Chartered Inst. Building Services Eng., London, U.K., 2006.

[10] *Degree-Days: Theory and Application (TM41)*, Chartered Inst. Building Services Eng., London, U.K., 2006.

[11] K. Lee, H.-J. Baek, and C. Cho, "The estimation of base temperature for heating and cooling degree-days for South Korea," *J. Appl. Meteorol. Climatol.*, vol. 53, no. 2, pp. 300–309, 2014.

[12] M. Shin and S. L. Do, "Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate," *Energy Buildings*, vol. 110, pp. 57–70, Jan. 2016.

[13] C. Ghiaus, "Experimental estimation of building energy performance by robust regression," *Energy Buildings*, vol. 38, no. 6, pp. 582–587, 2006.

[14] D. Lindelöf, "Bayesian estimation of a building's base temperature for the calculation of heating degree-days," *Energy Buildings*, vol. 134, pp. 154–161, Jan. 2017.

[15] Yuna Zhang, Z. O'Neill, B. Dong, and G. Augenbroe, "Comparisons of inverse modeling approaches for predicting building energy performance," *Building Environ.*, vol. 86, pp. 177–190, Apr. 2015.

[16] Q. Meng and M. Mourshed, "Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures," *Energy Buildings*, vol. 155, pp. 260–268, Nov. 2017.

[17] M. Molina-Solana, M. Ros, M. D. Ruiz, J. Gómez-Romero, and M. Martin-Bautista, "Data science for building energy management: A review," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 598–609, Apr. 2017.

[18] B. Cade, J. W. Terrell, and R. L. Schroeder, "Estimating effects of limiting factors with regression quantiles," *Ecology*, vol. 80, no. 1, pp. 311–323, 1999.

[19] U. Jensen and C. Lütkebohmert, "Change-point models," in *Encyclopedia of Statistics in Quality and Reliability*, Chichester, U.K.: Wiley, 2007.

[20] J. Kissock, J. Haberl, and D. Claridge, "Inverse modeling toolkit (1050RP): Numerical algorithms for best-fit variable-base degree-day and change-point models," *ASHRAE Trans.*, vol. 109, no. 18, pp. 425–434, 2003.

[21] *Measurement of Energy and Demand Savings (Guideline 14)*, Amer. Soc. Heating, Refrigerating Air-Conditioning Eng., Atlanta, GA, USA, 2002.

[22] *World Meteorological Organization*. Accessed: Feb. 10, 2017. [Online]. Available: http://www.wmo.int/pages/index_en.html

[23] Met Office, (2006). *MIDAS: UK Hourly Weather Observation Data*. NCAS British Atmospheric Data Centre. Accessed: Jan. 12, 2017. [Online]. Available: http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c

[24] Building Energy Efficiency Survey. *Building Energy Efficiency Survey 2014-15: Overaching Report*. Accessed: Nov. 16, 2016. [Online]. Available: https://www.gov.uk/government/publications/building-energy-efficiency-survey-bees

[25] R. Koenker and G. Bassett, Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[26] R. Koenker, *Quantile Regression*. New York, NY, USA: Cambridge Univ. Press, 2015.

[27] C. Li, Y. Wei, R. Chappell, and X. He, "Bent line quantile regression with application to an allometric study of land mammals' speed and mass," *Biometrics*, vol. 67, no. 1, pp. 242–249, 2011.

[28] L. Zhang, H. J. Wang, and Z. Zhu, "Composite change point estimation for bent line quantile regression," *Ann. Inst. Stat. Math.*, vol. 69, no. 1, pp. 145–168, 2017.

[29] C. Li, N. M. Dowling, and R. Chappell, "Quantile regression with a change-point model for longitudinal data: An application to the study of cognitive changes in preclinical alzheimer's disease," *Biometric*, vol. 71, no. 3, pp. 625–635, 2015.

[30] R. Koenker and K. Hallock, "Quantile regression: An introduction," *J. Econ. Perspect.*, vol. 15, no. 4, pp. 43–56, 2001.

[31] L. Hao and D. Q. Naiman, *Quantile Regression*, Thousand Oaks, CA, USA: Sage, 2007.

[32] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, 1979.

QINGLONG MENG received the B.S. degree in automation and the M.S. degree in control theory and engineering from the Xi'an University of Architecture and Technology, Xi'an, China, in 2002 and 2005, respectively, and the Ph.D. degree in power engineering and engineering thermophysics from Xi'an Jiaotong University, Xi'an, China, in 2010. From 2016 to 2018, he was a Visiting Academic with the School of Engineering, Cardiff University, Cardiff, U.K. He is currently an Associate Professor with the School of Civil Engineering, Chang'an University, Xi'an, China. His research interests include smart buildings, building energy analytics, and HVAC demand response.

MONJUR MOURSHED received the B.Arch. degree from the Bangladesh University of Engineering and Technology, Dhaka, in 1999, and the Ph.D. degree in civil and environmental engineering from the National University of Ireland, Cork, in 2006.

Since earning his first degree, he held various academic appointments with the Bangladesh University of Engineering and Technology, as a Lecturer, from 1999 to 2004, the University of Central Lancashire, as a Lecturer and a Senior Lecturer, from 2003 to 2007, and Loughborough University, as a Lecturer, from 2007 to 2013. He was a Senior Lecturer from 2013 to 2016 and has been a Professor of Sustainable Engineering with the School of Engineering, Cardiff University, U.K., since 2016. He has authored over 80 articles and co-chaired two international conferences. His research interests lie in the broader field of smart, resilient, and sustainable buildings and cities-linking environment (sustainable building and urban design), energy (building and district energy systems, smart grids, renewable energy, and decarbonization), and climate (energy meteorology and climate change).

SHEN WEI received the B.Eng. degree in building services engineering from the Beijing University of Technology, China, in 2006, and the M.Sc. degree in mechatronics engineering from Southern Denmark University, Denmark. He is currently pursuing the Ph.D. degree with Loughborough University, focusing on occupant adaptive behavior in buildings, including monitoring, modelling, and simulation. From 2015, he was a Senior Lecturer in building services engineering with Northumbria University, Newcastle, U.K. He has been a Lecturer in building services engineering with the Bartlett School of Construction and Project Management, University College London, since 2017. In the past five years, he has published over 30 peer-reviewed articles, including 12 journal papers. His main research interests include occupant behavior, thermal comfort & thermal storage (e.g. PCM) in buildings.