University of Reading

A comparative analysis of building energy estimation methods in the context of demand response

Article

Accepted Version

Curtis, M., Torriti, J. and Smith, S. T. (2018) A comparative analysis of building energy estimation methods in the context of demand response. Energy and Buildings, 174. pp. 13-25. ISSN 0378-7788 doi: https://doi.org/10.1016/j.enbuild.2018.06.004 Available at http://centaur.reading.ac.uk/78070/

Published version at: https://www.sciencedirect.com/science/article/pii/S0378778817336393

To link to this article DOI: http://dx.doi.org/10.1016/j.enbuild.2018.06.004

Publisher: Elsevier

Title: A Comparative Analysis of Building Energy Estimation Methods in the Context of Demand Response

Target Journal: Energy and Buildings

Authors: Mitchell Curtis [1, *], Jacopo Torriti [2], Stefan Thor Smith [2]
[1] Technologies for Sustainable Built Environments Centre, University of Reading, Reading, UK
[2] School of the Built Environment, University of Reading, Reading, UK
[*] Corresponding author at: Technologies for Sustainable Built Environments Centre, University of Reading, Reading, UK. E-mail m.r.curtis@pgr.reading.ac.uk

Abstract:
A critical element of assessing a building's suitability for Demand Side Response (DSR) is understanding its turndown potential to ensure that DSR participation will be financially viable. While research has been undertaken on site level DSR estimation methods, there is currently no research that compares the outcomes of these methods. This paper compares four non-domestic energy estimation methods used for understanding the DSR potential of electrical appliances in a building to provide insights about uncertainty levels based on input requirements. Each method is deployed to estimate the DSR potential of HVAC chiller assets at two UK hotels over two years. The results show the methods have a range of error levels from the highest Mean Average Percentage Error (MAPE) of 159% to the lowest MAPE of 39%. The input requirements followed a general trend of more complex informational inputs resulting in lower error values. The outcomes of this research enable users to make informed decisions in selecting DSR estimation methods based on information availability and acceptable estimation error levels.

Highlights
- Four DSR estimation methods were evaluated using empirical data from two hotels
- The DSR estimation methods were found to have a wide range of error levels
- The comparisons of methods allows for informed selection of a DSR estimation method based on available input information

Keywords:
- Demand Side Response
- Estimation
  Comparative Analysis
- Method Review
- Electricity demand

1 Introduction

Demand Side Response (DSR) programmes generally require a detailed understanding of the turndown potential of participating buildings to accurately forecast DSR capacity for electricity system balancing. This detail is needed as DSR programmes will apply penalties if contracted levels of turndown are missed. As an example, the UK Short Term Operation Reserve (STOR) programme requires participants to provide a guaranteed turndown kW amount for set periods of time of up to 14 hours per day (National Grid, 2016). If STOR participants underdeliver by more than 5%, then financial penalties are applied and progressively increased with the potential for ultimately removing non-performing participants from the programme if they fail in meeting guaranteed turn down levels too many times. The severity of penalties will vary by country and DSR programme. For example, the American San Diego Gas & Electric programme has a low severity based on payments being reduced proportionally to the contracted amount delivered (SDG&E, 2016). Whereas the Spanish programme is very strict with exclusion if the site fails to meet their obligations twice (SEDC, 2017). This means that correctly determining the long-term DSR potential of a building is important for appraising its suitability for DSR.

As DSR aggregators play a key role in provide 80% of DSR capacity (SEDC, 2017), this research has focused on the estimation methods aggregators apply when determining the turndown potential and suitability of buildings for DSR. DSR aggregators operate by combining small flexible loads from multiple buildings into a virtual single load and take responsibility for managing the DSR process. Research into how aggregators decide if a building is suitable shows that the key assessment tasks focus on determining the long-term DSR potential of a buildings' assets (Curtis, 2017). Therefore, the ability to correctly analyse a building's DSR potential is a critical element of an aggregator's business process. This is expressly important when dealing with small to medium enterprises with smaller overall levels of DSR potential as the ability to lowering the contracted amount of DSR to avoid penalties due to estimation uncertainty is limited. While an aggregator can perform building surveys to gain a detailed understanding of a building's DSR potential, surveys have a time and cost impact and therefore are normally only undertaken once an initial desktop assessment has been completed. However, performing a desktop assessment to determine a building's potential is often difficult as detailed usage information (from sub-meters for example) about the individual electrical assets that are being assessed for DSR is normally unavailable (Merry, 2017). Instead, the only information normally available is the building's overall electricity usage as recorded at half hourly (UK standard practice) or similar intervals by the building's utility supplier. Half hourly information will provide a usage profile that can be used for estimating the building's DSR potential and suitability if all electricity demand from the grid is reduced by either using backup generators or turning off all assets. For buildings that can only turndown a limited subset of assets for DSR, a building level profile is unable to provide the individual assets' usage patterns needed to understand their suitability for DSR. To gain this necessary level of detail requires additional analysis to try and determine what proportion of the building's usage is represented by individual assets.

Research on understanding energy usage in buildings is extensive, with a review by Borgstein, Lamberts, & Hensen (2016) identified five categories (Engineering calculations, Simulation, Statistical, Machine Learning, and Other) that each contained multiple approaches. However, research on application of these approaches for DSR estimation is limited and is influenced by whether the estimation is for implicit or instead explicit DSR (SEDC, 2016). Implicit DSR covers price-based measures, whereby demand might be reduced based on users responding to electricity price signals (for example, temporarily high electricity prices that encourages reduction in usage to reduce costs).

Explicit DSR covers incentive-driven measures, whereby demand reduction is specifically requested based on an external signal (for example, demand is reduced temporarily based on a site or its appliance receiving a signal in return for financial compensation for participation). As implicit DSR relies upon optional participation, research into estimating reduction potential focuses on how groups with similar DSR assets behave in response to different pricing signals, for the purposes of gaining an understanding of their combined potential. This is illustrated in research by Shen et al. (2016) where a genetic algorithm is used to estimate the DSR potential for a group of buildings based on time of use and dynamic pricing signals. The authors showed that if each building responds independently to pricing signals, then this can cause higher peak demand usage and therefore recommended that responses are coordinated across similar groups of buildings to achieve the desired peak reduction. Similarly, Chassin & Rondeau (2016) utilised the Random Utility Model to understand the potential provided by groups of fast-acting demand response loads under real-time pricing. In contrast to implicit DSR's reliance upon optional participation, explicit DSR participation is established by contract and the application of penalties where sites fail to respond to a specific reduction request or do not deliver pre-agreed levels of usage reduction. This means that estimation methods for explicit DSR focus on assessing the likely long-term potential of specific buildings to ensure that contractual commitments can be met. As 80% of DSR is currently provided by aggregators, who rely upon explicit DSR, this paper focuses on comparing only energy estimation methods used for explicit DSR (SEDC, 2017).

The majority of contributions to the research field of explicit DSR have originated from the Lawrence Berkeley National Laboratory - Demand Response Research Center (DRRC, 2017). Their research into DSR has covered several areas including methods for assessing the DSR potential of buildings. To help improve the assessment process the DRRC developed the Demand Response Quick Assessment Tool (DRQAT) (Yin & Black, 2015) which uses the EnergyPlus whole building energy simulation program (U.S. Department of Energy, 2017) to predict DSR potential using predefined building models and a limited set of user selectable variables. While the DRQAT program helps to make the assessment process easier, it introduces other limitations, notably that it will only work for predefined building models which are currently offices and retail buildings based in California. They also recognise that are still many input uncertainties like operational behaviour and space loads that are hard to capture in the DRQAT model. To overcome these uncertainties, they use metrics of peak demand (kW), absolute demand savings (kW), and relative demand savings (%) to compensate for differences in actual and forecasted load shapes. The DRRC have also looked into understanding the predictors that influence how well a building will perform when enabled for DSR (Mathieu et al., 2010; Piette et al., 2011). This research showed that the level of turndown potential could be linked to temperature if the DSR assets demonstrate varying levels of usage based on external weather conditions with prediction uncertainty being approximated using the standard error. The limitation of using this approach for assessing a building is the need for the building to have already been involved in DSR to have access to event outcomes for analysis. Another assessment approach proposed by Panapakidis et al. (2014) is to cluster electricity usage of a building into profiles that can then be used to ascertain DSR turndown opportunities based on variance between the profiles. They try to reduce uncertainty by testing a range of cluster lengths to find the optimal number to use that minimises the overall sum of squared errors. This method has the advantage of only needing the building's overall electricity usage records, yet is limited by the assumptions required when deciding what the differences between profiles mean in terms of individual asset usage. There are other proprietary commercially developed analysis methods that have not been published. One such method has been provided by an aggregator in association with this research. They have two approaches when performing building asset assessment for DSR. The first approach assumes that the asset will work at a set level all year. To help reduce the uncertainty of this estimation a second approach is used that analyses the building's overall electricity records for a year to create a baseload usage amount for 95% of the time. The aggregator then takes a proportion of this 95% to represent the asset usage. Using the baseload value reduces uncertainty by knowing that at least this amount of electricity is being used 95% of the time and therefore taking a proportion of it prevents over estimating the assets potential usage. The major limitation of both approaches is the assumed consistent usage of the asset across the year, which they recognise, but still use the method to provide an initial understanding of anticipated potential before deciding on further investigations.

The issue that faces aggregators and anyone trying to perform DSR estimations using these methods is knowing which one to use and how they compare in terms of uncertainty and cost to undertake. Therefore, the aim of this paper is to provide an understanding of uncertainty levels in current non-domestic DSR potential estimation methods based on the input requirements. By understanding the uncertainty levels and costs of DSR estimation methods this research hopes to increase usage of DSR from businesses that are currently excluded due to risk aversion resulting from not knowing the level of estimation uncertainty. The research is undertaken by examining and applying four DSR estimation methods to two UK hotels as described in Section 2. Section 3 sets out the research results and discusses these findings. Section 4 concludes by summarising the implications of this research.

2 Methods

Four DSR potential estimation methods were applied to two medium-sized UK hotels (~200 rooms) to evaluate outcome uncertainty against the level of information required for estimation. The four methods are: asset assessment; baseline comparison; historical event analysis; and building energy modelling. Figure 1 provides an overview of the explicit DSR estimation methods reviewed in this paper, including the primary data and parameter inputs and the analytical approaches used. The methods are to be used as part of an initial desktop assessment to determine the potential DSR of a building or business. The assessment provides a decision on whether further assessment or inclusion of the business in a DSR aggregation programme is valid. All methods estimate the half-hourly kW usage profile of electrical assets over a one-year period to assess if sufficient DSR potential exists. To explain how the methods were used and compared this section is divided into seven subsections. The first section describes the comparison of estimation method outcomes, followed by four sections describing the input requirements and calculation steps for each estimation method. Section six describes the sensitivity analysis approach used to highlight the influence of input parameter uncertainties on method outcomes. Finally, section seven describes the approach used to calculate the cost of using each method.

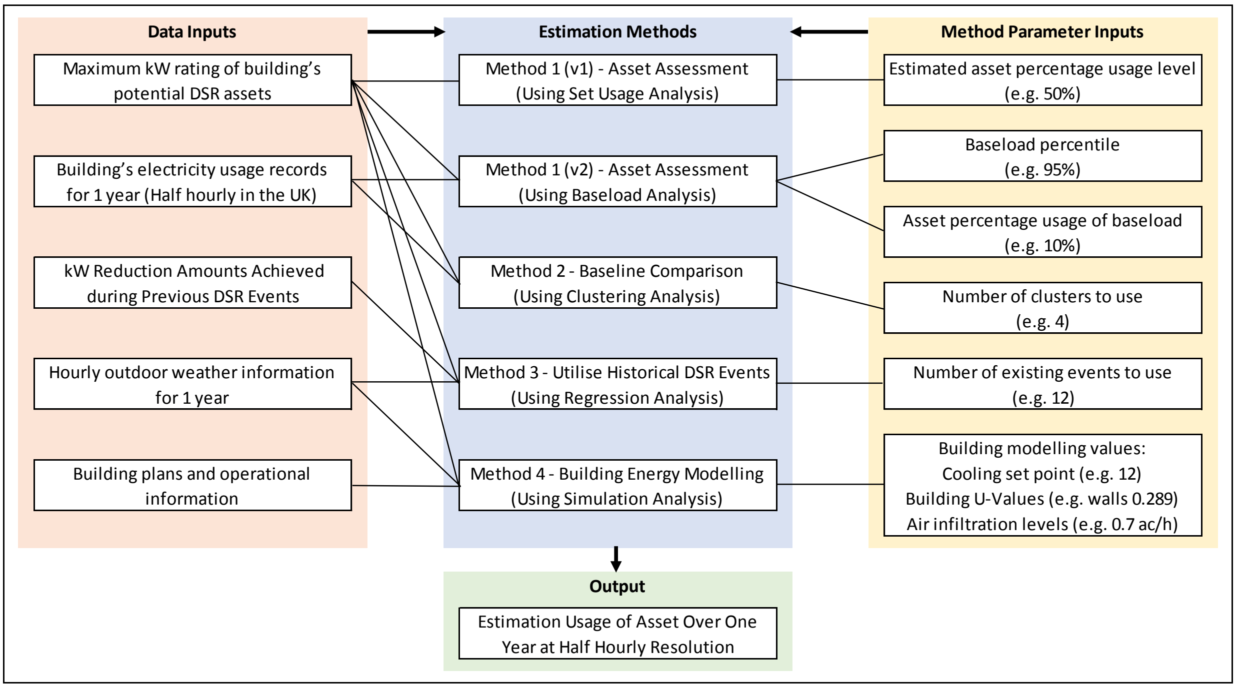

Figure 1 – Overview of DSR Estimation Methods

2.1 Comparison Approach

This comparison of estimation methods was undertaken by using each method to determine the usage profiles of HVAC chillers located at two UK hotels. Chillers are large centralised assets that cool water for distribution around each hotel's HVAC system to provide space cooling that were identified by Curtis et al. (2018) as being suitable for DSR due to the flexibility they offer through their ability to be temporally turned off without impacting end-users. The hotel chillers have a maximum rating of 333 kW for Hotel 1 and 290 kW for Hotel 2. The two hotels have been chosen due to having access to detailed information about each building's overall electricity usage as well as high-quality sub-metered electricity usage data for the chillers during the years 2013 and 2016 for Hotel 1 and 2015 and 2016 for Hotel 2. The sub-metered data enables a direct comparison of the estimation method outcomes against actual usage. While chillers are used as an example of an electrical appliance with DSR potential in this paper, its purpose is not to assess the suitability of chillers for DSR. Instead, the aim and focus of this research is to compare methods for estimating the potential levels of electricity usage by assets with potential for explicit DSR programmes, of which chillers are only one example. The resulting usage estimates for chillers, as a sample appliance, can then be used as an input for determining the specific DSR potential of a building based on the appliances characteristics and intended DSR programme requirements. The application of the estimate to a DSR programme is not covered in this paper as this is dependent on the ability of an appliance to meet specific programme requirements. Therefore, evaluation of the estimations is kept independent by using the Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE) methods.

The MAPE values provide an overall indication of the level of difference between the actual and predicted results while the MBE values indicate the direction of error with positive and negative results indicating over estimation and under estimation respectively. These methods were selected as De Gooijer & Hyndman (2006) define them as the most common measures to use for time series evaluation as they provide an easy to understand percentile value to indicate the level of forecasting error that can be used to compare uncertainty across the four estimation methods. They are also deemed suitable based on their general usage across the literature on DSR estimation methods (Aman et al., 2016; Larsen et al, 2015).

2.2 DSR Estimation Method 1 - Asset Assessment

The asset assessment method is based on a review of current estimation approaches undertaken at a UK DSR aggregator. This is the simplest of the four methods as it is based on using very limited information with two variations to the approach as follows (see Appendix A for detailed calculation steps):

- *Variation 1 – Minimum Information:* This approach uses only the maximum kW rating of the asset being assessed. The expected kW usage level of the asset across the year is calculated as a set percentage of the maximum rating. The set percentage can vary based on the assessor's prior knowledge of the asset type and building.
- *Variation 2 – Utilise Baseload Calculation:* This approach uses the building's overall electricity usage records over one year (in the UK this is provided in half-hourly intervals) to calculate its baseload usage. The baseload amount is calculated for each half-hourly period by taking all usage records for each period (i.e. 365 usage records for the 00:00 to 00:30 half-hour period), ordering the records by value, then finding the 5th percentile value. This provides half hour electricity usage values that the building will use at least 95% of the time over the year and is therefore classified as the baseload. The expected kW usage level of the asset across the year is then calculated by taking a percentage share of the baseload that is attributed to the asset to be used in DSR. Again, the percentage will be set according to prior knowledge of this type of asset and building.

2.3 DSR Estimation Method 2 - Baseline Comparison

The second estimation method utilises clustering techniques to identify DSR opportunities through comparison of each building's different usage profiles over a year. This method works on the basis that a building has different usage profiles throughout the year, and once profile clusters are identified, representative profiles of each cluster can be used to ascertain DSR turndown opportunities based on variance between the profiles. Panapakidis et al. (2014) reviewed a selection of clustering methods for electricity load curve analysis of buildings and identified that the k-means method offers a balanced approach for finding appropriate clusters that would be suitable for understanding building energy efficiency opportunities, including for DSR. However, they did not actually provide specific DSR estimation outcomes for the test building. Research by Van Wijk et al. (1999) also looked into how to use clustering techniques to identify patterns and trends on multiple timescales (days, weeks, seasons). They found that using k-means and then associating the resulting clusters to the different timescales allowed for identification and exploration of usage profiles. Their technique succeeds in identifying weekend vs weekday profiles and other significant periods, such as holidays. These clustering techniques show that building energy usage normally follows a small set of similar profiles. By identifying these profiles, it is then possible to understand different usage levels, which can then potentially be used to derive DSR estimations based on the business type.

The k-means cluster method is used for the baseline comparison (Sayad, 2017). The clustering method works by first selecting how many groups the usage dataset will be clustered into. For each group, a random point within the dataset is selected and deemed the centroid value. Each value in the dataset is assigned to the closest centroid. The mean of the values for each centroid is then calculated. The centroids are then moved to the mean position and the values are reassigned to the now closest centroids. This process is repeated until a pre-defined number of interactions is achieved or the level of centroid position change reaches a set tolerance. The number of clusters for the baseline comparison will vary for each building. One approach for determining the optimum number of k-means clusters to use is called 'elbow' method. This method works by repeating the k-means method using a range of clusters to determine each clusters percentage of variance. The percentage of variance (dependent variable) is plotted against number of clusters (independent variable) in order to find the 'elbow' of the curve that signifies the optimum number of clusters, as adding more will have limited benefit in reducing variance (Ketchen & Shook, 1996). The k-means elbow identification process is undertaken for each hotel's electricity usage data. The data within each cluster is then averaged by half-hourly period. The half-hourly averages in each cluster are then used to generate daily profiles at half-hourly resolution for each cluster of each hotel. Figure 2 provides an example of the daily profiles developed for the four identified clusters of a hotel.

Using the profiles to estimate DSR requires informed assumptions about what the profiles represent based on available information about the business. For the case of hotels, as in this study, information on energy sources related to heating and cooling (gas for heating, electricity for cooling), industry studies/reports on proportional breakdown of electricity use identifies that HVAC demand typically accounts for 34% of electricity demand in UK hotels (CIBSE, 2012). The consistent daily profiles of demand across all days of a week, consistent annual occupancy profiles found in hotels, and a high proportion of HVAC related demand provide the basis for assuming that variation in cluster profiles is a result of differing HVAC loads. It follows that the profile with the highest demand represents a high level of chiller usage, whilst the profile of lowest demand represents a baseline level of chiller usage.

For a different case, such as an office, where weekday and weekend profiles are likely to be represented in different clusters, a larger optimum set of clusters is likely. Identifying baseline level chiller usage would potentially be more difficult in such circumstances where greater variability in demand related activity is found. Determining what the profiles represent highlights the primary drawback of this method as it requires assumptions to be made on limited data. Incorrectly assuming what the profiles represent will result in incorrect DSR estimations and therefore this method needs to be used with caution.

Based on the assumption that profiles represent differences in chiller usage levels, the first step is to identify days associated with baseline use. In the context of the UK, chillers are not typically in use during the winter months. The baseline is, therefore, considered as days when the chiller is switched off during the heating season. The remaining clusters represent days when the chiller is in use. For this case, the kW usage levels of the chiller on these days is estimated by the difference between the cluster's usage value and the baseline value. Even in the case where the baseline cluster does not represent chiller switch-off the differences in usage could still be considered as representative of maximum available turndown. See Appendix B for detailed calculation steps used in this method.

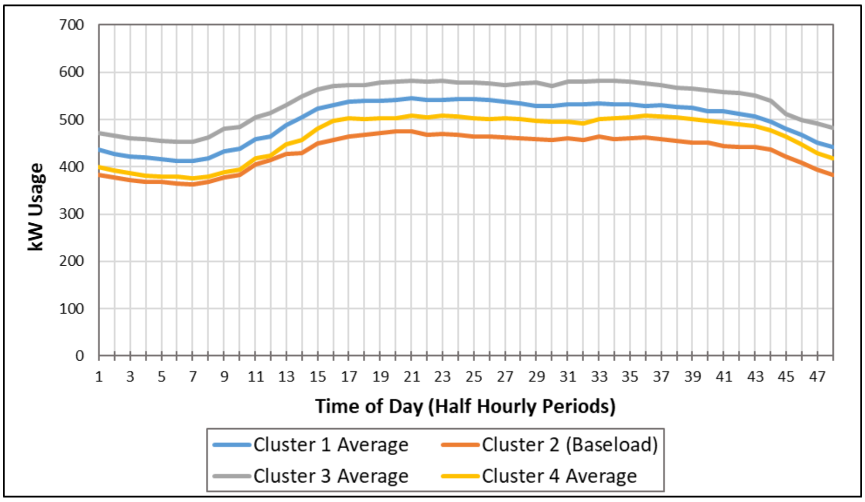

Figure 2 - Example Chart of Clustered Averages

2.4 DSR Estimation Method 3 - Utilise Historical DSR Event Outcomes

If a building has previously participated in DSR, then information gained on the kW amount reduction during each event can be utilised to estimate future performance. Research on this method has traditionally focused on confirming the DSR performance of a building by calculating the 'residual demand' (referred to as 'turndown' in this research), which is deemed as the difference between normal non-DSR building usage and the actual usage during a DSR event (Mathieu et al., 2010). Further research into understanding the expected level of residual demand using weather-based regression analysis was undertaken by Piette et al. (2011). They showed that the level of turndown potential could be linked to temperature if the DSR assets demonstrate varying levels of usage based on external weather conditions. This DSR estimation method utilises these concepts to identify a predictor that determines the expected turndown amount of historical DSR events. The predictor can then be utilised to determine the expected turndown amount at any time over a one-year period.

This method relies on access to historical DSR event outcomes for the building. To provide consistency for testing this method with both hotels, a set of 24 DSR events were randomly created. The DSR events were then matched to each hotel's actual chiller sub-metered data to provide real kW events outcome for each year of analysis (on the basis that during the event the chillers would have been temporarily turned off). Secondary data sources include any values that can be used for regression analysis to find a suitable predictor of the DSR event outcomes. For this research, the predictors selected for analysis were Outside Air Temperature, Building's Electricity Usage Level, Half Hour Period of Day, and Day of Week. The first step in this method is to calculate the R-squared value of each predictor against the historical DSR event outcomes to decide which predictor to use. The regression calculation results of Table 1 show that the Outside Air Temperature predictor achieved the highest r-squared score and therefore this predictor is selected for the next step. The second step then uses the Outside Air Temperature values for each half-hourly period of the year in conjunction with the predictors slope and y-intercept to calculated the DSR estimation potential for the buildings. See Appendix C for detailed calculation steps used in this method.

Table 1 - Method 3's R-squared Regression Results

Hotel / Year	Time of Day	Day of Week	Buildings Electricity Usage Level	Outside Air Temperature
Hotel 1 - 2013	0.003	0.036	0.273	0.722
Hotel 1 - 2016	0.003	0.040	0.087	0.636
Hotel 2 - 2015	0.007	0.017	0.046	0.434
Hotel 2 - 2016	0.019	0.028	0.066	0.447

2.5 DSR Estimation Method 4 - Building Energy Modelling

Building energy modelling provides insight into DSR potential by modelling the energy usage of building assets under different operational and environmental scenarios. Modelling gives insight into flexibility of asset usage that can then be used for DSR estimation. However, this is very time consuming in comparison to the previous estimation methods, and requires a very high level of information and specialised skills to complete. Utilising a database of archetypal building models for a building stock can help reduce the modelling burden for DSR, as demonstrated by Yin & Black (2015). The predefined model archetypes can be modified as necessary, but its success is dependent on the maturity of the database of archetypes and level of modification needed to provide results deemed of value to DSR estimation. Another issue with energy building models is the 'performance gaps' between model designs and actual performance of completed buildings, which can result in high levels of output uncertainty (Menezes, Cripps, Bouchlaghem, & Buswell, 2012). For this research, the building energy model DSR estimation method utilises the Yin & Black (2015) methodology by creating a building energy model of the test hotels using EnergyPlus. The outcome of the simulation includes the expected level of cooling in kW per half hour that will be used for DSR estimation.

To undertake this energy modelling approach, the building plans for each hotel were used to provide both accurate building dimensions as well as the fabric structure of the building (outlined in Table 2). The building plans are used to create a representative model of the building using the software package 'DesignBuilder' v5.0.2 (DesignBuilder, 2017b). The DesignBuilder program then utilises the EnergyPlus simulation program (U.S. Department of Energy, 2017) to estimate the buildings energy usage over one year at half hourly intervals. The simulated energy usage results of the modelled chiller units were then exported from DesignBuilder to provide the DSR estimation potential for each building. See Appendix D for detailed calculation steps used in this method.

Table 2 - Build Energy Model Components

Component	Hotel 1 Description	Hotel 2 Description
External Walls	400mm thick wall (formed of stone masonry, brick, glass wool insulation, and plasterboard) total U-Value of 0.289	300mm thick wall (formed of brick, polystyrene insulation, concrete, and plasterboard) total U-Value of 0.351
External Windows	Double glazed (formed of two 3mm panes with a 6mm air gap) total U-Value of 3.365	Double glazed (formed of two 3mm panes with a 6mm air gap) total U-Value of 3.365
Roof	400mm flat roof (formed of asphalt, glass wool insulation, air gap, plasterboard) total U-Value of 0.322	320mm Flat roof (formed of asphalt, glass wool insulation, air gap, plasterboard) total U-Value of 0.346

HVAC System	Fan Coil Unit (4-Pipe), 333kW air-cooled chiller with a cooling set point of 23°C	Fan Coil Unit (4-Pipe), 290kW air-cooled chiller with a cooling set point of 23°C
Property Details	7 stories, ~21,000 m^2 isolated building located in Bristol, UK.	6 stories, ~15,000 m^2 isolated building located in London, UK.
Weather File	Custom DesignBuilder weather data file created for each year of analysis (DesignBuilder, 2017a).	

2.6 Sensitivity Analysis

The accuracy of estimation method is an important factor in creating credible/robust DSR portfolios that can meet grid-operator needs. Appropriate interpretation of uncertainty in inputs to the proposed methods is, therefore, critical to DSR estimation. To understand the impact of each estimation method's input uncertainty on the DSR estimation, and so give insight as to where more accurate information should be sought, a one-at-a-time local sensitivity analysis test was carried out, as in Saltelli, Chan, & Scott (2008). The sensitivity results are compared using the HVAC chillers yearly MWh usage estimation output as generated by of the four methods, as this provides scale context to the test outcomes. In performing the sensitivity tests, each method was first run using base values for each input parameter, as described in Figure 1 and sections 2.2 to 2.5. Completing this step provides baseline outcomes for comparison against. Each input parameter was then adjusted from the base values, as outlined in Table 3, and the sensitivity test for each method re-run using the adjusted input parameter, generating the sensitivity comparison results. As estimation methods 1-3 only have one or two input variables, all inputs for each method are tested during the analysis. The detailed modelling approach of Method 4, however, has a wide range of input variables ranging from building form and structure, to operational schedules of appliances and occupancy profiles. In this instance, it is assumed that the availability of building plans and detailed information of HVAC and lighting infrastructure reduces uncertainty in many of the structural aspects of the model. Menberg, Heo, & Choudhary (2016) identified temperature set points, thermal conductivity, and air infiltration as having a significant impact on building energy model results. These three variables are the focus of our analysis for Method 4.

Table 3 - Summary of Estimation Method Sensitivity Analysis Input Parameters

Method	Base Values	Input Adjustment
1 (1)	50%	Adjust asset usage percentage by +/- 5 and 10 points
1 (2)	10%	Adjust asset percentage usage of baseload value by +/- 2.5 and 5 points
	5%	Adjust baseload percentile by +/- 1 and 2 points
2	4	Adjust number of clusters used by +/- 1 cluster
3	12	Adjust number of available existing events by -50%, +50%, +100%
4	23 °C	Adjust cooling set point by +/- 1 and 2 °C
	0.289 to 3.365	Adjust U-Values of External Walls, Windows, and Roof by +/- 10% and 20%
	0.7	Adjust air infiltration levels by +/- 0.1 and 0.2 ac/h

2.7 Determining the Cost of each Estimation Method

The final output of the review of DSR estimation methods is a comparison of each method's estimation errors in relation to its cost to run. This comparison is performed to provide context on the usage of each method in a business setting. It enables consideration of the cost/benefit selection of a higher error method that is cheaper or vice-versa. To calculate each method's cost to run in a business setting required estimating the time it would take an experienced user to perform the tasks needed to run the estimation method and the cost of any external data input requirements. Table 4 provides a summary of the expected time required and external cost (if any) for each informational input. The time estimations used in this table are necessarily subjective, as the actual time and cost required will depend on and vary by individuals and organisations. Given the potential for variability, creating a cost factor provides a means of understanding the representative scale of effort required to undertake each method. The figures used in this table provide a point of reference, comprising estimations based on experience gained through application of these methods within a UK aggregator for this research and observations of users. The time value includes both the time taken to obtain information about the building (this covers talking to the building representative to obtain the sites half hourly electricity usage data and information about the DSR assets) and the time required to format, analyse and interpret the data. Most external information has no direct cost, as it is obtained for free from the building users or other sources. The only externally sourced information incurring cost is historical weather observations (ECMWF, 2017), which has a fixed yearly fee of £5,000 and has been split into individual usage costs on the assumption of performing 500 assessments per year (£10 per usage).

Table 4 - Summary of Estimation Methods Information Input Costs

Information Input	Time to obtain / use (minutes)	Usage Cost (@ £20 per hour)
Maximum kW rating of building's DSR assets	30	£10
Building's electricity usage records for 1 year	60	£20
Previous DSR Event Outcomes	120	£40
Hourly outdoor weather information for 1 year	60	£20 + £10 (data)
Building plans and operational information	420	£140

To calculate the total cost of performing each method, the individual costs of gaining data for each input from Table 4 are associated with each method as per Table 5. This table shows the cumulative total running cost of each method, based on the information required. This information combined with the MAPE results from section **Error! Reference source not found.** enables a comparison of estimation error against method cost to be performed, as shown in section 3.3.

Table 5 – Summary of Costs to Perform Each Estimation Method

Information Input & Cost		**Information Usage and Cost per Method**				
		1 (1)	**1 (2)**	**2**	**3**	**4**
Maximum kW rating of building's DSR assets	£10	£10	£10	£10	£10	£10
Building's electricity usage records for 1 year	£20		£20	£20		
Previous DSR Event Outcomes	£40				£40	
Hourly outdoor weather information for 1 year	£30				£30	£30
Building plans and operational information	£140					£140
Total Cost per Method		**£10**	**£30**	**£30**	**£80**	**£180**

3 Results and Discussion

The results of applying the four DSR estimation methods to two hotels is reviewed and discussed over three sections. The first section reviews the initial outputs of each method by applying 'base case' values to the input variables, and comparing the estimation error between methods. The second section then reviews the sensitivity analysis results to understand the impact of input variables on the estimation error levels. Finally, the error levels are compared against the estimated cost of undertaking each method, to gain an understanding of how cost and error levels correspond.

3.1 Estimation Method MAPE and MBE Outcomes

The estimation errors of MAPE and MBE for each estimation method, when using default (base) values for input variables, are given in Table 6. The methods were applied to each hotel over two years to generate a predicted half hourly kW usage value for their HVAC chillers. The predicted kW values were then compared to the actual kW usage values (as recorded by sub-meters), and MAPE and MBE were calculated for annual estimation errors. The average, minimum, and maximum MAPE and MBE values were then calculated, as shown in Figure 3. The MAPE values provide an overall indication of the level of difference between the actual and predicted results. Figure 3 and Table 6 show a range of MAPE estimation errors across the methods, with M1-V1 'Asset Assessment' having the worst average level of error at 159%. In contrast, M3 'Utilise Historical DSR Event Outcomes' had the lowest average level of error at 39%.

The MBE values indicate the direction of error between the actual and prediction values, with positive and negative results indicating over estimation and under estimation respectively. Figure 3 shows that all methods, except M1-V1, under predict usage levels. As seen with the MAPE result, the M1-V1 outcome also has the highest average MBE value at 150%, which indicates that this method dramatically over predicted the expected usage of the HVAC chiller. In contrast, with an average MBE of -10%, M4 provides the closest prediction to actual usage.

Table 6 – Individual hotel summary of Estimation Method error levels

Method	Hotel 1 - 2013		Hotel 1 - 2016		Hotel 2 - 2015		Hotel 2 - 2016	
	MAPE	MBE	MAPE	MBE	MAPE	MBE	MAPE	MBE
M1-V1	193%	122%	250%	136%	98%	236%	96%	104%
M1-V2	35%	-46%	59%	-50%	71%	1%	75%	-38%
M2	57%	-41%	59%	-29%	40%	16%	70%	-12%
M3	33%	-15%	40%	-7%	36%	-6%	46%	-18%
M4	58%	-1%	63%	5%	39%	2%	45%	-31%

Abbreviation Key:

M1-V1 = Method 1- Variation 1 - Minimum information using set percentage of asset usage

M1-V2 = Method 1- Variation 2 - Utilise baseload calculation with set usage percentage

M2 = Method 2 - Baseline comparison using cluster analysis

M3 = Method 3 - Regression analysis utilising historical DSR event outcomes

M4 = Method 4 - Building energy modelling

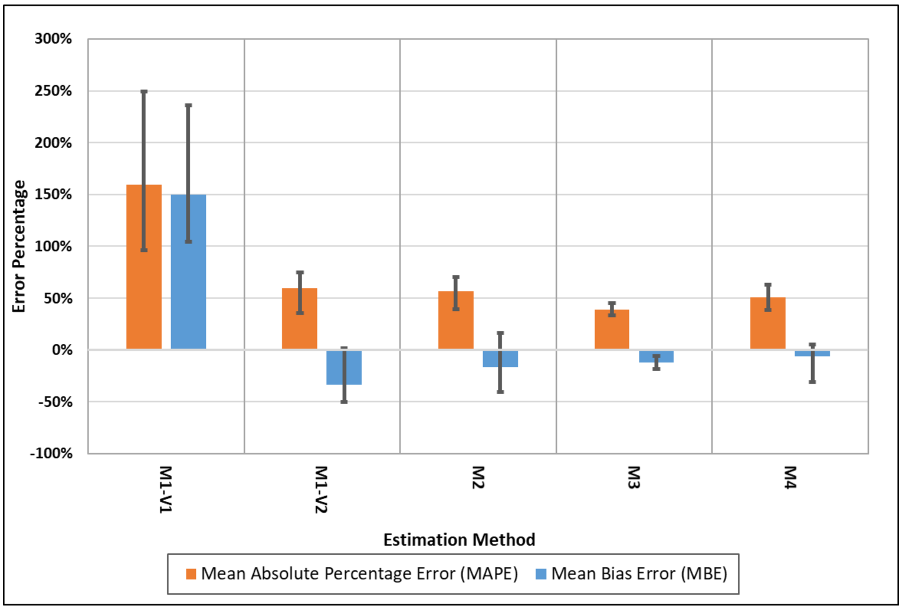

Figure 3 – Summary of Each Estimation Methods Error Levels

Considering the outcomes of each method: the two sub-variations of M1 had contrasting results with M1-V1 having the highest overall average error level at 159%, while M1-V2 had a considerably lower error level of 60%. The high uncertainty level of the M1-V1 method could be a result of it assuming a fixed usage level of a chiller when the actual sub-meter data shows a highly variable pattern based on a usage percentage mean of 20.8% with a variance of 252.5%. In contrast, M1-V2 uses the more variable input of the building's overall electricity usage levels for a year to first calculate the buildings baseload usage. A percentage (in this case 10%) of the baseload is then deemed to be used by the DSR asset, producing a much lower average MAPE value of 60%. This result is unexpectedly low considering the method still uses a fixed proportion of buildings usage, which only considers time of day variation and results in the same half hour prediction values being used for the entire year. The error level is still high due to this method only taking time of day variation into account and does not consider day of year variation which will impact the estimation results of a chiller that is highly influenced by seasonality

An average MAPE value of 56% placed M2 as the method with the second highest level of absolute error. Comparatively, however, the average MAPE is similar to the M1-V2 and M4 results. This outcome, which is based on the method outlined by Panapakidis et al. (2014), helps support usage of their profile clustering technique based on the DSR estimation results being comparable to the other methods. Caution however needs to be taken on assuming this method is comparable to M1-V2 and M4 due to its assumptions around the differences between profiles indicating usage of a particular electrical asset, which may be difficult to determine in different businesses.

The lowest MAPE of all the methods was M3 at 39%. The ranking of method suitability by MAPE supports research by Piette et al. (2011) where the inclusion of temperature dependency of DSR assets in predictors improves prediction. For non-weather impacted assets other potential regression parameters could be used including time of day, occupancy levels, or operational schedules. The drawback to this method is access to historical DSR events and obtaining suitable predictor data, which could be hard to come by.

An average MAPE value of 51% placed M4 as the method with the second lowest level of absolute error. It is possible to achieve lower levels of error as demonstrated by the researchers at the Lawrence Berkeley National Laboratory (Dudley, 2010) who used calibrated Energy Building Models for accurate DSR forecasting. However, the calibration methods require obtaining sub-metered data of key electrical assets which, if available, could be used directly for predicting the building's DSR usage, limiting the need for using an Energy Building Model. While this method achieves comparatively good error estimation levels even without calibration, it does have the drawbacks of requiring access to detailed plans of a building and the skill and time needed to construct the model.

3.2 Sensitivity Analysis of Estimation Methods

The previous review of the error in the estimation methods provides a comparative analysis of methods without accounting for the uncertainty in their input values. The error range in DSR estimation depends not only on the estimation methodology, however, but also on these input uncertainties and the sensitivity of method outcome to these uncertainties. Figure 4 summarises the sensitivity profiles for each method's inputs, as determined by re-running each method with adjusted inputs. To facilitate comparison of sensitivity between methods, the charts shown in Figure 4 have been normalised. Plotting change in input variable as a percentage of the base case value against the percentage difference in estimated energy use (MWh), Figure 4 shows varying sensitivity to inputs within and across the four methods. This section examines each method's sensitivity profiles to gain further insights into how they are influenced by input variation.

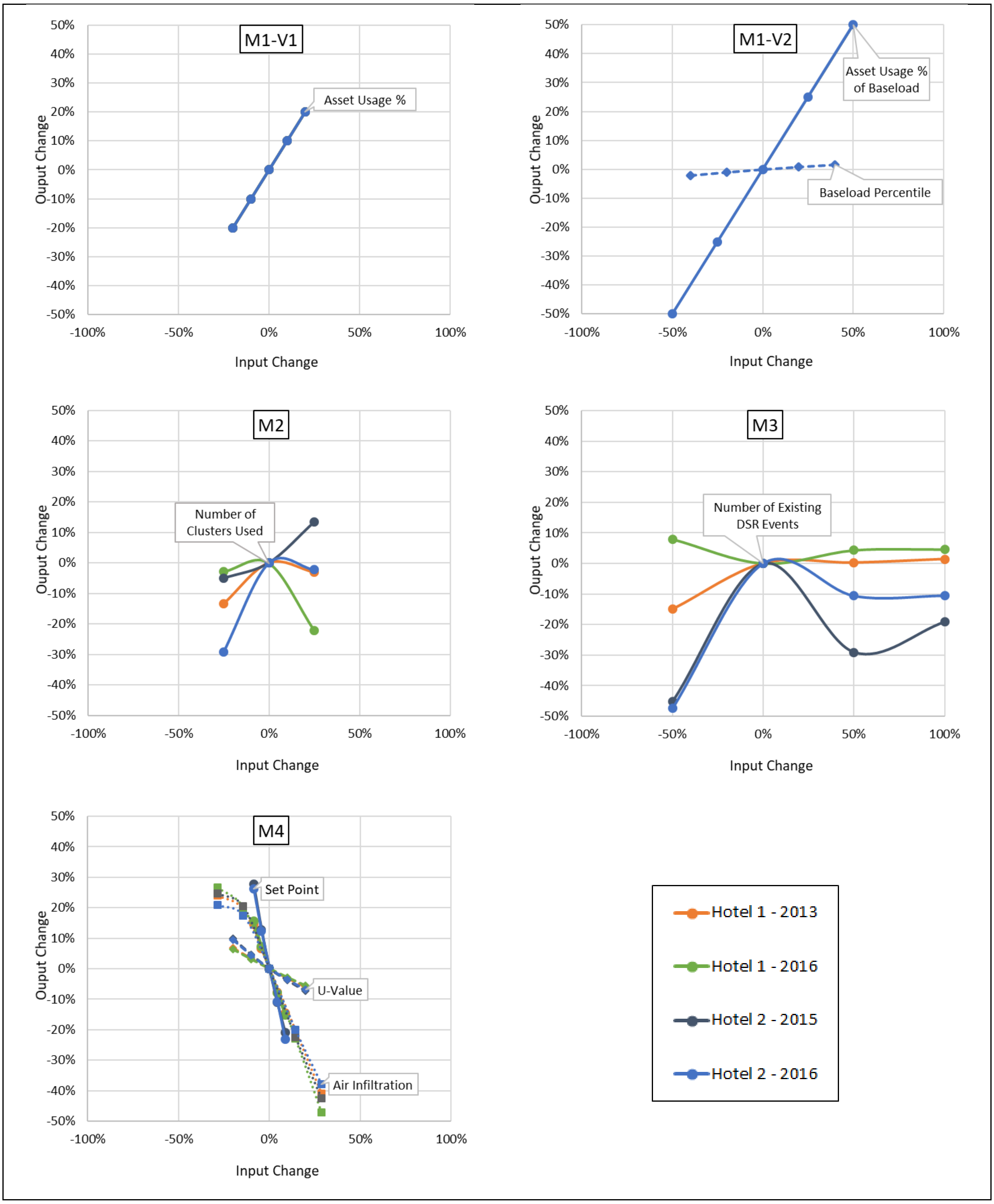

Figure 4 - Estimation Method Sensitivity Analysis Results

The asset usage percentage input gradients of M1-V1 (1:1) and M1-V2 (1:1) shows they are both sensitive to changes, whilst adjustments in the percentile value used for baseload estimation in M1-V2 has little effect (0.04:1). Altering the asset usage percentage input values for M1-V1 and M1-V2 however had different impacts on the resulting MAPE outcomes across both hotels and years. The M1-V1 MAPE outcomes varied from -28.2% to 29.5% with a consistent pattern of the MAPE value decreasing as the percentage of asset usage value lowered. This indicates that the base usage value of 50% is too high and a lower value should be used to better represent actual usage of the chillers. The M1-V2 MAPE outcomes had a greater variance level of -11.5% to 82.8% and in contrast to M1-V1, when the asset usage percentage of the baseload value is lowered the MAPE values increased.

However, when the usage percentage is increased MAPE values for Hotel 1 initially lower before increasing - indicating that the base value is close to optimal. MAPE values for Hotel 2 continue to decrease as the usage percentage increases indicating that a higher base value would be more appropriate. The other input for M1-V2, percentile baseload value, has negligible effect on the MAPE outcomes with a variance range of -1.0% to 1.6% across both hotels and years and therefore the base value of 10% is deemed appropriate.

M2 has a non-linear sensitivity profile, with each hotel and data year being impacted differently with no clear pattern. The percentage change in MAPE values resulting from the input changes has a variance range of -6% to 7% across both hotels and years. This level of MAPE variance implies that changing the number of clusters has only a small impact, and that the base value is appropriate for this application of the estimation method. The limited output variance could be the result of this method calculating the chiller usage values based on differences between cluster profiles that means adding or removing a single cluster will only cause the redistribution of input values into other similar clusters without causing major changes in the generated profiles.

M3 also has a non-linear sensitivity profile that varies differently between the two hotels. The general pattern of the profile shows that when the number of historical events is lowered by 50% from 12 to 6, this has the greatest impact on estimation outputs, with MAPE values increasing by 4% and 23% for Hotel 1, and 28% and 75% for Hotel 2. When the number of events is increased to 18 and 24, the profile shows a more consistent change, except for Hotel 2 -2015. When excluding Hotel 2 -2015, the MAPE values had a minimal change range of -2% to 5%. However, Hotel 2 - 2015 showed far greater changes, with the MAPE value increasing by 45% and 28%. A potential cause of this difference could be due to the facilities manager of Hotel-2 deciding when to turn the chiller system on and off during the year. In 2015 it was turned on in April and off in October, whereas in 2016 it was turned on in May but not turned off again. In contrast the Hotel-1 system is left running all year with output adjusted automatically as required to meet the set point conditions. Based on the overall results of this method it isit’s clear that reducing the number of historical events has a negative impact on the outcomes. Whereas the impact on increasing the number of events used is unclear due to the outcomes of Hotel 2 – 2015.

M4 has three different input variables of Cooling Setpoint, U-Value, and Air Infiltration. The Set Point Temperature and U-Value inputs have linear sensitivity profiles with gradients of (1:0.32) and (1:0.6) respectively. The Air Infiltration input range of 0.6 to 0.9 ac/h had a linear profile of (1:0.7), however the lowest input value of 0.5 ac/h was not linear with a smaller change in output compared to the linear values. Air Infiltration changes displayed the biggest impact on output and resulting MAPE values. This is shown with the MAPE values for Air Infiltration having a variance range of -18% to 54%. In contrast, the MAPE values range for the U-Value input was -8% to 8% and the Cooling Setpoint input range was -18% to 27%. The results show how changing the Set Point temperature and Air Infiltration rates have significant impacts on the chiller usage compared to only a minor impact from changing U-Values. This could reflect the usage of mechanical space cooling, which actively responses to temperature requirements and causes pressurised losses through Air Infiltration. The Air Infiltration input having the biggest impact does raise concern for this type of estimation method, as this is one of the hardest parameters to determine when constructing the energy building model. The other inputs can be obtained with relatively high accuracy by obtaining the Set Point directly from the building's current setup and the U-Values from visual inspections of the existing construction and building plans. In contrast, the Air Infiltration rate can only be accurately obtained through a building pressure test which would be infeasible for a building of this size. Therefore, the default building model Air Infiltration rates will need to be used, and caution taken on the final outputs.

3.3 Cost versus Method Estimation Errors

The final set of results compares the cost of running each method against the expected level of estimation error. This comparison helps provide context to usage of the methods when balancing cost against acceptable error levels. Figure 5 maps out the links between each method's average MAPE results as per Table 6 and the estimated cost to run as per Table 5. The figure shows a rough trending direction of a higher method cost resulting in lower estimation errors. This is reflected in the lowest cost method M1-V1 having the highest error level while the lowest error level M3 has the second highest cost. Each method will be further examined to understand the implications of method costs and input requirements on error outcomes.

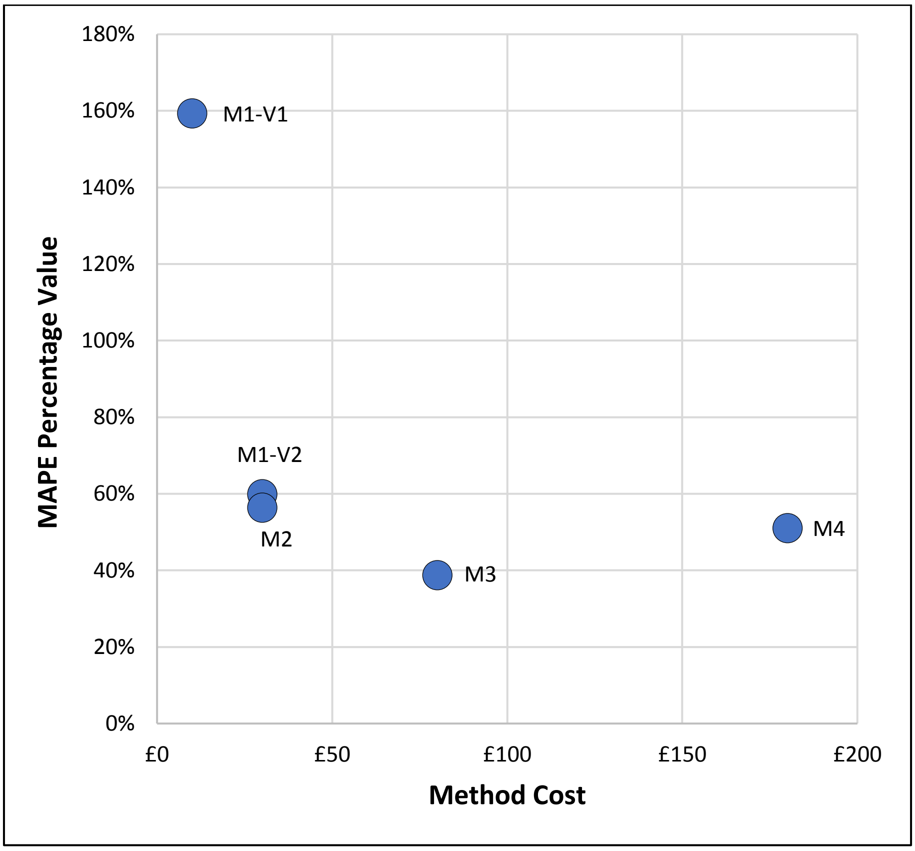

Figure 5 - Comparison of Estimation Method Error versus Cost

M1-V1 has the distinction of being the cheapest estimation method with the worst error level. This can be directly related to the input requirement of only needing to know the asset's maximum kW rating, and then using a percentage of this for the estimation. This requires minimal time for a person to undertake, both in collecting the required information and using it to calculate the estimation. Unfortunately, the high error level means that this method can only be used for a very rough and quick estimation before proceeding with a lower error method. In comparison, M1-V2 reduces the error level by two thirds compared to M1-V1 while costing 3 times more to run. While M1-V2 is more expensive than M1-V1, it is still comparatively cheap compared to all the methods tested. This method also uses relatively accessible data of the building's electricity usage records, which in the UK is available in half hourly format for any business with peak electricity usage of 100 kW or greater.

M2 is the third equally cheapest method to run due to the primary input requirement being the building's half hourly electricity usage records. It also has the fourth lowest error level and therefore, of the methods analysed, provides a representatively balanced error to cost ratio, which makes it a potentially suitable approach. However, as discussed previously, this method's usage of clustering means that care needs to be taken on its application to suitable buildings and assets.

M3 achieved the lowest error level of all methods tested at 39%. However, it also has the second highest cost at £80, which is a result of requiring two expensive input requirements. Firstly, it uses detailed historical air temperature readings over a year for the building's location, which requires paying for access to the necessary weather archive. Secondly, it uses previous DSR event outcomes which require time to obtain from the building users, and then formatting and verifying before using. It is also anticipated that obtaining previous DSR event outcomes could be difficult, due to the limited current uptake of DSR and even if the client has participated, then it could be difficult for them to provide the necessary information based on how it has been provided from their current aggregator.

M4 had the highest cost at £180 with the second lowest error level of 51%. The high cost is primarily due to the time required to model the building in the building energy modelling tool. As the resulting error level is similar to M1-V1, M2 and M3 methods, which are significantly cheaper to run, this method is not recommended. Although a potential justification for using this method would be if multiple assets within one building were being estimated, thereby reducing the individual assessment costs while providing a combined view of the building's potential.

4 Conclusion

This paper has undertaken an examination and comparison of four non-domestic DSR estimation methods to provide insights into uncertainty levels based on the input requirements. The examination was performed by using each method to estimate the DSR potential of HVAC chiller assets at two hotels over two years. The estimation outcomes were then compared against the chiller's actual sub-metered usage records by calculating MAPE and MBE values to understand each method's level of estimation error. The results showed a wide range of estimation errors. Method 1 - Sub-variation 1 yields the highest error level MAPE of 159%, while the lowest error level MAPE of 39% was achieved with method 3. While method 3 could be a recommended approach based on its low error level alone, its usage is restricted by information input considerations. The primary limitations of this research were a reliance on usage of one electrical appliance (HVAC chillers) and business (hotels) type, uncertainty of the method usage time and cost input variables due to the subjectively of how each organisation could apply them, and being restricted to using only known estimation methods that excludes unpublished proprietary approaches. Based on this paper's findings, each method requires review to understand the implications of input requirements on outcome uncertainty. These findings can be summarised as follows:

- **Method 1** sub-variation 1 has the lowest informational requirement and cost of £10 to use based on only needing to know the maximum kW rating of the asset being assessed to apply this method. However, the penalty of this low informational requirement is the highest error level of all methods at 159%. Sub-variation 2 achieved a much lower error level of 60% by using the building's half-hourly electricity records that increases the usage costs to £30. The sensitivity results for this method showed a high impact on the outcomes based on variations of the inputs. This means that the error results might differ substantially when used in other scenarios. Therefore, the error levels reported in this research for method 1 need to be used with care when deciding on suitable assessment approaches.

- **Method 2** had the second worst error level of 56% while being the third cheapest to run at £30 through clustering of the building's half hourly electricity usage data. The sensitivity analysis of this method showed a medium to low impact on error levels arising from changes in the primary input of how many clusters are used. These results indicate that baseline comparison is a suitable method for assessment though it has two limitations that need to be fully understood by users to ensure valid results. Firstly, it requires the user to select the appropriate number of clusters, which is open to individual interpretation. Secondly, this method will only work on electrical assets that have enough variation within the building's overall usage to be identified by the clustering.

- **Method 3** had the lowest overall error level of 39% with the second highest cost of £80. The low error level makes its utilisation of historical DSR event outcomes an attractive method. However, its practical usage is limited as it requires the building to have previously undertaken DSR and have access to historical DSR events outcomes. The sensitivity analysis also showed a significant increase in error if less than 12 historical event records over a year are available for analysis. In new DSR markets these limitations may restrict usage of this method. Even in established markets it could be difficult or time consuming to obtain any adequate historical information from the existing DSR aggregator.

- **Method 4** had the second lowest error level at 51% but had the highest cost of £180, which is over twice that of method 3, the next most expensive, as a consequence of the amount of time required to develop a building energy model. While this method had the second lowest error level, it is only slightly lower than many other cheaper options and method 2, for example, costs 6 times less with only a slightly higher error level of 56%. The usage requirements of this method also restrict its practical application given its reliance on detailed building plans and the skills to develop building models. The importance of having the right information and skills is highlighted by the sensitivity analysis, which showed major impacts from variations in temperature set-points and air infiltration model values.

These findings have three key implications on the selection of DSR estimation methods. Firstly, the wide range of error levels means the outputs of these methods will need to be carefully considered when being used to make decisions about the suitability of buildings for DSR. Secondly, care needs to be taken in ensuring accurate input selection as sensitivity analysis demonstrates that adjusting the inputs on most methods will result in large variations to the outputs. Thirdly, this research tested four methods using HVAC chillers in hotels only. Therefore, other assets and businesses may result in different error outcomes and caution needs to be taken before this research is used to select estimation methods outside of this scope. This final implication highlights a potential future area for research which would entail re-running the method comparisons on different DSR assets and businesses to understand the different impacts on estimation outcomes.

5 Funding

This work has been supported and funded by the Technologies for Sustainable Built Environments (TSBE) centre, Reading University, in conjunction with the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/G037787/1].

6 Appendix A

The following steps outline the calculations performed for Method 1 - Asset Assessment:

1. **Variation 1 – Minimum Information**
   1.1. An anticipated set percentage usage amount of the asset is selected based on either a default 50%, or another amount if the assessor has prior knowledge of the type of asset and site.
   1.2. The expected kW usage level of the asset is calculated for each half-hour of a year by multiplying the anticipated percentage usage amount by the maximum rating of the asset, with the resulting values being saved into a DSR asset usage estimation dataset.

2. **Variation 2 – Utilise Baseload Calculation**
   2.1. Using the site's Metered Electricity Usage Records, a baseload value is calculated by obtaining the 5th percentile kW value for each half-hour period of the day based on one year's worth of data as per formula (1) (e.g. for each half-hour period of a day, the 365 daily values for the year are obtained and then ranked before determining the 5th Percentile value).

$$n_{HH} = \left\lceil \frac{P}{100} \times N_{HH} \right\rceil \qquad (1)$$

Where:
$n =$ kW value of percentile for selected half-hour
$P =$ Percentile
$N =$ Ordered list of kW values for selected half-hour (sorted from least to greatest)
$HH =$ Selected half-hour

   2.2. A percentage value is then selected that represents how much of the baseload is expected to be used by the asset. This can either be a default 10%, or another amount if the assessor has prior knowledge of the asset type and site.
   2.3. The expected kW usage level of the asset is calculated for every half-hour period in a year by multiplying the anticipated percentage usage amount against the baseload kW value, with the resulting values being saved into a DSR asset usage estimation dataset.

2.4. If the usage outcome is higher than the maximum usage rating of the assets, then the previous step is re-run with a lower percentage.

7 Appendix B

The following steps outline the calculations performed for Method 2 - Baseline Comparison:

1. The k-means cluster method is used for the baseline comparison (Sayad, 2017). This clustering method works by first selecting how many groups the usage dataset will be clustered into. For each group, a random point within the dataset is selected and deemed the centroid value. Each value in the dataset is assigned to the closest centroid. The mean of the values for each centroid is then calculated. The centroids are then moved to the mean position and the values are reassigned to the now closest centroids. This process is repeated until a pre-defined number of interactions is achieved or the level of centroid position change reaches a set tolerance.
2. The number of clusters for the baseline comparison will vary for each site. For this analysis the 'elbow' method for determining the optimum number of k-means clusters is used. This method works by repeating the k-means method using a range of clusters to determine each cluster's percentage of variance. The percentage of variance (dependent variable) is plotted against the number of clusters (independent variable) in order to find the 'elbow' of the curve, which signifies the optimum number of clusters as adding more will have limited benefit in reducing variance (Ketchen & Shook, 1996). Figure provides an example of identified `elbow' for clustering of one hotel's daily electricity usage profiles over one year. The main recognised limitations of the elbow method is its reliance on a manual decision-making process to determine where the elbow sits, and that the chart might not have a recognisable elbow if the line is consistent across the clusters (Ketchen & Shook, 1996). The elbow method calculation is performed by:

    i. Calculating the percentage of variance explained for a range of clusters (normally 1-15) using the equation (Imran, 2015).
    ii. Create a line chart with markers that shows each cluster's percentage of variance as shown in Figure for Hotel 1 in 2016
    iii. Determine the elbow based on the chart and record the cluster number.

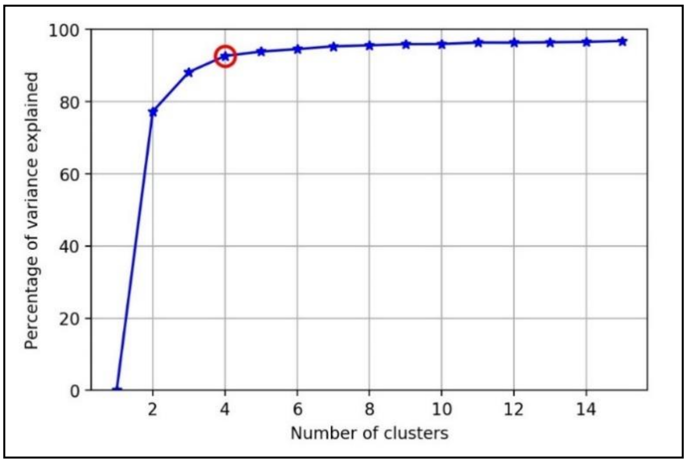

Figure 6 - Example of Cluster Identification using the Elbow Method (with the Elbow being indicated by the red circle)

3. Once the number of clusters to be used has been decided, then the k-means method as shown in equation (2) (Sayad, 2017) can be used to group the Site's Half Hourly Electricity dataset into similar days. The dataset is then updated with a new column 49 containing a value that represents which cluster each day belongs to.

$$J_n = \sum_{j=1}^{K} \sum_{i=1}^{n} \left(x_i - c_j\right)^2 \tag{2}$$

   Where:
   $n =$ Objects being clustered
   $J_n =$ Cluster outcome for n value
   $K =$ Clusters
   $c_j =$ Centroid for cluster *j*
   $x_i =$ Object *i*

4. The half-hourly averages in each cluster are then used to generate daily profiles at half-hourly resolution for each cluster of each hotel. **Error! Reference source not found.** provides an example of the daily profiles developed for the four identified clusters of a hotel.
5. The baseline profile is then identified based on the assumption that the profiles represent differences in chiller usage levels. In the context of the UK, chillers are not typically in use during the winter months. Therefore, the baseline is considered as days when the chiller is switched off during the heating season and, as a result, profile cluster 2 in **Error! Reference source not found.** comprises the baseline profile as it has the lowest usage values. The remaining cluster profiles then represent days when the chiller is in use.
6. A new dataset is created that covers all half-hourly periods for one year, and has an additional column identifying which cluster profile is associated with each day of the year. For each day in the dataset, the kW usage levels of the chiller is estimated by the calculating the difference between that day's cluster profile usage value and the baseline value. If a day in the new dataset is associated with the baseline cluster, then the chiller is deemed to be off during this day, so the expected usage is set to 0.
7. The dataset now represents the DSR asset usage estimation dataset of the chiller. The results are then checked to verify that no values are greater than the maximum usage rating of the chiller asset. If there are, then the values are adjusted down to the maximum rating or, if the values are consistently too high, then this method is rejected if the assessor believes the method is providing unrealistic results based on the assessor's (or their colleagues') prior knowledge of customary usage for this type of asset.

8 Appendix C

The following steps outline the calculations performed for Method 3 - Utilise Historical DSR Event Outcomes:

1. The first step is to determine what variables are available for predicting the event turndown amount. For this example, the variables of Outside Air Temperature, Site Electricity Usage, Half Hour Period of Day, and Day of Week are used.
2. For each variable, a two-column dataset is created for each year of data with the first column containing the event turndown results, and the second column containing the predicting variable value.
3. Using equation (3) the R-squared / coefficient of determination for each dataset is calculated.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y})^2} \quad (3)$$

Where:

R^2 = R-squared / coefficient of determination

y_i = Current value from event data set

$\overline{y}$ = Mean of event data set values

f_i = Predicted value for y_i

4. The R-squared values of each variable used as shown in Table 1 are compared, and the highest value selected as the predictor variable to be used for estimating DSR asset usage. In this case the Outside Air Temperature has the highest values.
5. The Outside Air Temperature values for each half-hourly period of the year in conjunction with the predictor's slope and y-intercept are used to calculate the DSR estimation potential for the hotels.

9 Appendix D

The following steps outline the calculations performed for Method 4 - Building Energy Modelling:

1. The building plans for each hotel were used to provide both accurate building dimensions as well as the fabric structure of the building (outlined in Table 2). The building plans are used to create a representative model of the building using the software package 'DesignBuilder' v5.0.2 (DesignBuilder, 2017b). The DesignBuilder program then utilises the EnergyPlus simulation program (U.S. Department of Energy, 2017) to estimate the building's energy usage over one year at half-hourly intervals.
2. Customised weather files were generated for each hotel for the years 2013 and 2016 and loaded into DesignBuilder. These were created using MIDAS weather data (UK Met Office, 2017) that was then converted into an EnergyPlus formatted hourly weather data.epw file using the process outlined on the DesignBuilder online help (DesignBuilder, 2017a)
3. Each model's energy usage was then simulated at half-hour intervals for one year using DesignBuilder/EnergyPlus, with the results of the chiller assets electricity usage being extracted to provide the DSR estimation potential for each hotel.

10 References

Aman, S., Frincu, M., Chelmis, C., Noor, M., Simmhan, Y., & Prasanna, V. K. (2016). Prediction models for dynamic demand response: Requirements, challenges, and insights. *2015 IEEE International Conference on Smart Grid Communications, SmartGridComm 2015*, 338–343. https://doi.org/10.1109/SmartGridComm.2015.7436323

Borgstein, E. H., Lamberts, R., & Hensen, J. L. M. (2016). Evaluating energy performance in non-domestic buildings: A review. *Energy and Buildings*, *128*, 734–755. https://doi.org/10.1016/j.enbuild.2016.07.018

Chassin, D. P., & Rondeau, D. (2016). Aggregate modeling of fast-acting demand response and control under real-time pricing. *Applied Energy*, *181*, 288–298. https://doi.org/10.1016/j.apenergy.2016.08.071

CIBSE. (2012). *CIBSE Guide F*. London: CIBSE Publications.

Curtis, M. (2017). Demand side response aggregators: How they decide customer suitability. In *2017 14th International Conference on the European Energy Market (EEM)* (pp. 1–6). IEEE. https://doi.org/10.1109/EEM.2017.7981909

Curtis, M., Torriti, J., & Smith, S. T. (2018). Demand Side Flexibility and Responsiveness: Moving Demand in Time Through Technology. In A. Hui, R. Day, & G. Walker (Eds.), *Demanding Energy: Space, Time and Change* (pp. 283–312). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-61991-0_13

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, *22*(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

DesignBuilder. (2017a). DesignBuilder - Edit / Translate Hourly Weather Data. Retrieved February 3, 2017, from http://www.designbuilder.co.uk/helpv5.0/Content/_Edit_hourly_weather_data.htm

DesignBuilder. (2017b). DesignBuilder V5.0.2. Retrieved from https://designbuilder.co.uk/software/product-overview

DRRC. (2017). Demand Response Research Center. Retrieved August 5, 2016, from https://drrc.lbl.gov/about-us

Dudley, J. H. (2010). Comparison of Demand Response Performance with an EnergyPlus Model in a Low Energy Campus Building. Retrieved from http://www.escholarship.org/uc/item/9zw4m12g

ECMWF. (2017). European Centre for Medium-Range Weather Forecasts - Meteorological archive. Retrieved May 2, 2017, from https://www.ecmwf.int/en/forecasts/accessing-forecasts/order-historical-datasets

Imran. (2015). What is “Within cluster sum of squares by cluster” in K-means. Retrieved March 2, 2017, from https://discuss.analyticsvidhya.com/t/what-is-within-cluster-sum-of-squares-by-cluster-in-k-means/2706/2

Ketchen, D., & Shook, C. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

Larsen, E. M., Pinson, P., Leimgruber, F., & Judex, F. (2015). From demand response evaluation to forecasting - Methods and results from the EcoGrid EU experiment. *Submitted to IEEE Transactions on Power Systems*, *10*, 75–83. https://doi.org/10.1016/j.segan.2017.03.001

Mathieu, J. L., Gadgil, A. J., Callaway, D. S., Price, P. N., & Kiliccote, S. (2010). Characterizing the Response of Commercial and Industrial Facilities to Dynamic Pricing Signals From the Utility. *ASME 2010 4th International Conference on Energy Sustainability, Volume 1*, 1019–1028. https://doi.org/10.1115/ES2010-90266

Menberg, K., Heo, Y., & Choudhary, R. (2016). Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings*, *133*, 433–445. https://doi.org/10.1016/j.enbuild.2016.10.005

Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, *97*, 355–364. https://doi.org/10.1016/j.apenergy.2011.11.075

Merry, J. (2017). Inferior submetering is destroying performance. Retrieved April 8, 2017, from https://www.cibsejournal.com/opinion/inferior-submetering-means-substandard-performance/

National Grid. (2016). Short Term Operating Reserve Information. Retrieved January 30, 2017, from http://www2.nationalgrid.com/UK/Services/Balancing-services/Reserve-services/Short-Term-Operating-Reserve/Short-Term-Operating-Reserve-Information/

Panapakidis, I. P., Papadopoulos, T. A., Christoforidis, G. C., & Papagiannis, G. K. (2014). Pattern recognition algorithms for electricity load curve analysis of buildings. *Energy and Buildings*, *73*, 137–145. https://doi.org/10.1016/j.enbuild.2014.01.002

Piette, M. A., Mathieu, J. L., Price, P. N., & Kiliccote, S. (2011). Quantifying changes in building electricity use, with application to demand response. *IEEE Transactions on Smart Grid*, *2*(3), 507–518. https://doi.org/10.1109/TSG.2011.2145010

Saltelli, A., Chan, K., & Scott, E. (2008). *Sensitivity Analysis*. Chichester: Wiley.

Sayad, S. (2017). K-Means Clustering. Retrieved March 2, 2017, from http://www.saedsayad.com/clustering_kmeans.htm

SDG&E. (2016). *San Diego Gas & Electric - Capacity Bidding Program*. Retrieved from http://regarchive.sdge.com/tm2/pdf/ELEC_ELEC-SCHEDS_CBP.pdf

SEDC. (2016). *Explicit and Implicit Demand-Side Flexibility Complementary Approaches for an Efficient Energy System*.

SEDC. (2017). *Explicit Demand Response in Europe Mapping the Markets 2017*.

Shen, L., Li, Z., & Sun, Y. (2016). Performance evaluation of conventional demand response at building-group-level under different electricity pricings. *Energy and Buildings*, *128*, 143–154. https://doi.org/10.1016/j.enbuild.2016.06.082

U.S. Department of Energy. (2017). EnergyPlus. Retrieved April 3, 2017, from https://energyplus.net/

UK Met Office. (2017). MIDAS Data Guide. Retrieved February 15, 2017, from https://badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html

Van Wijk, J. J., Van Selow, E. R., Wijk, J. J. Van, & Selow, E. R. Van. (1999). Cluster and calendar based visualization of time series data. *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis’99)*, 4. https://doi.org/10.1109/INFVIS.1999.801851

Yin, R., & Black, D. (2015). *Improvement of Demand Response Quick Assessment Tool (Drqat) And Tool Validation Case Studies*.