Energy performance optimization of buildings using data mining techniques

Kai Corten^{1,2,*}, Eric Willems², Shalika Walker¹, and Wim Zeiler^{1,*}

¹Department of the Built Environment, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands ²Huygen Engineers & Consultants, P.O. Box 27, 6160 MB Geleen, The Netherlands

Abstract. The operational energy consumption of buildings often does not match with the predicted results from the design. One of the most dominant causes for these so-called energy performance gaps is the poor operational practice of the heating, ventilation and air conditioning (HVAC) systems. To improve underperforming HVAC systems, analysis of operational data collected by the building management system (BMS) can provide valuable information. In order to completely use and interpret operational data, the building sector is urging for methods and tools. Data mining (DM) is identified as an emerging powerful technique with great potential for discovering hidden knowledge in large data sets. In this study, the performance of HVAC systems was analysed using regression analysis as DM technique. This leads to valuable insights to control and improve the building energy performance. The results show that a reduction of 7-13% on the heating demand and 41-70% on the cooling demand can be obtained.

1 Introduction

In order to achieve the climate and energy targets of the European Union, existing energy-intensive buildings need to be renovated or replaced by high-performance nearly zero-energy buildings (nZEB). However, the operational energy consumption of these buildings often does not match with the design, which is termed as the 'energy performance gap' [1]. Evidence is emerging that buildings use on average 1.5 to 2.5 more energy than predicted in the design phase [2]. The most dominant causes for these gaps are uncertainty in modelling during the design phase, unknown occupant behaviour and poor operational practice of heating, ventilation and air conditioning (HVAC) systems [3]. In Europe, due to the low renovation and replacement rates of the existing building stock, the improvement of the operational energy performance of buildings is of significant importance [4].

Nowadays, modern non-residential buildings are generally equipped with an advanced building management system (BMS) which collects and stores operational data, such as temperature, flow rate, pressure, control signals and status of equipment in the HVAC system. The number of devices in current buildings able to capture data is as massive as the connections of human activities with the internet. This new paradigm is in many cases considered as the era of the internet of things (IoT), or the era of the big data (BD). This data can be used to control and improve the operational performance of buildings. However, the BMS can only perform simple data analysis and visualizations. In addition, the quality of the data analysis is heavily dependent on the knowledge and experience of the particular investigator. By extracting the data from a BMS and performing more advanced data analysis procedures, an unused potential of knowledge of building performance can be revealed.

Therefore, the building sector urgently needs advanced methods and tools to analyse the massive data collected by the BMS. Data mining (DM) technology is a promising approach for extracting useful insights from large data sets [5]. It is the process of turning data into information and gaining knowledge from this information. DM techniques are already successfully applied in various research fields. However, the application of DM frameworks for building energy consumption and operational data is still in an elementary phase [6].

Based on this information, this paper tries to analyse and identify energy reduction possibilities of HVAC systems of two case study buildings in the Netherlands, using regression analysis as the DM technique. In section 2, the methodology is presented. Section 3 and 4 scrutinize the case study buildings and the analysed results are presented. Finally, the discussion and the conclusion of the study are presented in section 5 and 6.

2 Methodology

2.1 Three main steps

Although there are studies performed on causes of energy gaps and ways to avoid these causes, a structured analysis which leads to the explanation of a specific energy gap is still missing [7]. In optimization and system recommissioning studies, high energy consumption

^{*} Corresponding authors: <u>k.corten@huygen.net</u>, <u>w.zeiler@tue.nl</u>

situations were detected followed by a quite labourintensive check of the building services systems [8]. In this research, the Pareto analysis [9] and LEAN Energy Analysis (L.E.A.) [10, 11] are used to perform such a study on the time series data of the building services installations. This approach is based on the previous work conducted by Huls [12] and Schoenmaker [13] for Pareto analysis and Vink [14] for L.E.A. [15]. The study consists of three main steps which are discussed below in detail.

2.1.1 Step 1: Most influential parameters

Since modern non-residential buildings have large data sets, it is important to find out which data is the most important to analyse. Therefore, the first step of the study is the identification of the most influential parameters of the HVAC system including the occupant behaviour by means of the Pareto analysis.

First, the two case study buildings are modelled with building energy simulation (BES) program the EnergyPlus [16]. Then, the influence of ten selected parameters on the heating and cooling demand is determined by simulating 10% higher and lower values than designed. This 10% deviation is determined in order to compare realistic possible values. After the simulations, the parameters are ranked, scored and grouped based on their maximum influence on the energy performance gap. According to the Pareto analysis, 20% of the selected parameters with the largest influence causes roughly 80% of the energy performance gap [9]. The application of this principle is visualized in Fig. 1, where the solid black line indicates the cumulative percentage. The challenge is to identify the 20% of the parameters which are mainly responsible for the underperformance of the HVAC system.



Fig. 1. Visualisation of the Pareto analysis used in this study.

2.1.2 Step 2: Performance analysis

The second step is the performance analysis focusing on the identified most influential parameters.

From both case study buildings, operational data regarding the installations were measured by sensors connected with the BMS and were logged for one whole year. When the obtained data is complete and consistent, the operational performance is analysed by means of a top-down approach. This implies that the performance is analysed first on building level, then on system level and finally on component level. For the analysis, the program RStudio is used which is developed for statistical computing and graphics [17].

2.1.3 Step 3: Energy saving potential

The third step is the investigation of the potential level of energy performance of both case study buildings. This is analysed using the L.E.A. [10, 11]. L.E.A. systematically analyses and identifies energy saving potential based on trends and patterns.

This results in developing a benchmark model using regression analysis based on empirical data. The benchmark model can be obtained in various ways. In this study, historical data of each building is used to create the benchmark model. After the model is validated, the performance is predicted and compared with measured data. The difference between the predicted energy consumption by the benchmark model and the actual measured energy consumption represents an indication of the energy saving potential.

2.2 Case study buildings

The study is based on two case study buildings: a nursing home and an office building located in the Netherlands (NL) (Table 1). The nursing home (case study 1) is constructed in 2016. This building has a gross floor area (GFA) of 7,070 m². The office building (case study 2) is constructed in 1993 and is renovated in 2009. This building has a GFA of 1,650 m².

	Case study building 1	Case study building 2
Building function	Nursing home	Office building
Location	Maastricht (NL)	Breda (NL)
Year of construction	2016	1993
Year of renovation	-	2009
GFA	7,070 m ²	1,650 m ²

Table 1. Specifications of the two case study buildings.

3 Case study 1

Case study 1 is the nursing home which contains a stateof-the-art HVAC system to provide a pleasant and healthy indoor climate.

The heating and cooling demand of the building is provided by an aquifer thermal energy storage (ATES) system with a heat pump. For heating peaks, the building is equipped with two gas-fired boilers. The cooling demand is generally provided directly by the ATES system without using the heat pump (passive cooling). For cooling peaks, the heat pump is used as a cooling machine to assist (active cooling). The heat and cold are supplied by three air handling units (AHUs) and by three concrete core activation (CCA) systems which are controlled by heating curves. On floor level, the supplied air is reheated by electrical duct heaters. In addition, the rooms are heated by radiators.

3.1 Most influential parameters

The most influential parameters are determined by ranking the maximum influence (positive or negative) of each parameter on the energy performance gap as mentioned in section 2.1.1. For the heating demand, the most important parameters are the maximum heating supply temperature (41%) and the efficiency of heat recovery (23%) (Fig. 2). On the cooling demand, the cooling set-point temperature (59%) and the minimum cooling supply temperature (23%) have the largest influence (Fig. 3).



Fig. 2. Ranking of the parameters based on their influence on the energy performance gap regarding heating demand.



Fig. 3. Ranking of the parameters based on their influence on the energy performance gap regarding cooling demand.

3.2 Performance analysis

3.2.1 Energy conservation

The heating and cooling demand of the building is mostly provided by the ATES system (Fig. 4). As usual, the extracted amount of energy is highly dependent on the outdoor temperature (Te). When the outdoor temperature is lower than about 10° C solely heat is extracted. In addition, cold is only extracted when the outdoor temperature is higher than about 19° C. In the figure, the slope of the points is a function of the building envelope, ventilation and infiltration air, the efficiency of the heating or cooling system and occupant behaviour. The lower the slope, the more energy-efficient the building. Between 10° C and 19° C it is important that the ATES system does not switch quickly between heating and cooling mode.

It has been laid down by Dutch water law that the ATES systems must not disturb the thermal balance in the subsoil. Since the heating and cooling demand is dependent on weather conditions and building use, which change every year, the government requires an equilibrium energy supply over a period of five years [18]. For this reason, it is important to determine the thermal balance of the ATES system. This is calculated for one year by means of the following equation.

$$Balance_{ATES} = \frac{Q_{cold} - Q_{heat}}{Q_{cold} + Q_{heat}}$$
(1)

where Q_{cold} denotes the cold extraction [kWh] and Q_{heat} denotes the heat extraction [kWh] by the ATES system.

During the whole measured period, it was found that the actual thermal balance is negative. This means that more heat is extracted than cold by the ATES system, which leads to a total cold surplus of 38%. The thermal imbalance can be solved by reducing the amount of extracted heat by the system. However, the generation of more heat by the gas-fired boilers is not sustainable. Therefore, a more suitable solution is to provide more heat to the warm well during the summer by means of a dry cooler on the roof of the building. This principle is called regeneration.



Fig. 4. Thermal energy extraction by the ATES system in relation to the outdoor temperature.

The temperature of the warm and cold well varies throughout the year due to the influence of heat losses to the surroundings and the amount of injection and extraction of heat and cold to and from the ground (Fig. 5). According to the design, the injection temperature into the warm well is about 18°C and the injection temperature into the cold well is about 7°C. However, during cooling mode, heat is injected with a temperature between 13°C and 18°C into the warm well. During heating mode, cold is injected into the cold well with a temperature between 9°C and 15°C. Therefore, the actual injection temperature into the warm well is too low and the injection temperature into the cold well is too high. This is caused by the energy imbalance.



Time [Month/yr] • Warm well • Cold well Fig. 5. Daily mean injection temperature into the warm and cold

01/18

03/18

05/18

07/18

3.2.2 Energy supply

well.

09/17

11/17

The AHUs and CCA systems are largely responsible for keeping the desired indoor temperature level. The supply temperature of AHU-1 during the year including the density of the measured points is illustrated in Fig. 6. In the figure, the solid red line presents the designed heating curve of AHU-1. However, the measurement data shows that the supply temperature is not in line with the heating curve and there is a considerable deviation. In comparison with AHU-1, the supply temperatures of AHU-2 and AHU-3 are much more in line with the heating curve. Consequently, the operational performance of AHU-1 has to be analysed more in detail and AHU-2 and AHU-3 are not analysed on component level.

Similarly, the supply temperature of CCA-1 including the density of the measured points is illustrated in Fig. 7, where the solid red line also indicates the designed heating curve of the system. CCA-1 is responsible for the same part of the building as AHU-1. It is notable during heating mode that this system (CCA-1) provides a lower supply temperature than the heating curve and AHU-1 provides a higher supply temperature than the heating curve. Therefore, one possibility is to reduce the supply temperature of AHU-1 when the supply temperature of CCA-1 is increased. This system optimization ensures that the indoor temperature remains virtually the same.



Fig. 6. Supply temperature of AHU-1 in relation to the outdoor temperature including the density of the measured points.



Fig. 7. Supply temperature of CCA-1 in relation to the outdoor temperature including the density of the measured points.

3.3 Energy saving potential

The main function of the thermal energy extraction is to bridge the difference between the outdoor temperature and the required indoor temperature and to provide hot water for hygiene purposes. Therefore, as a first attempt the relationship between the thermal energy extraction (dependent variable) and the outdoor temperature (independent variable) is determined using regression analysis. The results are evaluated by means of the coefficient of determination (R^2) which indicates to what extent the dependent variable is predictable. The closer the R^2 value is to 1, the better the relationship between the variables. Though there is no universal standard for a minimum acceptable R^2 value, 0.75 is often considered as a reasonable indicator of a good causal relationship [19].

In order to reduce the influence of time-dependent parameters (such as internal heat and solar irradiance), the daily mean thermal energy extraction is used. The extracted heat correlates linearly with the outdoor temperature (Fig. 8). This relationship is strong since the R^2 is 0.93. The cold extraction correlates non-linear with the outdoor temperature (Fig. 9). This leads to a still acceptable relationship with a R^2 of 0.76. Therefore, these models can be used as benchmark models. The difference between the predicted and measured energy demand indicates a yearly energy saving potential of 7% on the heating demand and 70% on the cooling demand.



Fig. 8. Regression analysis for the daily mean heat extraction as a function of the outdoor temperature.



Fig. 9. Regression analysis for the daily mean cold extraction as a function of the outdoor temperature.

4 Case study 2

Case study 2 is the office building which contains a commonly used HVAC system to provide a pleasant and healthy indoor climate.

The heating demand is generated by a gas-fired boiler which provides heat to the AHU and the radiators. The cooling demand is provided to the AHU by a cooling machine. The supply air temperature of the AHU is centrally controlled by means of a heating curve which is based on the outdoor temperature. In addition, the air is humidified when needed.

4.1 Most influential parameters

The most influential parameters are determined by ranking the maximum influence (positive or negative) of each parameter on the energy performance gap as mentioned in section 2.1.1. For the heating demand, the most important parameters are the heating set-point temperature (35%) and the maximum heating supply temperature (23%) (Fig. 10). On the cooling demand, in particular the cooling set-point temperature (89%), as well as the heating set-point temperature (4%), have the largest influence (Fig. 11). This means that the most influential parameters of this office building are also only related to the HVAC system.

https://doi.org/10.1051/e3sconf/201911105016



Fig. 10. Ranking of the parameters based on their influence on the energy performance gap regarding heating demand.



Fig. 11. Ranking of the parameters based on their influence on the energy performance gap regarding cooling demand.

4.2 Performance analysis

4.2.1 Energy conservation

Since this building is not 24/7 in operation, the heating and cooling demand is divided into two periods. This can provide insight into energy waste. The first period relates to the hours when the building is occupied, namely on weekdays between 09:00-17:00h (Fig. 12). The second period is related to the remaining hours when the building is not occupied (Fig. 13).

This results in a linear correlation between the thermal energy demand and the outdoor temperature during occupied hours. During unoccupied hours, especially the heating demand is fluctuating in relation to the outdoor temperature. Since this could be energy waste, the heat supply by means of the AHU is analysed.



Fig. 12. Thermal energy demand during occupied hours (weekdays between 09:00-17:00h) in relation to the outdoor temperature.



Fig. 13. Thermal energy demand during unoccupied hours (weekdays before 09:00h, weekdays after 17:00h and weekend) in relation to the outdoor temperature.

4.2.2 Energy supply

The AHU is largely responsible for keeping the required indoor temperature level. The supply temperature of the AHU during the year including the density of the measured points is illustrated in Fig. 14. Most of the time, the supply temperature is higher than the designed heating curve. Therefore, the performance of the AHU has to be analysed more in detail.

The supply and indoor temperatures are analysed by means of a typical week profile in autumn when the mean outdoor temperature is about 13°C (Fig. 15). This week profile clearly shows that the indoor temperature during occupancy remains within the desired range of 21-24°C. However, the supply temperature is relatively high during the night and weekend. This leads to the unnecessary heating demand during unoccupied hours. Therefore, the regulation of the system can be adjusted while still considering the indoor climate of the office building. Moreover, the larger energy peaks that will occur at the beginning of each workday must be taken into account when the indoor temperature during the night and the weekend are lower.



Fig. 14. Supply temperature of the AHU in relation to the outdoor temperature including the density of the measured points.



Fig. 15. Supply temperature of the AHU in combination with the indoor temperature during a typical week in autumn.

4.3 Energy saving potential

The heating and cooling demand fluctuates during the day due to time-dependent parameters, such as internal heat (occupants, equipment and lighting), solar irradiance, day and night set-points and stored energy in the thermal mass. In order to only involve data during occupancy, the regression analysis is based on data of weekdays between 09:00-17:00h. In addition, the daily mean demand during these hours is used in order to better include the dynamic behaviour of the HVAC system.

The heating demand correlates strongly with the outdoor temperature since the R^2 is 0.86 (Fig. 16). For the cooling demand, the relationship with the outdoor temperature leads to a R^2 of 0.66, which can be referred to as moderate (Fig. 17). Therefore, the model for cooling demand is less reliable than the model for heating demand. Nevertheless, these models can be used as benchmark models. As mentioned earlier, the difference between the predicted and measured energy demand is an indication of the energy saving potential of the building. This results in a yearly energy saving potential of 13% on the heating demand and 41% on the cooling demand.



Fig. 16. Regression analysis for the daily mean heating demand as a function of the outdoor temperature.



Daily mean cooling demand vs. Te | 09-17h

Fig. 17. Regression analysis for the daily mean cooling demand as a function of the outdoor temperature.

5 Discussion

5.1 Most influential parameters

According to the Pareto analysis, 20% of the most influential parameters accounts for roughly 80% of the energy performance gap. However, in this study not all cases meet this condition. The main reason for this is that the building-related parameters such as U-values and gvalues have been left out of consideration in order to remain in line with the scope of the study.

Although both case study buildings have almost the same most influential parameters, more buildings need to be investigated in order to conclude under which conditions the applied parameters in this study are generally applicable. Nevertheless, the Pareto analysis contributes to staying focused on the most important parameters of buildings.

5.2 Performance analysis

This study is performed by only using existing sensors of the BMS of both case study buildings including electricity and gas sensors, so that the method is directly applicable to other similar non-residential buildings. As expected, these sensors are sufficient for drawing conclusions about the overall operational HVAC performance. However, in order to draw more detailed conclusions about parts of the HVAC system, the number of sensors needs to be extended. Sometimes, the limited number of sensors can be compensated by calculations. However, real measurements are preferred for accuracy reasons.

For the pre-processing and cleaning of the target data, it is necessary to obtain information about the time of use of the building and maintenance of the HVAC system, for example adjustment in settings. This is essential to properly assess deviations in historical building data. Moreover, the data has to be considered as realistic, since it has found that the sensors of the BMS can be inaccurate or unreliable. Even though the data has been critically analysed, some inaccuracies are not quantifiable because calibration and maintenance of sensors and BMS are not considered in this study.

5.3 Energy saving potential

In this study, the energy saving potential is investigated based on the L.E.A. which creates benchmark models using data of other buildings with similar type, use, and other characteristics if available. Another possibility is to compare data before and after a large retrofit. However, collecting this data was not an objective of this study. Therefore, to stay in line with the L.E.A. approach, the benchmark models in both case studies are created by using their historical data. The performance modelling is conducted by comparing these benchmark models with measured data. This shows deviations from the regression line which gives an indication of the energy saving potential. However, the results are dependent on the accuracy of the created models. In this study, the models for cooling are less reliable than the models for heating. More research is needed to improve these models.

6 Conclusion

The process of building performance optimization starts with the decision of the building owner. Therefore, it is very important that building owners become more aware of how energy efficient their buildings can operate by controlling and improving controllable parameters. This study provided three main steps in order to systematically improve the underperforming HVAC systems.

The first step is the identification of the most influential parameters of the HVAC system including the occupant behaviour. Since the BMS in buildings contains massive data sets which can be stored unlimited in a database, the most influential parameters ensure to remain focused on parameters that are important for the underperforming installations. This leads to an effective performance analysis. In this study, the most influential parameters are obtained by the Pareto analysis which has proved to be an efficient approach.

The second step is the analysis of operational data regarding the most influential parameters of the HVAC system. It is advisable to determine during the design which sensors are essential for the analysis. If the sensors of the BMS do not measure the data related to these parameters, additional measurements are needed. Moreover, if the history of the logged data is too short, the data set is not suitable to see trends and obtain knowledge about the operational building performance. Therefore, it is necessary to log the data for at least a few months, preferably over seasons. By means of the logged data, the energy performance can be analysed. It is recommended to analyse the HVAC system by means of a top-down approach. This leads to an efficient process to execute. Only when there is a specific complaint from the building occupants, a bottom-up approach could be more appropriate. To interpret the data analysis, the data measured by the sensors has to be assessed for accuracy and reliability.

The third step is to investigate the energy saving potential of the underperforming HVAC system. For this, DM techniques have shown to be very valuable. In this study, regression analysis is used to create the benchmark models. This analysis is used in order to avoid a complex and time-consuming process compared to data models based on more advanced algorithms such as decision tree, artificial neural network and support vector machine. These models can be made in different ways. The models created in this study are based on historical data of the relevant building. Comparing the benchmark models with measured data leads to an indication of the potential reduction of the energy performance gap. This comparison is highly dependent on the reliability of the models. The results in this study show that a reduction of 7-13% on the heating demand and 41-70% on the cooling demand can be obtained. In addition, these models can be used to control the energy consumption.

However, regression analysis reduces the added value of large data sets because only a small number of variables is used to develop the model. This leads to a limitation of the discovered knowledge for practical application. Therefore, future research is needed to discover more insights by means of other DM techniques.

Acknowledgements

The authors would like to thank the involved staff of Huygen Engineers & Consultants and Eindhoven University of Technology for their contribution and support.

References

- International Energy Agency, Total Energy Use in Buildings: Analysis and Evaluation Methods (Annex 53), (2013).
- 2. CarbonBuzz, Retrieved from www.carbonbuzz.org, (2018).
- C. v. Dronkelaar, M. Dowson, C. Spataru, D. Mumovic, A Review of the Regulatory energy Performance Gap and its Underlying Causes in Nondomestic Buildings, *Frontiers in Mechanical Engineering*, (2016).
- 4. European Commission, Energy Efficiency and its contribution to energy security and the 2030 Framework for climate and energy policy, (2014).

- 5. C. Fan, F. Xiao, Assessment of building operational performance using data mining techniques: a case study, *Energy Procedia* **111**, 1070-1078, (2017).
- 6. S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, *Energy and Buildings* **88**, 395-408, (2015).
- 7. P. de Wilde, The gap between predicted and measured energy performance of buildings: A framework for investigation, *Automation in Construction* **41**, 40-49, (2014).
- 8. Oregon Department of Energy, Retro-commissioning in Silver Falls School District, Retrieved from www.oregon.gov, (2014).
- 9. B.M. Bartlett, Powerful Strategic Planning Rules That Could Revolutionize Your Business And Life, Retrieved from http://benmbartlett.com, (2018).
- 10. B. Abels, F. Sever, K. Kissock, D. Ayele, Understanding Industrial Energy Use through Lean Energy Analysis, (2011).
- 11. M. Donnelly, J. Kummer, K. Drees, LEAN Energy Analysis: Using Regression Analysis to Assess Building Energy Performance, (2013).
- 12. A.J. Huls, Systematic energy performance assessment in operating office buildings, (2016).
- 13. I. Schoenmakers, W. Zeiler, G. Boxem, KPI's energy consumption of isolation rooms in hospitals, *Proceedings REHVA World Congress CLIMA 2016*, (2016).
- 14. W. Vink, W. Zeiler, Re-Commissioning the HVAC of an Office Building: Systematic Performance Evaluation according to L.E.A., *Proceedings CIBSE Technical Symposium 2018*, (2018).
- 15. A.J. Huls, B. Lops, W. Zeiler, Time series Analysis for Re-Commissioning of Building Service installations, *Proceedings ISTE conference 2018*, (2018).
- EnergyPlus, Retrieved from http://energyplus.net, (2018).
- 17. RStudio, Retrieved from http://rstudio.com, (2018).
- 18. Agentschap NL, Energiezuinig koelen met warmteen koudeopslag, (2011).
- 19. G.R. Ruiz, C.F. Bandera, Validation of Calibrated Energy Models: Common Errors, (2017).