Project acronym	B4B
Project full name	Brains for Building's Energy Systems
Grant No	MOOI32004
Project duration	May 1, 2021 – December 31, 2025

Deliverable 1.06
Tools for Continuous Commissioning with a) Pareto LEAN energy analysis and b) data trends of continuous monitoring, for detection and GBS plug-ins
Srinivasan Gopalan (TU/e), Agata Rijs (Spie), Shobhit Chitkara (TU/e), Anand Thamban (TUe), Rick Kramer (TUe)

Work package	1
Result	2
Lead beneficiary	TU/e
Due Date	1 January 2025
Deliverable Status	Final
File name	B4B WP1 D1.06 Tools for Continuous Commissioning_draft.docx

The project was carried out with a Top Sector Energy subsidy from the Ministry of Economic Affairs and Climate executed by the Netherlands Enterprise Agency. The specific subsidy for this project concerns MOOI subsidy round 2020

SUMMARY[1]

Fault impact analysis is an important step in developing Fault Detection and Diagnosis (FDD) tools for Heating, Ventilation and Air-Conditioning (HVAC) systems. It provides an insight into how HVAC systems react in the presence of operational faults and helps to prioritise which faults to focus on. The research in the existing literature has mostly focused on modelling several faults occurring in various HVAC systems and components to evaluate the effect of the fault on energy performance and the predicted thermal comfort of occupants. However, the real frequency of fault occurrence, an important factor which affects the overall impact a fault has on the system's performance, has barely been addressed in the reviewed literature. This paper addresses this gap by considering real fault occurrence frequency data in addition to the effect of the faults on the energy performance of the building to assess their total energy impact. A real office building in the Netherlands was modelled using DesignBuilder, and its energy performance was simulated using EnergyPlus. Air Handling Unit (AHU) faults were introduced using the native fault object and Energy Management System (EMS) features available in EnergyPlus. The fault occurrence data was extracted from AHU work orders using text mining. The fault modelling and text mining results were combined to get the total energy impact of the fault, and subsequently, the faults were prioritised using the Pareto principle. The research identified that fan failure, Heat Recovery Wheel (HRW) failure, fan stuck at 50%, and Heating Coil Valve (HCV) stuck at 0% are the priority faults for winter and fan failure is the priority fault for summer.

[1] This deliverable is also published as a journal paper Energy and Buildings 319C (2024) 114476 "Fault prioritization for Air Handling units using fault modelling and actual fault occurrence data" https://doi.org/10.1016/j.enbuild.2024.114476

TABLE OF CONTENT

Summary 2
Table of Content 3
1 Introduction 4
2 Methodology 6
2.1 Case study building: Modelling and calibration 6
2.2 Selection of faults for focus and fault modelling 9
2.3 Effect of fault on energy performance 10
2.4 Fault occurrence frequency analysis 10
2.5 Fault impact and fault prioritisation using the Pareto principle 12
3 Results 13
3.1 Effect of faults on energy performance 13
3.2 Frequency of fault occurrence from text mining 16
3.3 Total energy impact and fault prioritisation 17
3.4 List of BMS alerts 18
4 Discussion 19
4.1 Reflection on results 19
4.2 Comparison of results with literature 20
4.3 Contributions, limitations and future research 20
5 Conclusions 22
Appendix A: Case study building model construction information 23
Appendix B: HVAC operation setpoints of the simulation model 24
Appendix C: Lexicon used in fault frequency analysis 25
Appendix D: Effect of faults on energy exchanged at the heating and cooling coils 27
Appendix E: Reaction of master–slave controllers to sensor bias and incorrect setpoint faults 28
Appendix F. Supplementary data 30
References 31

1 INTRODUCTION

Buildings account for 36 % of the total energy consumption worldwide [1]. Operational faults that occur in the Heating Ventilation and Air-Conditioning (HVAC) systems of buildings, in addition to reducing the thermal comfort of occupants, could lead to an energy waste of up to 30 % of this energy consumption [2]. In The Netherlands, buildings constituted 37 % of the final energy consumption in 2022 [3]. A 30 % energy wastage corresponds to 49.68 billion kWh, around 1.05 times the electricity generated by renewables in 2022 in The Netherlands (46.88 billion kWh as retrieved from Statistics Netherlands [4]). Preventing this energy waste will help reduce the investment needed to install the infrastructure needed for energy generation, transmission and distribution.

Fault Detection and Diagnosis (FDD) is a process that identifies these operational anomalies and reduces energy waste, service costs and equipment downtime [5]. Over two decades, the research in this field has grown with a focus on data-driven FDD [6]. However, there is a lack of widespread uptake of FDD tools in practice due to (i) insufficient sensor granularity, (ii) inadequate field testing, validation, and real-time implementation and (iii) lack of scalability & transferability [2,5–7]. In this regard, the domain of simulation-based fault impact analysis could play a very important role in accelerating the development and uptake of FDD tools. These studies could help to throw light on the impact of faults on multiple Key Performance Indicators (KPIs) such as energy performance, thermal comfort etc. Subsequently, by prioritising a small proportion of faults that account for a very high share of the impact on KPIs, the development of FDD tools could be accelerated. In the context of The Netherlands, fault prioritisation helps to identify which faults account for the energy waste equivalent to the electricity generated from renewables. It helps develop FDD tools to detect and diagnose these faults. A literature review of existing research on simulation-based fault impact analysis was performed. The reviewed research articles were analysed on multiple aspects as presented below:

i. *Number of faults studied*: The number of faults studied varies greatly amongst the reviewed articles. For example, Lu et al. [8] studied various faults relating to sensors, ducts, pipes, valves, coils, dampers, chiller, cooling tower, boiler, pump, fan, control & scheduling. Similarly, Li and O'Neill [9] also studied various faults covering most of the above-mentioned system components. In papers studying multiple faults, the trend has been to rank the faults based on their effect. In contrast, there has also been research where only one fault has been studied. Li et al. [10] studied the effect of thermostat bias on three different HVAC supply systems: variable refrigerant flow, ground source heat pump & chiller-boiler systems.

ii. Software employed: EnergyPlus is the most widely used simulation software in the literature. The research by Lee and Yik [11], Basarkar et al. [12], Wang and Hong [13], Zhang and Hong [14], Li and O'Neill [9], Li et al. [10], Zhong et al. [15] use EnergyPlus. While other software such as Modelica in Lu et al. [8], TRNSYS in Khire and Trcka [16], and Simulink in Ginestet et al. [17] have also been used, co-simulations such as EnergyPlus-CONTAM in Lu et al. [18], EnergyPlus-Modelica in Huang et al. [19], and Simulink-TRNSYS in Verhelst et al. [20] have also been employed. As summarized by Li and O'Neill [21], EnergyPlus offers greater fault modelling capabilities, Modelica offers flexibility for the development of dynamic fault models, and other software (Trnsys, HVACSIM+, DOE-2, & EnergyPro) have limited capabilities for modelling faults.

iii. Case study building used: The US Department of Energy (DOE) prototype small, medium & large office buildings reported in Deru et al. [22] were used by Zhang and Hong [14], Li and O'Neill [9], Lu et al. [8], Huang et al. [19], Basarkar et al. [12], and Wang and Hong [13]. The medium office building with adapted, local weather conditions was also used for research outside the US, such as by Li et al. [10] and Zhong et al. [15]. Other specific case study buildings were also used, such as in the research of Lu et al. [18], Verhelst et al. [20], Lee and Yik [11], Khire and Trcka [16], Otto et al. [23] and Ginestet et al. [17]. While the use of the DOE prototype buildings offers the possibility to draw generic conclusions regarding the impact of faults, the use of specific case study buildings offers the possibility of experimental validation.

iv. Indicator studied: Energy is the most widely used performance indicator to study the effect of faults. It has been used in all reviewed studies in the form of one or more of the following- total energy, total HVAC energy, energy of individual systems such as cooling and heating, energy of components such as boiler, pumps, heat recovery etc. Other KPIs like Predictive Mean Vote (PMV), Predicted Percentage of Dissatisfied (PPD) and outdoor air ratio have also been frequently used. Lu et al. [8] used a comprehensive set of 13 KPIs to study the effect of faults. Li and O'Neill [9] and Lu et al. [18] used the Sobol index to analyse the sensitivity of KPIs such as energy, PPD & outdoor air ratio to the implemented faults. The general trend is to rank the simulated faults based on the employed KPIs.

v. Study of simultaneous faults: The literature also researches the effects of multiple faults occurring together. Zhong et al. [15] concluded that multiple faults can cause either synergetic or antagonistic effects. Khire and Trcka [16] demonstrated complex coupling effects of faults. Li and O'Neill [9] combined the studied faults' occurrence into fault modes, constituting a simulation input file.

vi. Fault mode occurrence: Finally, research has also been performed on incorporating probability distribution functions for fault intensity/failure mode occurrence. This was outlined as a recommendation for future research by Li and O'Neill [21]. The research of Otto et al. [23], Li and O'Neill [9] and Lu et al. [18] incorporated mathematical functions for the studied faults into fault modes which constituted a simulation input file.

Based on the reviewed body of literature, it was identified that there is a lack of inclusion of real-world fault occurrence frequency data from maintenance work orders in a simulation-based fault impact analysis study. The fault frequency is an important parameter which affects the energy impact of faults on buildings. A fault that has a minimal effect on energy consumption but occurs very frequently can have a much greater impact on energy consumption. Therefore, the primary objective of this research is to combine real-world fault occurrence frequency data extracted from maintenance work orders with the effect of faults obtained through building energy simulations to quantify the total energy impact of faults. Additionally, the reviewed studies have not made a selection of important faults in their research through prioritisation. Such a selection would help to narrow down the focus to a small number of faults which have a very high impact. Consequently, the secondary objective of this research is to select a set of prioritised faults based on their total energy impact. Furthermore, this paper links the results of the fault prioritisation obtained with those of engineering practice. When these faults have to be detected and diagnosed, an alert must be raised to signal their diagnosis. Thus, the tertiary objective of this research is to study the Building Management System (BMS) alerts available in practice and identify if any new alerts need to be developed for the studied faults.

The technique of text mining has been employed to obtain real fault occurrence frequency from maintenance work orders because the fields containing information relevant to the faults are free text fields. Also, there has been an increase in scientific research focusing on extracting useful information from maintenance work orders using text mining recently. A comprehensive literature review by Shamshiri et al. [24] identifies multiple articles using text mining to obtain information such as frequency of warning and failure generation [25], condition of building systems such as electrical, plumbing, HVAC, fire and elevator [26], identifying locations of fault occurrences and systems that require frequent maintenance [27], severity of service requests [28], to benchmark buildings' operational performance [29], and to benchmark maintenance operations [30]. This research also contributes to the growing literature on text mining of HVAC maintenance work orders.

The chosen system for study is the Air Handling Unit (AHU) due to its importance in the HVAC system. The AHU provides heating and cooling along with air distribution to meet the indoor space conditioning requirements [31]. Its customised nature and absence of quality system integration could lead to the failure of hardware and controller faults [31]. Since the number of AHUs is 20 times that of the available maintenance personnel and complex control strategies require expensive external consultancy for diagnosing issues [32], there is a high potential for implementing Automated FDD (AFDD) and predictive maintenance solutions for AHUs. Thus, this study focuses on prioritising AHU faults.

2 METHODOLOGY

The methodology to prioritize faults comprises three parts: (i) Fault modelling and simulation to study the effect of faults on energy performance; (ii) text mining to obtain fault occurrence frequency value; and (iii) assessing the total energy impact by combining the effect on energy performance and fault occurrence frequency followed by prioritisation. Fig. 1 provides an overview of the methodology.

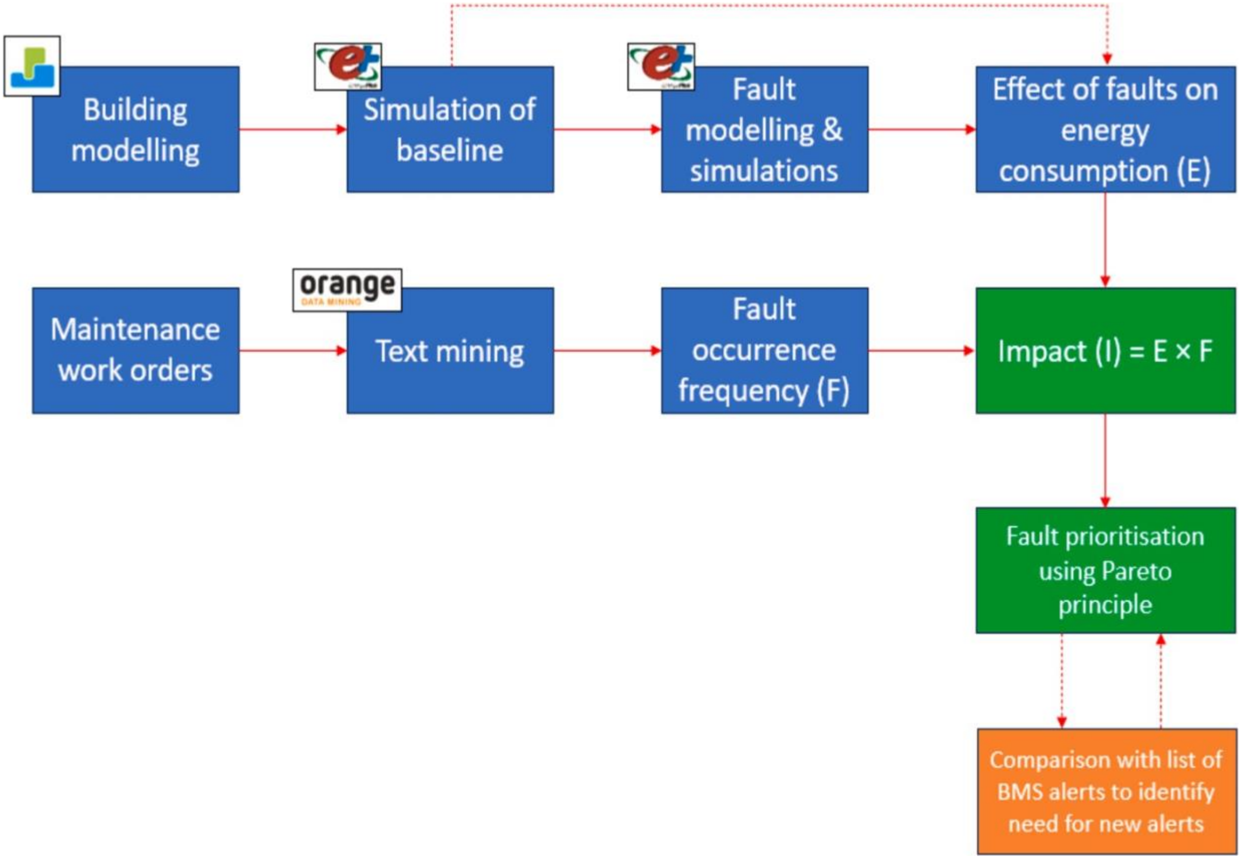

Fig. 1. The overall methodology to determine the impact of faults.

The chosen case study building was first modelled using the software DesignBuilder, and its energy performance was simulated using EnergyPlus. This simulation represented the fault-free scenario and will henceforth be referred to as baseline simulation. Subsequently, the studied HVAC faults were modelled using EnergyPlus, and the energy performance of the case study building was simulated under the influence of these faults. The effect of each fault on the energy performance of the building was determined by comparing it with the energy performance of the baseline simulation. Simultaneously, maintenance work orders were used to obtain the frequency of occurrence of the studied faults through text mining. The total energy impact of each fault was calculated by combining the effect on energy performance and its frequency of occurrence. Subsequently, the faults were prioritised based on the calculated impact values. Additionally, the list of prioritised faults was compared against available BMS alerts to determine if any new alert is needed when a prioritised fault has been diagnosed in future using an AFDD tool.

2.1 Case study building: Modelling and calibration

The chosen case study building is an office building of the building services company Kropman, located in the city of Breda, The Netherlands. This building is a representative office building in Western Europe. Furthermore, this building is a living laboratory, serving as a testbed for various research activities in the past and the present. Additionally, relevant sensor and energy meter data is available for the calibration of its model leading to the choice of this building as the case study. This building was built in 1993 and renovated in 2009. The building is divided into three HVAC zones: North, South and 1.05. Fig. 2(a) illustrates the external appearance of the building, and Fig. 2(b) illustrates how the zones are laid out on the first floor of the building. The North & South zones contain offices on three floors: ground floor, first floor and second floor. Zone 1.05 exists only on the first floor. The space below zone 1.05 on the ground floor contains a canteen and a meeting room which are a part of zone North.

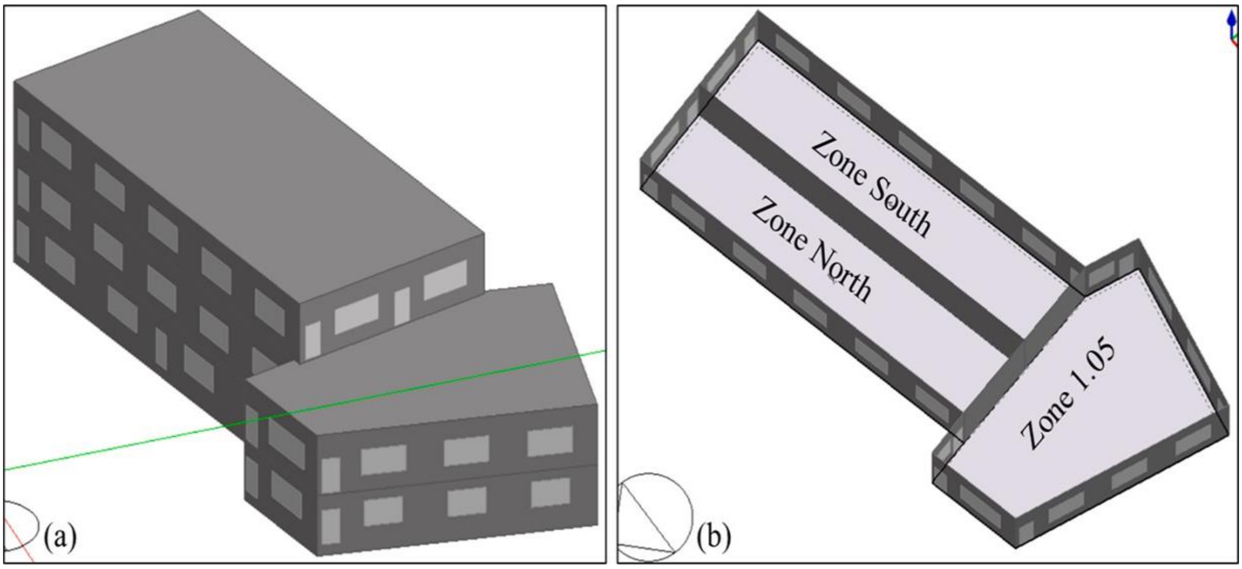

Fig. 2. (a) External view of the case study building and (b) division of the building into zones on the first floor (adapted from Thamban [34]).

The air supply and distribution are done through a central AHU, which contains supply and return fans, supply and return filters, a Heat Recovery Wheel (HRW), and a heating coil. The cooling coil is not present in the AHU itself but rather as three after-coolers, providing cooling for the three zones. Fig. 3 illustrates the AHU and the after-coolers for the three zones. The energy supply is provided by a gas-fired boiler for the heating season and a chiller for the cooling season. In addition to the heating coil in the AHU, the space heating demand is also met through radiators present in all three zones. A detailed explanation of the HVAC system & equipment characteristics can be found in Chitkara [33].

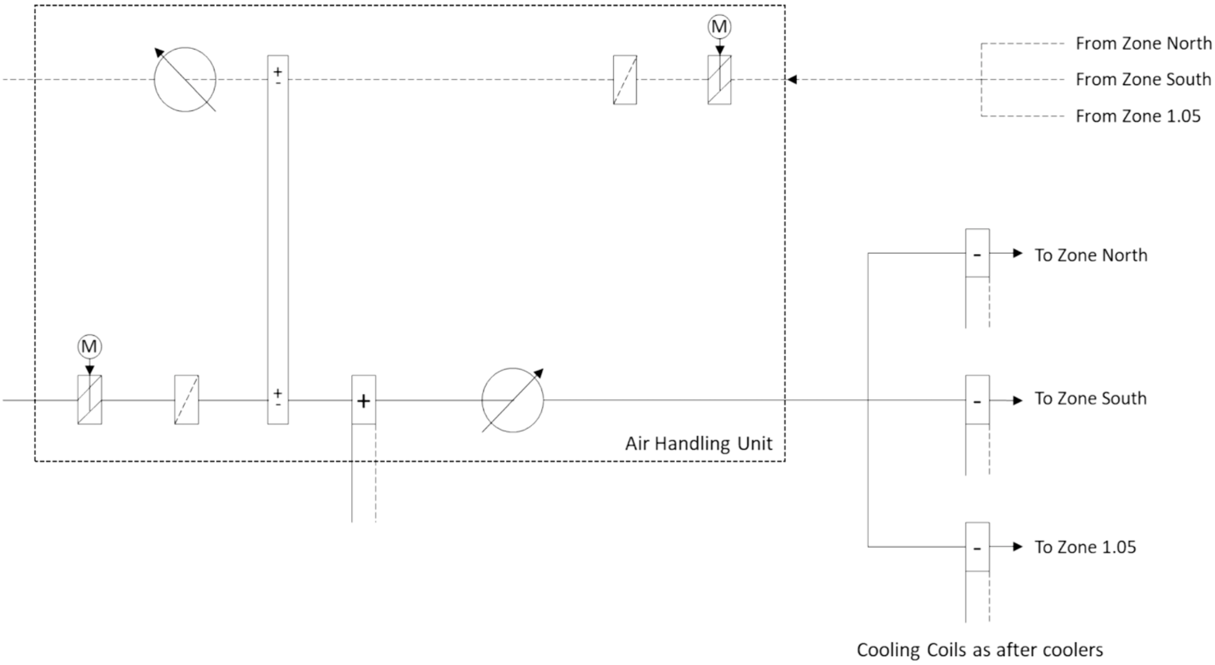

Fig. 3. An overview of the HVAC system at the case study building Kropman Breda.

The building was modelled using DesignBuilder, and its energy performance was simulated using EnergyPlus. Appendix A lists the construction information used for the model of the building, and Appendix B lists the parameters used in the HVAC system modelling. The design of the heating season model included an AHU with HRW, a supply fan, a return fan, and a heating coil. Due to the limitations of modelling the after-coolers in DesignBuilder, the cooling coils were modelled by placing a fan coil unit in the respective zones and the simulation calculated the energy performance of the system by considering the air delivered at the zone air terminal to pass through the fan coil unit before being supplied to the zones. However, the cooling coils in this model lack controllers, an object that adjusts the cooling coil valve (CCV) opening depending on the value of the supply air temperature (SAT), leaving the cooling coil to meet the setpoint. The presence of these controllers is important for adjusting valve opening depending on the type and intensity of fault introduced and, consequently, to simulate the energy performance of the HVAC system under the presence of faults during the cooling season. Therefore, a second HVAC model was designed to represent the cooling season operation of the case study building where three AHUs were placed in parallel, with each AHU containing a cooling coil placed after the supply fan. This model accurately represents the airflow streams to and from the three zones: cooling coils, the chilled water flow system (along with the pumps), and the chiller of the building. Fig. 4(a) illustrates the placement of the cooling coil after the supply fan in the AHU. Fig. 4(b) illustrates the layout used for the cooling season model in DesignBuilder.

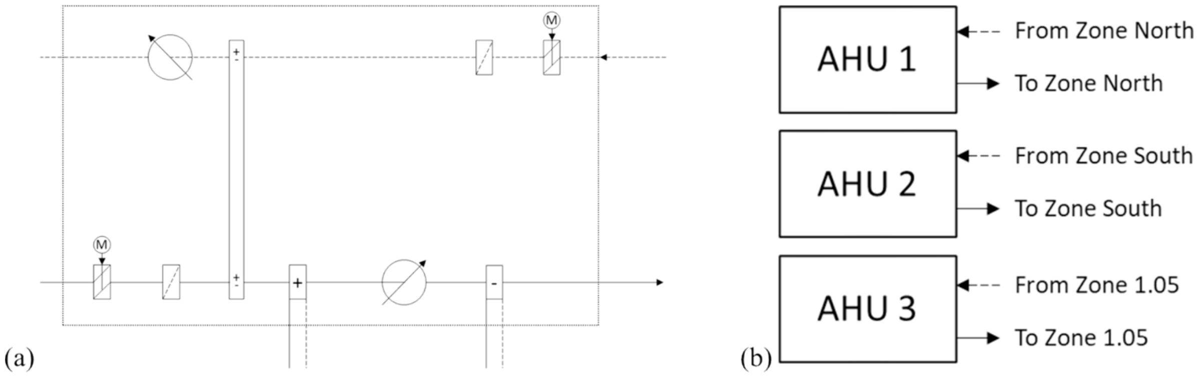

Fig. 4. (a) AHU with cooling coil placed after supply fan (b) HVAC model layout of three AHUs in parallel for summer (cooling season) simulation.

Information provided by experts indicated that the median HVAC maintenance request was resolved within 21 hours after it was reported. In future, if a BMS containing an AFDD add-on raises an alert after diagnosing a fault, the impact of the fault will persist for 21 hours before being fixed. Therefore, to study the effects of the various faults on the building's energy consumption, it was decided to simulate the performance of the building over 24 hours each for winter and summer. The chosen dates for fault simulation were 9th February 2023 for the winter season and 15th June 2023 for the summer season. The data recorded at the weather station in Gilze-Rijen (located 15 km from the building) was retrieved from the Royal Netherlands Meteorological Institute's database and was converted to EnergyPlus weather files. The baseline models were calibrated according to the ASHRAE (American Society of Heating Refrigerating and Air-conditioning Engineers) Guideline-14 using the error metrics Normalised Mean Bias Error (NMBE) and Coefficient of Variation of the Root Mean Square Error (CV-RMSE) whose formulae, taken from Gestwick and Love [35] and ASHRAE [36], are provided in Eq. (1) and Eq. (2):

$$\mathrm{NMBE} = \frac{\sum_1^n (y_i - \hat{y}_i)}{(n-p) \times \bar{y}} \times 100 \qquad \text{Eq. (1)}$$

$$\mathrm{CV-RMSE} = \frac{\left[\sum_1^n ((y_i - \hat{y}_i)^2)/(n-p)\right]^{1/2}}{\bar{y}} \times 100 \qquad \text{Eq. (2)}$$

Where

y_i is the measured data point.

y_i is the simulated data point.

y is the mean value of the measured data points.

n is the number of data points.

p is the number of parameters in the model (0 for NMBE & 1 for CV- RMSE) [35].

The BMS of the case study building logged data of the hourly electricity consumption of the chiller, hourly electricity consumption of the rest of the HVAC system (including fans, HRW & pumps), and hourly natural gas consumption. Therefore, the calibration was performed result, the baseline model could not be calibrated for natural gas consumption. The simulated natural gas consumption value for 9th February was 944.46 kWh, and the effects of heating season faults on the energy performance of the building were studied by considering this performance as the baseline. The obtained error metrics for the total HVAC electricity of the heating season model (NMBE: 4.34 %, CV-RMSE 11.34 %) and the chiller electricity consumption of the cooling season model (NMBE: -4.30 %, CV-RMSE: 29.61 %) were within the limits of ± 10 % for NMBE & 30 % for CV-RMSE specified in ASHRAE Guideline 14–2014 for calibration using hourly data. Fig. 5 shows the simulated and measured total HVAC electricity consumption for the simulated winter day. Similarly, Fig. 6 shows the simulated and actual chiller electricity consumption for the simulated summer day. It must be noted that the night ventilation control strategy was not implemented for the cooling day simulation.

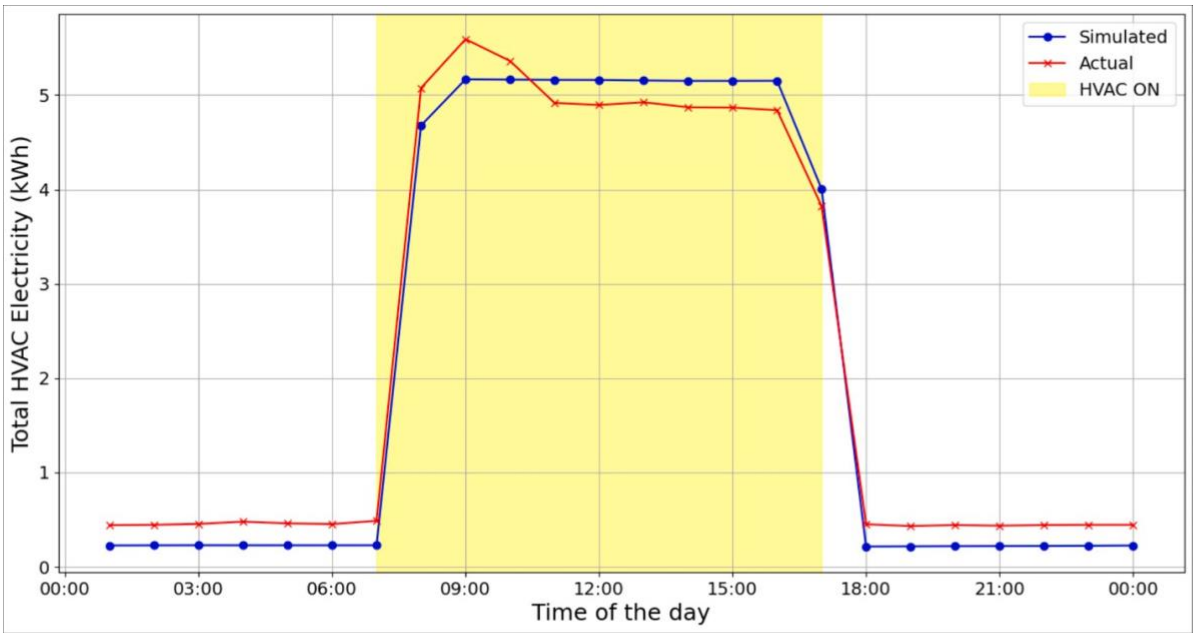

Fig. 5. A comparison of the simulated and measured total HVAC electricity consumption for the winter day.

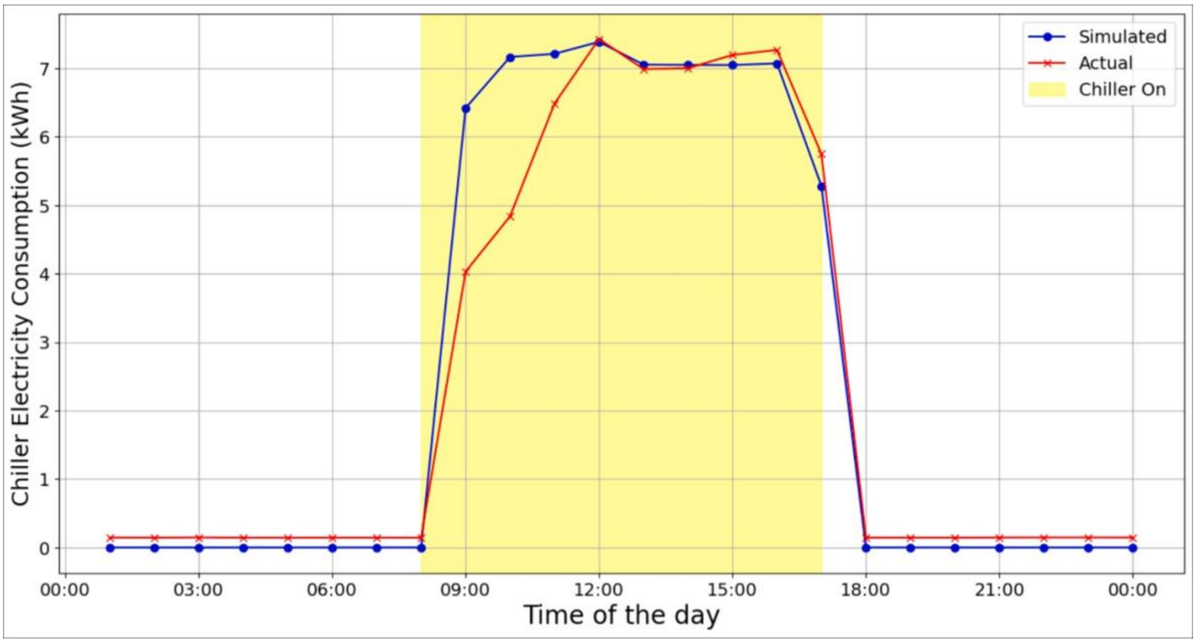

Fig. 6. A comparison of the simulated and measured chiller energy consumption for the summer day.

2.2 Selection of faults for focus and fault modelling

Gunay, Shen and Yang [37], provided an overview of the AHU faults (along with Variable Air Volume (VAV) terminal faults) studied in the literature over twenty years. These faults can be grouped into the categories of design, acute and degradation faults. The scope of this research is limited to only non-degradational, acute faults concerning the heating coil, cooling coil, fans, HRW, sensor bias and control setpoint. Applying this criterion, a selection of faults was made to be studied. Table 1 lists the faults, the fault severity settings studied, and how they were modelled in EnergyPlus. The fault severity settings for Heating Coil Valve (HCV), CCV & fan stuck faults were chosen based on the work of Li and O'Neill [9], who obtained the values from a combination of literature review and modelling. The sensor bias and control setpoint fault severity setting were chosen based on other reviewed literature.

Table 1 List of AHU faults chosen to be studied.

Component	Fault	Fault settings studied	EnergyPlus feature used for fault modelling
Heating coil & valve actuator	Heating coil valve stuck	0 %, 100 %	Energy Management System (EMS): Coil valve stuck- the minimum, maximum & actual mass flow rate of water through the coil was set at a value corresponding to the valve stuck % Coil valve leak- the minimum mass flow rate of water through the coil was set at a value corresponding to the valve leak % Fan failure/stuck- the air mass flow rate set at a value corresponding to the airflow %
	Heating coil valve leak	0 %, 40 %	
Cooling coil & valve actuator	Cooling coil valve leak	0 %, 50 %	
	Cooling coil valve stuck	0 %, 40 %	
Fans	Fans failure	0 % airflow	
	Fans stuck	50 % airflow, 85 % airflow	
Heat recovery wheel (only for heating day)	Wheel failure	0 % rotation	Heat Recovery (availability schedule set to 'Off 24/7')
Sensor bias	Supply air temperature sensor bias	±2°C bias	Operational Faults (offset of ± 2°C introduced)
	Indoor air temperature sensor (thermostat) bias	±2°C bias	
Control setpoint	Incorrect supply air temperature setpoint (high)	+2°C	Setpoint Managers: minimum & maximum setpoint values changed for the heating model;
	Incorrect supply air temperature setpoint (low)	-2°C	Setpoint Managers: setpoint values at outdoor low and high temperatures changed for the cooling model

2.3 Effect of fault on energy performance

The effect of each fault on energy performance was obtained by comparing the energy consumption under the faulty condition against the energy consumption of the baseline model. While the effect on thermal comfort is not in the scope of this study and forms part of another upcoming study, the absolute value of the energy deviation was considered for the effect on energy performance due to which faults that lead to lower energy consumption but reduce thermal comfort such as fan failure, fan stuck at a low speed, HCV and CCV stuck closed/at a very low level etc. were also considered as very important faults (depending on the magnitude of deviation). For the winter day faults, the KPI used to study the effect on energy performance is the total HVAC energy consumption. For the summer day faults, the KPI used to study the effect on energy performance is the sum of the chiller energy consumption and pump energy consumption.

2.4 Fault occurrence frequency analysis

This section presents the methodology adopted to obtain fault occurrence frequency values using text mining. Fig. 7 illustrates the methodology developed for the text mining process using the software Orange.

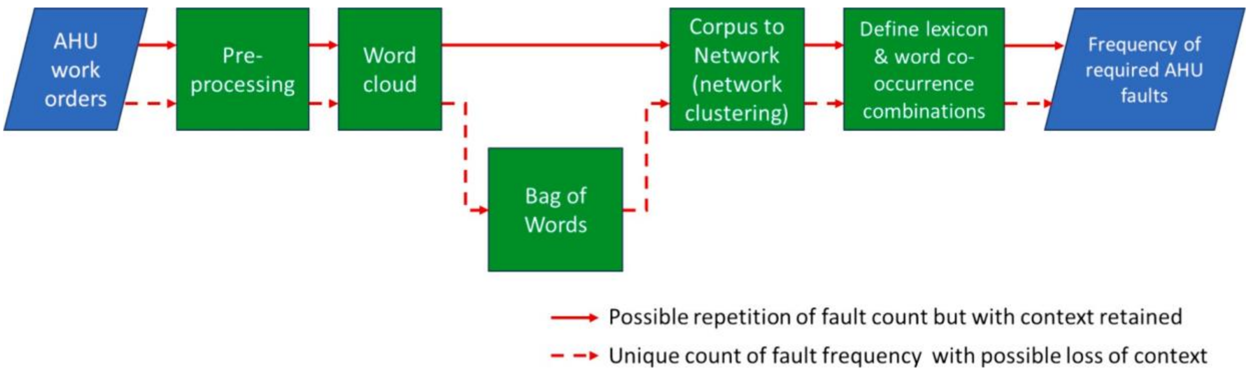

Fig. 7. An overview of the methodology adopted to obtain fault frequency using text mining.

The work orders that were used as input for the text mining were 11,255 non-annual AHU maintenance work orders raised with SPIE Building Solutions B.V., a building services company. These work orders were raised over 5 years between May 2018 to April 2023 and pertain to 1,822 non-residential buildings (municipality buildings, train stations, schools, supermarkets, production halls, museums, office buildings, university buildings, healthcare buildings, airports, shops). Out of 76 fields available to be filled in each maintenance work order, the data filled in five fields that could provide relevant information about the faults were chosen for analysis. These fields are: (i) subject of the reported issue, (ii) work order text, (iii) cause, (iv) follow-up action, and (v) work carried out. It is important to note that these work orders were filled in Dutch.

Following the selection of the relevant fields, pre-processing was performed, which involved the removal of standard Dutch stop words, tokenisation, and transformation to convert all letters to lowercase and remove accents. The pre-processing was performed iteratively by observing the word cloud obtained at the end of each iteration. Fig. 8 shows the word cloud obtained at the end of the final iteration. The word cloud has been translated into English.

The word cloud obtained in Fig. 8 contains non-HVAC-related words such as west, campus, why etc. It also contains HVAC-related terms outside the scope of the faults being studied, such as frost thermostat, pump, etc. To extract results relevant to the studied faults, a lexicon of 50 keywords, which could be used to describe the studied AHU faults, was developed. Appendix C lists the used keywords and their English translation. These keywords were obtained from how the faults have been referred to in the literature and translated to obtain their Dutch language equivalent for use in the lexicon. For example, if the fault 'heating coil valve stuck' is considered, the term 'heating coil valve' is represented as 'regelafsluiter verwarming' in Dutch and the word 'stuck' is represented by the word 'vast' in Dutch. Sometimes there could be a word variation that could be used, e.g. verwarmer (heater) instead of verwarming (heating). Considering all such possible variations, the lexicon of 50 keywords was developed for the studied faults.

Fig. 8. The word cloud obtained at the end of pre-processing.

To calculate the frequency of fault occurrence, the 'Corpus to Network' feature (also known and hereafter referred to as network clustering) was used. This feature provides information on how many times two given words occur in a given window [38]. The count of co-occurrence of specific combinations of keywords from the defined lexicon indicates how many times a particular fault may have occurred. There are three hyper-parameters in the network clustering feature driving the obtained results: threshold (the minimum number of times two words appear in a window of 2 × window size + 1), frequency threshold (the minimum frequency each word must have to be included in the network results), and window size [38]. The values of both threshold and frequency threshold were set to 1 to consider all word occurrences and co-occurrence counts. A window size of 7 was used based on the research of Levy et al. [39]. This means that for every given word, the algorithm searches for the occurrence of another word within a window of 7 positions to its right and 7 positions to its left. Specific combinations of two keywords, the co-occurrence of which could be construed as a possibility a fault could have occurred, were developed for all the studied AHU faults. These combinations have been listed in the supplementary material provided in this article. The values of all co-occurrences of the listed combinations per fault were obtained and summed up to acquire the total count (or frequency) of the number of times these faults may have occurred in the analysed work orders.

It must be noted that the network clustering performed as outlined above gives a possible value for fault frequency but does not guarantee that each of the keywords occurred only once in each work order. Since the studied fields are free-text fields, it is possible that one or more keywords could have occurred more than once in the work order and there is a chance that the co-occurrence for a given combination of keywords is more than one in the analysed window size. Therefore, to overcome this, an additional feature called 'Bag of Words' [40] was simultaneously applied which extracted the unique words occurring in each work order. Subsequently, a corpus was created where each work order contained only the extracted unique words before performing network clustering. The window size for the network clustering was adjusted to include the length of the full work order. The obtained fault frequency results through this process can also be interpreted as the number of work orders issued for a specific fault. Eq. (3) presents the formula for fault frequency.

$$\text{Fault frequency (F)} = \sum \text{values of all co} - \text{occurrences of the listed combinations for the fault} \qquad \text{Eq. (3)}$$

When divided by 5 (the number of years for which maintenance work orders were considered), this number gives the annual work orders raised for each fault. This result is subsequently used for calculating the total energy impact.

2.5 Fault impact and fault prioritisation using the Pareto principle

Once the effects of faults on energy consumption and the frequency of faults were obtained, the overall impact of the fault was calculated by multiplying the energy effect and frequency, as indicated in Eq. (4) below.

$$\mathrm{I} = \mathrm{E} \times \mathrm{F} \qquad \text{Eq. (4)}$$

where,

E is the effect of the fault on energy consumption expressed in kWh per day

F is the fault frequency expressed as the number of work orders per year (normalised fault frequency = fault frequency / total number of work orders analysed), and

I is the fault impact, expressed in kWh per day per year

Subsequently, the Pareto Principle was used to prioritise the faults. First introduced by Vilfredo Pareto in 1896–1897 and later pointed out by Dr. Joseph Juran in the 1950s to be a universal principle, it states that 20 % of the causes lead to 80 % of the effects [41]. The calculated total energy impact values of the faults were first arranged in descending order. Subsequently, the (first 20 % of) faults that cumulatively led to a total energy impact of 80 % compared to the baseline scenario were prioritized and recommended for monitoring by future FDD tools. Applying the Pareto Principle to prioritise faults represents a novel contribution to the literature on fault impact analysis because the general practice in the reviewed literature has been to rank the studied faults based on their sensitivity to specific, chosen KPIs. Through the Pareto Principle, a specific threshold (80 %) is used as the cutoff point to select faults for further focus in subsequent steps.

Once the faults have been prioritised, the subsequent logical step the presence of these high-priority faults which also requires the presence of alerts to raise a signal if the faults have been diagnosed. Therefore, a list of BMS alerts available in practice was obtained from the PULSE Core platform of SPIE Building Solutions B.V. to determine if the development of any additional alerts was needed for the prioritized faults.

3 RESULTS

This section presents the effects of AHU faults on the energy performance of the case study building, the results of the fault frequency analysis, and the overall impact of each fault. Subsequently, the prioritised faults for both the heating and cooling seasons, after applying the Pareto principle, are also presented. Finally, commonly available BMS alerts are also presented.

3.1 Effect of faults on energy performance

In this section, the effects of the studied faults on the energy performance of the case study building are presented. The indicator used for energy performance is the total HVAC energy consumption for the winter day simulation and the total energy consumption of the chiller and pumps for the summer day simulation. Table 2 presents the deviation in the energy performance indicator under faulty operations for the winter and summer simulation days.

For the heating season, the HRW failure leads to the highest increase in energy consumption followed by the IAT sensor bias (-2 °C) fault. Fan failure leads to the highest decrease in energy consumption, followed by IAT sensor bias (+2 °C), fan stuck (50 %) and HCV stuck (0 %). For the cooling season, CCV stuck (50 %) and CCV leak (40 %) lead to the highest increase in energy consumption whereas fan failure and CCV stuck (0 %) lead to the highest decrease in energy consumption.

Table 2 Overview of the effects of faults on energy consumption.

Heating - baseline energy: 999.14 kWh		Cooling - baseline energy: 67.02 kWh	
Fault	Energy deviation (kWh)	Fault	Energy deviation (kWh)
HCV stuck (0 %)	-112.37	CCV stuck (0 %)	-67.02
HCV stuck (100 %)	6.75	CCV stuck (50 %)	14.80
HCV leak (0 %)	0.00	CCV leak (0 %)	-0.01
HCV leak (40 %)	-1.31	CCV leak (40 %)	14.23
Fans stuck (85 %)	-31.49	Supply fan stuck (85 %)	-0.95
Fans stuck (50 %)	-118.11	Supply fan stuck (50 %)	-15.40
SAT sensor bias (+2 °C)	-20.57	SAT sensor bias (+2 °C)	-0.94
SAT sensor bias(-2 °C)	5.71	SAT sensor bias(-2 °C)	-2.55
IAT sensor bias (+2 °C)	-131.99	IAT sensor bias (+2 °C)	-0.10
IAT sensor bias(-2 °C)	80.94	IAT sensor bias(-2 °C)	0.13
Incorrect SAT setpoint (+2 °C)	1.63	Incorrect SAT setpoint (+2 °C)	-2.73
Incorrect SAT setpoint (-2 °C)	-20.32	Incorrect SAT setpoint (-2 °C)	-0.94
HRW failure	317.03		

3.1.1 Effect of sensor bias and incorrect setpoint faults on energy exchanged at heating and cooling coils

It can be observed from Table 2 that the sensor bias faults (except for IAT sensor bias in the cooling season) and incorrect setpoint faults have different magnitudes of effects when the fault severities are in opposite directions. To understand these results further, the energies exchanged at the heating and cooling coils were also analysed for the respective sensor bias and incorrect setpoint faults. These results are presented in Table 3 for the heating season & Table 4 for the cooling season. The energies exchanged at the coils for the rest of the simulated faults are shown in Appendix D.

Table 3: Energy exchanged at the heating coil and Total HVAC energy for selected faults.

Fault	Energy exchanged at the heating coil (kWh)	Deviation (kWh)	Total HVAC energy consumption (kWh)	Deviation (kWh)
Baseline	282.04	-	999.14	-
SAT Sensor Bias (+2°C)	220.16	-61.88	978.57	-20.57
SAT Sensor Bias (-2°C)	311.15	29.11	1004.85	5.71
IAT Sensor Bias (+2°C)	256.47	-25.57	867.15	-131.99
IAT Sensor Bias (-2°C)	254.36	-27.68	1,080.08	80.94
Incorrect SAT Setpoint (+2◦ C)	295.16	13.12	1,000.77	1.63
Incorrect SAT Setpoint (-2◦ C)	220.17	-61.87	978.82	-20.32

Table 4: Energy exchanged at the cooling coils and energy consumption of chiller and pump for selected faults.

Fault	South zone	North zone	1.05 zone	Total Energy Exchanged	Deviation	Energy Consumption (Chiller &	Deviation
	(kWh)	(kWh)	(kWh)	(kWh)	(kWh)	Pump)	(kWh)
Baseline	83.16	101.09	41.94	226.19	-	67.02	-
SAT Sensor Bias (+2°C)	87.23	99.89	42.99	230.11	3.92	66.08	-0.94
SAT Sensor Bias (-2°C)	70.47	91.22	37.06	198.75	-27.44	64.47	-2.55
IAT Sensor Bias (+2°C)	82.66	100.68	41.93	225.27	-0.92	66.92	-0.10
IAT Sensor Bias (-2°C)	84.18	101.49	41.85	227.52	1.33	67.15	0.13
Incorrect SAT Setpoint (+2°C)	70.71	91.45	37.18	199.34	-26.85	64.29	-2.73
Incorrect SAT Setpoint (+2°C)	87.22	99.91	42.99	230.12	3.93	66.08	-0.94

EnergyPlus uses a control strategy that uses the model's knowledge of how much energy enters/leaves a zone as a function of the zone's temperature. It uses a 'predictive system energy balance' method first to calculate how much energy needs to be delivered by the HVAC system to maintain the required zone temperature and then updates the actual zone temperature based on how much energy is actually delivered by the HVAC system [42]. The controller for the water coil (heating/ cooling) is a solution inverter which iterates through various mass flow rates through the coil to find the required mass flow rate to meet the coil SAT [42]. The HVAC system in the case study building accomplishes its control strategy through a master-slave PI controller where the master controller adjusts the SAT setpoint in response to the measured Indoor Air Temperature (IAT), and the slave controller adjusts the opening of the HCV/CCV in response to the SAT setpoint to be met. Table 5 lists the expected end outcome regarding the energy transferred at the heating and cooling coils for these six faults in the case study building. These expected outcomes were deduced from how the master-slave controller would react to such a fault from the moment of introduction in the HVAC system. The details of the full reaction that led to these final expected outcomes can be found in Appendix E.

The results of the EnergyPlus fault simulations were compared against the expected outcome qualitatively explained in Table 5 to understand how close the simulation is to the expected reality. Comparing the expected behaviour under the master-slave control with the simulation results for the heating season faults in Table 3, it can be seen that except for the IAT sensor bias (-2°C) fault, the energy transferred at the heating coil for the remaining faults followed the expected trend in Table 5 and the direction of deviation of total HVAC energy is the same as that of the energy transferred at the heating coil. For the IAT sensor bias (-2°C) fault, the deviation of total HVAC energy is in the opposite direction of the energy transferred at the heating coil.

When the comparison was made for the simulation results of the cooling season faults in Table 4 against the expected results, it was observed that the SAT sensor bias (-2°C) and the incorrect SAT setpoint (+2°C) faults follow the expected trend and the direction of deviation of chiller and pump energy consumption is the same as that at the cooling coils. For the IAT sensor bias (+2°C) the simulated result is a negligible reduction in energy (both coils and chiller + pumps); the expected controller reaction is a negligible increase. Nevertheless, the magnitude of deviation is negligible. For the IAT sensor bias (-2°C), the energy transferred has increased very slightly at the cooling coils of the South and North zones and decreased very negligibly at the cooling coil for zone 1.05. This is against the result expected with the master–slave controller. For the incorrect SAT setpoint (-2°C) fault, the energy ex-change at the cooling coils of zones South and 1.05 follow the expected trend of a small increase whereas it is the opposite for the North zone cooling coil. There is a very slight decrease in the energy consumption of chiller and pumps. Finally, for the SAT sensor bias (+2°C) fault, the direction of deviation of energy exchanges at the cooling coils of zones South and 1.05 increase as expected whereas it decreases at the cooling coil in zone North although the magnitude of deviation and the offset effect caused by the master–slave controller cannot be estimated.

Table 5: Expected change in energy transfer at heating and cooling coils due to master–slave controller.

Fault	Master-slave controller reaction: heating season	Master-slave controller reaction: cooling season
SAT sensor bias (+2°C)	Overall, there is a decrease in energy transfer across the heating coil.	The effect of this fault should be (partly) compensated by the master–slave controller. Overall, there is an increase in energy transfer across the cooling coils.
SAT sensor bias (-2°C)	The effect of this fault should be (partly) compensated by the master–slave controller. Overall, there is an increase in energy transfer at the heating coil.	Overall, there will be a decrease in cooling energy.
IAT sensor bias (+2°C)	Overall, there will be a decrease in energy transferred at the heating coil.	There is expected to be a small (negligible) increase in energy transfer at the cooling coils.
IAT sensor bias (-2°C)	There is expected to be a negligible change (increase) in energy transfer at the heating coil.	Overall, the energy transferred across the cooling coils will decrease and the opening of the cooling coils will depend on how much the IAT value increases
Incorrect SAT setpoint (+2°C)	Overall, there is expected to be a minimal change (increase) in energy transferred at the heating coil if any.	Overall, there is expected to be a decrease in energy transfer at the heating coil.
Incorrect SAT setpoint (-2°C)	Overall, there is expected to be a decrease in energy transfer across the cooling coils.	Overall, there is expected to be a negligible/small increase in energy transfer at the cooling coils.

3.1.2 Energy performance result for impact calculation

As mentioned in section 2.3, the absolute value of the deviation in energy consumption is chosen to represent the effect on energy performance. Fig. 9 presents the absolute deviation in total HVAC energy consumption in decreasing order for the heating season faults. HRW failure and fan failure have the highest effect on energy performance, followed by IAT sensor bias (+2°C), fan stuck (50 %), HCV stuck (0 %) and IAT sensor bias (-2°C).

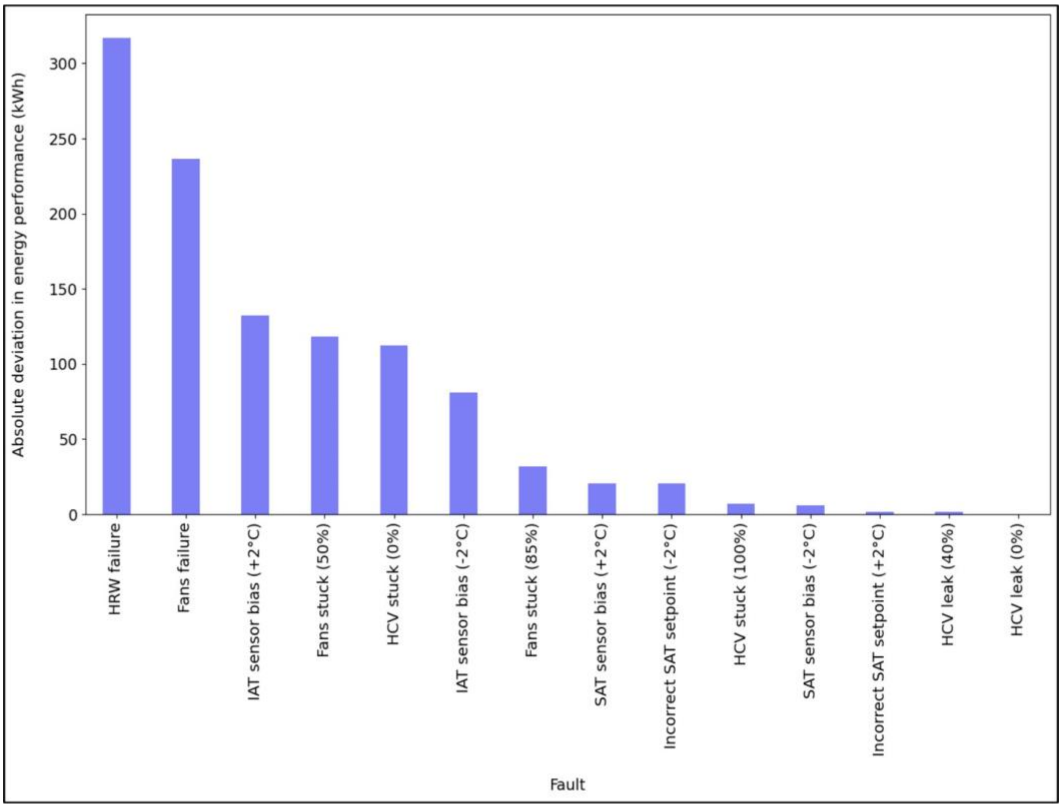

Fig. 9. Effect of heating season faults on the energy performance of the case study building.

Similarly, Fig. 10 presents the absolute deviation in energy consumption of the chiller and pumps in decreasing order for the cooling season faults. The CCV stuck (0 %) fault and fan failure have the joint highest effect on energy performance, followed by fan stuck (50 %), CCV stuck (50 %), and CCV leak (40 %).

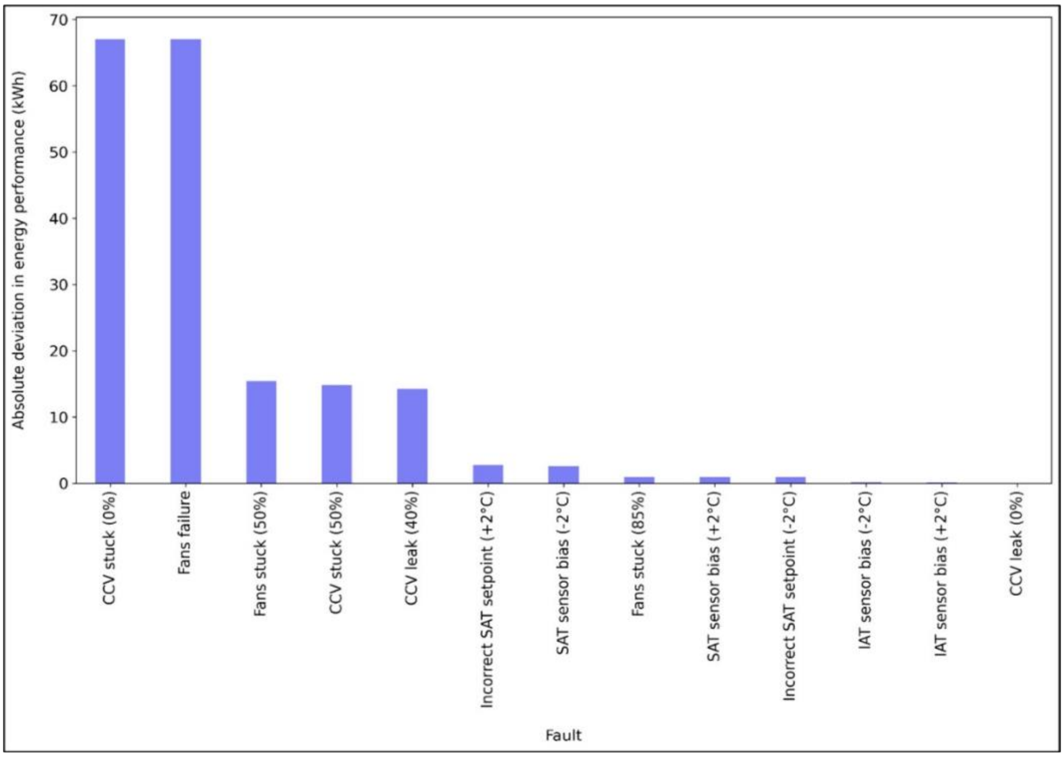

Fig. 10. Effect of cooling season faults on the energy performance of the case study building.

3.2 Frequency of fault occurrence from text mining

Fig. 11 presents the results of the fault frequency values obtained through the network clustering feature. The figure presents the results for the two different analyses: clustering without applying the bag of words feature and clustering after applying the bag of words feature. The results are presented as the total count of relevant word co-occurrences obtained across all 11,255 work orders examined. Fan failure has the highest frequency of occurrence and its value is much higher than the fault with the second highest frequency. The CCV stuck fault has the least frequency of occurrence. For the HCV stuck fault, the obtained count is the same for both the analysed cases. For six faults (fan failure, incorrect SAT setpoint high, HCV leak, HRW failure, CCV leak and CCV stuck), the count obtained without applying the bag of words feature is higher than the count obtained after applying the bag of words feature (although for the CCV leak fault, the difference is only 1 and for the CCV stuck fault, the difference is only 8). However, it is interesting to note that the value of the total count obtained through network clustering after applying the bag of words feature is higher than the value obtained without using the bag of words feature for four faults (incorrect SAT setpoint low, SAT sensor bias, fan stuck, and IAT sensor bias).

As mentioned in section 2.4, for calculating the total fault impact, the results obtained after applying the bag of words feature were used. In this way, the results of the word co-occurrence count consider only one occurrence of each keyword per work order and can also be interpreted as the possible number of times the faults occurred over the analysed 5 years. When this value is divided by 5, the annual occurrence rate of each fault can be obtained and used for impact calculation.

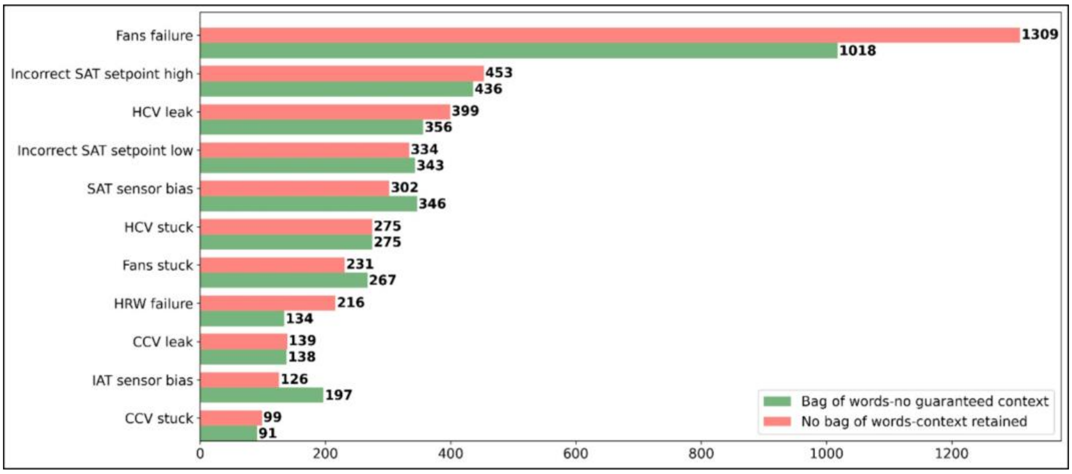

Fig. 11. Results for fault frequency occurrence obtained through network clustering for both cases (with and without applying the bag of words feature).

3.3 Total energy impact and fault prioritisation

In this section, the total energy impact values of the AHU faults obtained using Eq. (4), are presented. Fig. 12 shows the total energy impact of the heating season faults along with the Pareto curve (referred to as 'Cumulative Impact (%)' in the figure). The fan failure fault has the highest impact, followed by HRW failure, fan stuck (50 %), and HCV stuck (0 %). These four faults comprise the list of prioritised faults for the heating season when the Pareto principle is applied. Fig. 13 presents the total energy impact of the cooling season faults along with the Pareto curve. The fan failure fault has the highest total energy impact on the case study building and solely constitutes the list of prioritised faults for the cooling season.

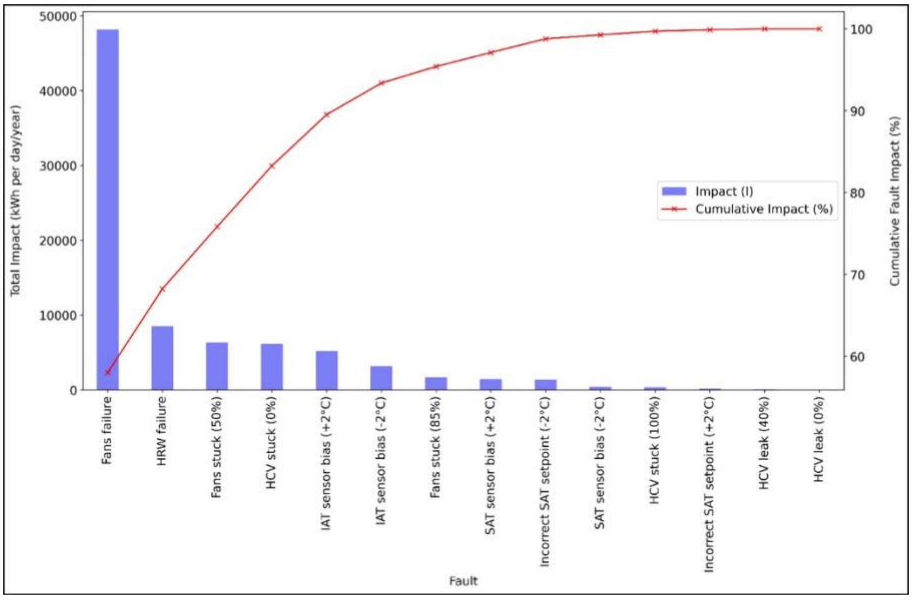

Fig. 12. Total energy impact values and Pareto curve for the heating season faults.

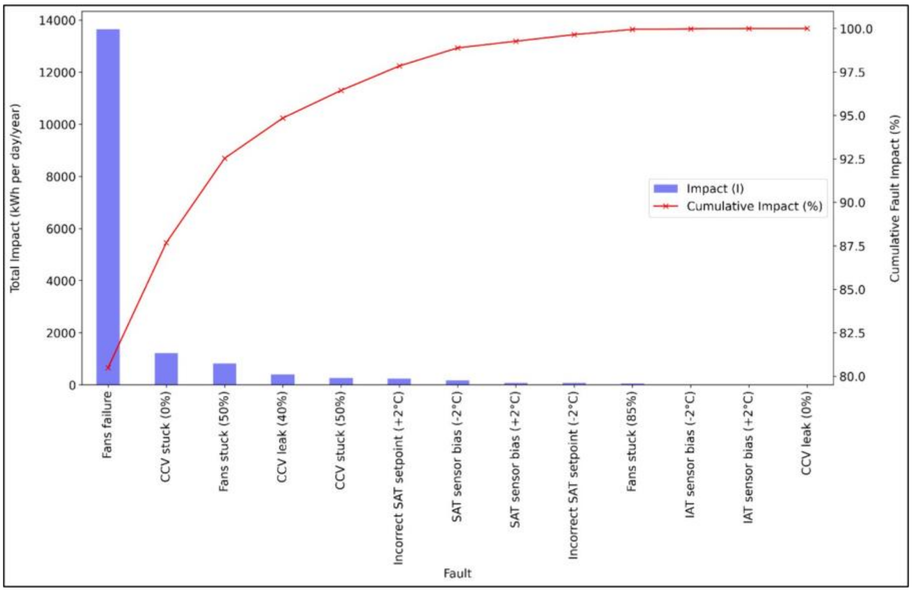

Fig. 13. Total energy impact values and Pareto curve for the cooling season faults.

3.4 List of BMS alerts

A list of existing BMS alerts related to the studied AHU components and faults was obtained to assess what faults can be detected using already available alerts in the BMS. The availability of these alerts was classified as either 'often' or 'occasionally' depending on how often these alerts were observed in practice. The broken fan drive alert (based on motor failure, digital command alert due to incorrect fan operational status returned compared to the sent command signal, delta pressure transmitter) and HRW failure alert (based on motor failure & digital command alert due to incorrect motor operational status returned compared to the sent command signal) are often available. Similarly, alerts are often available for situations when sensor values of SAT, supply water temperature to coils, IAT, return air temperature and supply air pressure exceed the setpoint limits for a specific period. However, these alerts do not indicate if a sensor has a bias/offset or if the setpoint itself is wrongly configured. So, these alerts may not help directly to detect the sensor offset and setpoint faults that were simulated in this study. For some components like the CCV and the HCV, their end positions (0 % & 100 % open) are monitored using a digital command alert based on an electrical contact. If the valves are stuck at either 0 % or 100 %, then the available alerts can be used to diagnose the presence of the stuck valve faults. However, there is no alert available for the valves being stuck in between 0 % & 100 % open. Finally, the fan stuck fault and valve leak fault do not have an alert.

4 DISCUSSION

The main objective of this research was to combine the actual frequency of fault occurrence with the simulated effect on energy performance into a total fault impact analysis framework. The Pareto principle was employed to prioritise the faults based on their total energy impact. Analyzing the list of existing alerts helped to understand how often alerts are available and can be triggered when the studied AHU faults (at the studied severity levels) are diagnosed.

4.1 Reflection on results

In this sub-section, a discussion of the obtained results for energy performance, fault frequency, total energy impact and BMS alerts is performed.

4.1.1 Effect on energy performance

The KPI used to study the effect of summer faults on energy performance was the sum of chiller and pump energy consumption. The fan and HRW energy consumption of the cooling season model were not included in the KPI because it does not reflect the actual energy consumption of those components in the building. As a result, the effect on energy performance for the fan failure fault was the same as that for the CCV stuck at 0 %. When the accurate HVAC system energy value is considered for the KPI, the fan failure fault will have a much higher effect on energy performance than the CCV stuck (0 %) fault. The following brief calculation analyses what this additional energy performance deviation would be for fan failure. For the summer season, the supply fan power is around 2.8 kW when operational. Considering the return AHU fan too, the total fan power is 5.6 kW. The HRW power is 3.7 kW. When the AHU is operational for 10 h a day, the energy consumed by these two fans and the HRW put together is 93 kWh. Therefore, the deviation of total HVAC energy is 160 kWh. In the case of the fan stuck fault, the fan power consumption is reduced by 50 % and there is no change in the power consumption of HRW. So, the absolute energy deviation from the baseline of total HVAC energy consumption for the fan stuck (50

4.1.1.1. Effect of night ventilation strategy.

The simulated model did not contain a night ventilation strategy (implemented between 23:00 and 07:00 in the case study building) for the cooling season. If night ventilation were in effect, the maximum additional consumption of the HVAC system is (roughly) 45 kWh if the night ventilation is switched on throughout the night. So, the absolute energy deviation of total HVAC energy consumption for fan failure could vary between 160 kWh and 205 kWh, depending on the duration of the night ventilation operation. Similarly, for the fan stuck fault, the absolute energy deviation varies between 43.5 kWh and 66 kWh, depending on the duration of the night ventilation operation.

Since the operation of night ventilation depends on the IAT value, the IAT sensor bias also affects the operation of night ventilation. A positive IAT sensor bias could lead to a maximum effect of night ventilation being switched on throughout the night and vice-versa for a negative bias (roughly 45 kWh absolute energy deviation for both cases). Consequently, this will also affect the chiller energy consumption during the day which cannot be quantified without a simulation of a model with the night ventilation strategy.

4.1.2 Frequency of fault occurrence

The frequency of fault occurrence has played an important role in determining the impact of faults. For example, in prioritising cooling season faults, the CCV stuck (0 %) fault and the fan failure fault contribute equally to the effect on the studied energy performance KPI. However, the fan failure fault occurs most frequently among the studied faults in the cooling season. This compounds the impact of the fault when both the energy effect and frequency are considered together (more than 80 % of the cumulative total energy impact of all faults). In contrast, the CCV stuck (0 %) has the least frequency of occurrence, and consequently, it contributes to just above 7 % of the cumulative total energy impact of all faults. In prioritising heating season faults, the HRW failure has the highest effect on energy performance but occurs with the least frequency. Therefore, the frequency value tempered the total energy impact the fault had on the building (just over 10 % of the cumulative total energy impact of all faults).

The values obtained for the fault frequency indicated that for four faults, the frequency obtained after applying the bag of words feature was higher than the frequency obtained without applying the bag of words feature. This observation is contrary to the initial expectation. One possible reason could be that in the case using the bag of words feature, the window of search for word co-occurrence was expanded to include the whole work order thereby returning relevant results even if the two words were found outside the window size of 7.

However, the specific reason for this result and all possible circumstances that would lead to this result were not investigated as it requires further study is required outside the scope of this research.

4.1.3 Total energy impact and fault prioritisation

The results obtained for the cooling season indicated that fan failure alone constitutes the list of prioritised faults. If the total HVAC energy were to be considered (as discussed in section 4.1.1), the fan failure would still be the sole prioritised fault both when the night ventilation operation is considered and not considered. While the HRW failure was not simulated for the cooling season, it had the highest effect on energy performance for the heating season. Therefore, it is recommended to consider the HRW failure as well when developing AFDD tools for cooling season faults.

4.1.4 BMS alerts

A comparison of the faults prioritised in this research and the BMS alerts available in the market indicated that except for the fan stuck fault (heating season), an alert is available for the rest of the prioritised faults for the studied severity settings. An alert must be developed for the fan stuck fault. Additionally, when considering the rest of the faults, the stuck valve fault (for a value between 0 % and 100 % open), the leaky valve fault, the SAT sensor bias fault, the IAT sensor bias fault and the incorrect SAT setpoint (low & high) faults require the development of alerts to be raised when diagnosed by AFDD tools.

4.2 Comparison of results with literature

The results obtained in this study were compared with the results found in the reviewed literature. Across multiple studies, the faults relating to airflow such as outside air damper fully open/stuck [8,9,13,15,16], VAV box damper fully open/stuck [9,11,23], fan stuck at full speed [15] had the highest effect on energy consumption. In some cases, the CCV/HCV faults [8] and sensor bias faults [8,11–13,15] also had a considerable/comparable effect. In line with the literature, the results obtained from this study also indicate that fan failure, fan stuck, and valve (CCV/HCV) stuck faults have a significant effect on energy performance. The study of HRW failure is an addition to the list of faults studied in the literature and purely in terms of effect on energy performance, it has a higher effect than the other faults during the heating season. When considering the total energy impact (after incorporating fault occurrence frequency), fan failure, fan stuck, HCV stuck and HRW failure constitute the prioritised faults (when the Pareto principle is applied).

4.3 Contributions, limitations and future research

This research has made the following contributions to the scientific literature on simulation-based fault impact research:

- Incorporation of real fault occurrence frequency: To the best of the authors' knowledge, this paper represents the first research which has combined real fault occurrence frequency data from maintenance logbooks with a simulation-based fault impact analysis. This helps to quantify the total energy impact of faults.
- Prioritisation of a small set of faults: The Pareto principle has been used to prioritise faults based on their total energy impact. This provides a cutoff criterion to select a small number of priority faults that have a very high impact on further focus through AFDD tools.
- Study of existing BMS alerts: The prioritised faults have been compared with a list of existing BMS alerts to verify if sufficient alerts are present to notify maintenance teams about the diagnosis of those faults using an AFDD tool. This represents an important step in bridging the gap between research and practical implementation.

Based on the results of this study, the next step is to develop an AFDD tool to detect and diagnose the studied faults, with a particular focus on the prioritised faults. The development of such a tool will expedite condition-dependent maintenance.

While the study has made significant contributions to the literature of fault impact analysis study, there are some limitations which must be addressed in subsequent studies.

- The scope of this study was to prioritise AHU faults based on energy and frequency. However, faults can also affect the thermal comfort of occupants. Therefore, the inclusion of thermal comfort in this framework is a recommended future study.
- For the word co-occurrence analysis, the hyper-parameter 'window size' was chosen based on research that focused on text written in grammatically correct English whereas the maintenance work orders were not only in Dutch but also need not necessarily have contained grammatically correct and full sentences.

To address this, a study on how to adapt the network clustering analysis to maintenance orders must be performed.

- The fault frequency was mined from free-text fields in the maintenance orders. Different members of the maintenance team could describe the same maintenance problem differently. Additionally, the lexicon of keywords was deduced from Dutch translations of how the faults are named in the scientific literature in English. This could have resulted in some relevant results not being considered. The lack of standardised fault names could later hamper a meta-analysis of faults reported by FDD tools Chen et al. [43]. Therefore, it is recommended to first standardise the names of faults and subsequently incorporate fields with drop-down lists with the standardised names in maintenance orders so that all relevant results can be obtained through text mining.
- The fault frequency was studied for faults pertaining to limited components of AHU. Extending the list of faults to not just more AHU components (e.g.: filters, humidistat, frost thermostat, dampers), but also to the entire HVAC system will provide crucial data about frequently occurring faults across generation, distribution, and transmission systems.
- While the Pareto principle helped to prioritise faults, a key limitation of this method is that the faults that form a part of the prioritised list (through 80 % cumulative total energy impact) are subject to the number of faults simulated and their respective severities. One way to overcome this is to incorporate the probabilities of occurrence for the range of all plausible severities of all the studied faults, as performed by Li and O'Neill [9], so that the total of all energy deviations due to all studied faults (based on which the 80 % threshold is determined) is fixed. Thus, it is recommended to identify and include probabilities of occurrence of various severities for the simulated faults in future research.

5 CONCLUSIONS

A fault impact analysis study was performed to prioritise acute AHU faults concerning CCV, HCV, fans, HRW, sensor bias and incorrect set-point. A building energy simulation model of a case study was developed using DesignBuilder and the studied faults were introduced and simulated using EnergyPlus. Fault occurrence frequency values were obtained from maintenance records over 5 years using text mining. The impact of the fault was defined as the product of the effect on energy performance and fault frequency. The Pareto principle was applied to prioritise the most important faults based on their impacts. A list of available BMS alerts was obtained and cross-verified with the prioritised list of faults. The main conclusions are as follows:

- Fan failure, HRW failure, Fan stuck (50 %) and HCV stuck (0 %) are the prioritised heating season faults.
- Fan failure is the prioritised cooling season fault.
- From the list of prioritised faults, only the fan stuck at 50 % fault for the heating season doesn't have any alerts in the BMS; the rest of the specific fault severity settings of the prioritised faults have an alert available in the BMS used in practice.

APPENDIX A: CASE STUDY BUILDING MODEL CONSTRUCTION INFORMATION

Table A1 Ground and internal floor.

Material	Thickness (m)	Thermal conductivity (W/m-K)	Density (kg/m3)	Specific heat capacity (J/kg-K)
Urea formaldehyde foam	0.0916	0.04	10	1,400

Table A2: External floor.

Material	Thickness (m)	Thermal conductivity (W/m-K)	Density (kg/m3)	Specific heat capacity (J/kg-K)
Outermost layer	0.025	0.5	1300	1000
MW stone wool (rolls)	0.1482	0.04	30	840
Timber flooring (innermost)	0.005	0.14	650	1200

Table A3 Roof.

Material	Thickness (m)	Thermal conductivity (W/m-K)	Density (kg/m3)	Specific heat capacity (J/kg-K)
Board insulation (Glass fibre board)	0.0829	0.036	160	840
Metal deck	0.01	45.28	7,824	500

Table A4: External wall.

Material	Thickness (m)	Thermal conductivity (W/m-K)	Density (kg/m3)	Specific heat capacity (J/kg-K)
Brickwork outer	0.105	0.84	1,700	800
MW stone wool	0.05	0.038	40	840
Concrete block (lightweight)	0.15	0.19	60	1,000
Plaster (lightweight)	0.013	0.16	600	1,000

Table A5: Internal walls.

Material	Thickness (m)	Thermal conductivity (W/m-K)	Density (kg/m3)	Specific heat capacity (J/kg-K)
Gypsum plasterboard	0.025	0.25	900	1,000
Air gap 10 mm Gypsum plasterboard	0.025	0.25	900	1,000

Table A6: Windows.

U-Factor (W/m2-K)	Solar Heat Gain Coefficient	Visible Transmittance
3.2	0.691	0.744

APPENDIX B: HVAC OPERATION SETPOINTS OF THE SIMULATION MODEL

Winter Day Operation:

- Air flow rate: 2.73 m^3/s.
- AHU maximum SAT (at the location after supply fan): 26 °C. North zone indoor temperature setpoint: 22.5 °C.
- South zone indoor temperature setpoint: 22 °C.
- 1.05 zone indoor temperature setpoint: 20.5 °C.

Summer Day Operation:

- Air flow rate: 3 m^3/s.
- Minimum SAT (at the location after cooling coils) at outdoor temperatures of 20 °C and 30 °C respectively: North: 18 °C and 17 °C.
- South: 18 °C and 17 °C.
- 1.05: 18 °C and 16 °C.
- Zone indoor temperature setpoint (all 3 zones): 23.5 °C.

APPENDIX C: LEXICON USED IN FAULT FREQUENCY ANALYSIS

Table C1 Lexicon used to represent the studied AHU faults.

Keyword in Dutch	English translation
regelafsluiter	control valve
hoog	high
regeling	rule/regulation/control
laag	low
cv	central heating
klep	valve
druk	pressure
toevoerkanaal	supply channel
toevoervent	supply channel
afzuigvent	exhaust channel
aanzuigkanaal	supply channel
setpoint	setpoint
opnemers	sensors
ventilator	fan
toevoerventilator	supply fan
snaarbreuk	belt broken
ruimtetemperatuur	room temperature
verwarmer	heater
afzuigventilator	return fan
sensor	sensor
sensoren	sensors
fout	fault
lekkage	leakage
warmtewiel	heat recovery wheel
inblaastemperatuur	supply air temperature
gewenste	desired
regelen	rules
storing	failure
verwarming	heating
opnemer	sensor
koeling	cooling
metingen	measurements
kapot	broken
afsluiter	valve
ventilatie	ventilation
gewenst	desired
koeler	cooler
thermostaat	thermostat
vast	stuck

Keyword in Dutch	English translation
snaar	belt
hoge	high
lek	leak
snaren	belts
lage	low
meting	measurement
afwijking	deviation
afwijkingen	deviations
verkeerd	incorrect
verkeerde	incorrect
offset	offset

APPENDIX D: EFFECT OF FAULTS ON ENERGY EXCHANGED AT THE HEATING AND COOLING COILS

Table D1: Energy exchanged at the heating coil and Total HVAC consumption for selected faults.

Fault	Energy exchanged at the heating coil (kWh)	Deviation (kWh)	Total HVAC energy consumption (kWh)	Deviation (kWh)
Baseline	282.04	-	999.14	-
HCV Stuck (0 %)	0.00	-282.04	886.77	-112.37
HCV Stuck (100 %)	311.15	29.11	1005.89	6.75
HCV Leak (0 %)	282.04	0.00	999.14	0.00
HCV Leak (40 %)	282.04	0.00	997.83	-1.31
Fan Failure	0.00	-282.04	762.67	-236.47
Fan Stuck (85 %)	241.06	-40.98	967.65	-31.49
Fan Stuck (50 %)	144.03	-138.01	881.03	-118.11
HRW Failure	418.43	136.39	1316.17	317.03

Table D2: Energy exchanged at cooling coils and energy consumption of chiller and pump for selected faults.

Fault	South zone	North zone	1.05 zone	Total Energy Exchanged	Deviation	Energy Consumption (Chiller &	Deviation
	(kWh)	(kWh)	(kWh)	(kWh)	(kWh)	Pump)	(kWh)
Baseline	83.16	101.09	41.94	226.19	-	67.02	-
CCV Stuck (0%)	0.00	0.00	0.00	0.00	-226.19	0.00	-67.02
CCV Stuck (50%)	86.64	100.56	43.54	230.74	4.55	81.82	14.80
CCV Leak (0 %)	83.16	101.09	41.94	226.19	0.00	67.01	-0.01
CCV Leak (40%)	85.86	101.19	42.93	229.98	3.79	81.25	14.23
Fan Failure	0.00	0.00	0.00	0.00	-226.19	0.00	-67.02
Fan Stuck (85%)	75.80	97.11	39.52	212.43	-13.76	66.07	-0.95
Fan Stuck (50%)	50.60	64.80	26.24	141.64	-84.55	51.62	-15.40

APPENDIX E: REACTION OF MASTER–SLAVE CONTROLLERS TO SENSOR BIAS AND INCORRECT SETPOINT FAULTS

Table E1 The reaction of master–slave controllers to sensor bias and incorrect SAT setpoint faults during the heating season.

Fault	Master-slave controller reaction: heating season
SAT Sensor Bias (+2°C)	Bias increases SAT value → Slave controller reduces HCV opening to reduce SAT → Energy transfer at heating coil decreases → When IAT decreases, master controller will increase SAT setpoint → If maximum desired SAT is reached, HCV won't be opened further Overall, there is a decrease in energy transfer across the heating coil.
SAT Sensor Bias (-2°C)	Bias decreases SAT value → Slave controller increases HCV opening to increase SAT → When IAT increases beyond setpoint, master controller decreases SAT setpoint → HCV opening reduces The effect of this fault should be (partly) compensated by the master-slave controller. Overall, energy transfer at the heating coil increases.
IAT Sensor Bias (+2°C)	Bias increases IAT value → Master controller reduces SAT setpoint → Slave controller reduces HCV opening → Due to its bias, IAT value perceived to be above setpoint → No further change in HCV opening & SAT setpoint unless needed to maintain IAT setpoint Overall, there will be a decrease in energy transferred at the heating coil.
IAT Sensor Bias (-2°C)	Bias decreases IAT value → Master controller increases SAT setpoint → Slave controller increases HCV opening → If SAT setpoint already at maximum desired SAT, HCV opens further only to maintain this SAT Overall, it is expected that there will be a negligible change (increase) in energy transfer at the heating coil.
Incorrect SAT Setpoint (+2°C)	Maximum desired SAT value increased → SAT setpoint signalled by master controller will only be slightly higher (if needed) → Slave controller increases HCV opening only slightly higher (as needed) Overall, it is expected that there will only be a minimal change (increase) in energy transferred at the heating coil if any.
Incorrect SAT Setpoint (-2°C)	Maximum desired SAT value decreased → SAT setpoint signalled by master controller cannot exceed maximum desired SAT → Slave controller responds with a decreased HCV opening → No further change in HCV opening unless needed to maintain SAT setpoint Overall, it is expected that there will be a decrease in energy transfer at the heating coil.

Table E2: he reaction of master–slave controllers to sensor bias and incorrect SAT setpoint faults during the cooling season.

Fault	Master-slave controller reaction: cooling season
SAT Sensor Bias (+2°C)	Bias decreases SAT value → Slave controller reduces CCV opening to increase SAT → Energy transfer at cooling coil decreases → When IAT increases, master controller will decrease SAT setpoint → If minimum desired SAT is reached, CCV won't be opened further
	The effect of this fault should be (partly) compensated by the master–slave controller. Overall, energy transfer at the heating coil increases.
SAT Sensor Bias (+2°C)	Bias decreases SAT value → Slave controller reduces CCV opening to increase SAT → Energy transfer at cooling coil decreases → When IAT increases, master controller will decrease SAT setpoint → If minimum desired SAT is reached, CCV won't be opened further
	Overall, there will be a decrease in energy transfer at the cooling coil.
IAT Sensor Bias (+2°C)	Bias increases IAT value → Master controller decreases SAT setpoint → Slave controller increases CCV opening → If SAT setpoint already at minimum desired SAT, CCV opens further only to maintain this SAT
	Overall, the energy transferred across the cooling coils will decrease and if the cooling coils will open will depend on how much the IAT value increases.
Incorrect SAT Setpoint (+2°C)	Minimum desired SAT values increased → Master controller increases SAT setpoint → Slave controller reduces CCV opening → No further change in CCV opening unless needed to maintain SAT setpoint
	Overall, it is expected that there will be a decrease in energy transfer across the cooling coils
Incorrect SAT Setpoint (-2°C)	Minimum desired SAT values decreased → SAT setpoint signalled by master controller will only be slightly lower (if needed) → Slave controller increases CCV opening only slightly higher (as needed)
	Overall, it is expected that there will be a negligible/small increase in energy transfer at the cooling coils.

APPENDIX F. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2024.114476.

REFERENCES

[1] United Nations Environment Programme, ‘2021 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION Towards a zero-emissions, efficient and resilient buildings and construction sector’, 2021. [Online]. Available: <www.glob

alabc.org>.

[2] J. Granderson, G. Lin, R. Singla, E. Mayhorn, P. Ehrlich, D. Vrabie, ‘Commercial Fault Detection and Diagnostics Tools: What They Offer, How They Differ, and What’s Still Needed’, 2018, Accessed: Jul. 07, 2024. [Online]. Available: <htt

ps://doi.org/10.20357/B7V88H>.

[3] RvO Nederland, ‘Monitor Verduurzaming Gebouwde Omgeving 2023’, 2023. Accessed: Jun. 03, 2024. [Online]. Available: <https://www.rvo.nl/sites/defaul t/files/2023-12/monitor-verduurzaming-gebouwde-omgeving-2023.pdf>.

[4] CBS, ‘Renewable electricity; production and capacity’. Accessed: Jun. 03, 2024. [Online]. Available: <https://www.cbs.nl/en-gb/figures/detail/82610ENG>.

[5] Y. Zhao, T. Li, X. Zhang, C. Zhang, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future, Renew. Sustain. Energy Rev. (2019), https://doi.org/10.1016/j. rser.2019.04.021.

[6] W. Kim, S. Katipamula, A review of fault detection and diagnostics methods for building systems, Science and Technology for the Built Environment 24 (1) (2018) 3–21, https://doi.org/10.1080/23744731.2017.1318008.

[7] Z. Chen, Z. O’Neill, J. Wen, O. Pradhan, T. Yang, X. Lu, G. Lin, S. Miyata, S. Lee,

C. Shen, R. Chiosa, M.S. Piscitelli, A. Capozzoli, F. Hengel, A. Kührer, M. Pritoni,

W. Liu, J. Clauß, Y. Chen, T. Herr, A review of data-driven fault detection and diagnostics for building HVAC systems, Appl. Energy 339 (2023), https://doi.org/ 10.1016/j.apenergy.2023.121030.

[8] X. Lu, Y. Fu, Z. O’Neill, J. Wen, A holistic fault impact analysis of the high- performance sequences of operation for HVAC systems: Modelica-based case study in a medium-office building, Energ. Buildings 252 (2021), https://doi.org/ 10.1016/j.enbuild.2021.111448.

[9] Y. Li, Z. O’Neill, An innovative fault impact analysis framework for enhancing building operations, Energ. Buildings 199 (2019) 311–331, https://doi.org/ 10.1016/j.enbuild.2019.07.011.

[10] G. Li, Q. Yao, Y. Hu, X. Fang, L. Wang, Investigating thermostat sensor offset impacts on operating performance and thermal comfort of three different HVAC systems in Wuhan, China. Case Studies in Thermal Engineering 31 (2022), https:// doi.org/10.1016/j.csite.2022.101788.

[11] S.H. Lee, F.W.H. Yik, A study on the energy penalty of various air-side system faults in buildings, Energ. Buildings 42 (1) (2010) 2–10, https://doi.org/10.1016/j. enbuild.2009.07.004.

[12] M. Basarkar, X. Pang, L. Wang, P. Haves, T. Hong, ‘Modeling and simulation of HVAC faults in EnergyPlus’, Sydney, Nov. 2011. https://doi.org/10.26868

/25222708.2011.1925.

[13] L. Wang, T. Hong, ‘Modeling and Simulation of HVAC Faulty Operations and Performance Degradation due to Maintenance Issues’, Berkeley, Jan. 2013.

[14] R. Zhang, T. Hong, Modeling of HVAC operational faults in building performance simulation, Appl. Energy 202 (2017) 178–188, https://doi.org/10.1016/j. apenergy.2017.05.153.

[15] F. Zhong, J.K. Calautit, Y. Wu, Assessment of HVAC system operational fault impacts and multiple faults interactions under climate change, Energy 258 (2022), https://doi.org/10.1016/j.energy.2022.124762.

[16] R. Khire, M. Trcka, ‘Model based failure mode effect analysis on whole building energy performance’, Chambe´ry, Aug. 2013.

[17] S. Ginestet, D. Marchio, O. Morisot, Evaluation of faults impacts on energy consumption and indoor air quality on an air handling unit, Energ. Buildings 40 (1) (2008) 51–57, https://doi.org/10.1016/j.enbuild.2007.01.012.

[18] X. Lu, Z. O’Neill, Y. Li, F. Niu, A novel simulation-based framework for sensor error impact analysis in smart building systems: A case study for a demand-controlled ventilation system, Appl. Energy 263 (2020), https://doi.org/10.1016/j. apenergy.2020.114638.

[19] S. Huang, W. Zuo, D. Vrabie, R. Xu, Modelica-based system modeling for studying control-related faults in chiller plants and boiler plants serving large office buildings, Journal of Building Engineering 44 (2021), https://doi.org/10.1016/j. jobe.2021.102654.

[20] J. Verhelst, G.V. Ham, D. Saelens, L. Helsen, Economic impact of persistent sensor and actuator faults in concrete core activated office buildings, Energ. Buildings 142 (2017) 111–127, https://doi.org/10.1016/j.enbuild.2017.02.052.

[21] Y. Li, Z. O'Neill, A critical review of fault modeling of HVAC systems in buildings, Build. Simul. (2018), https://doi.org/10.1007/s12273-018-0458-4.

[22] M. Deru et al., 'U.S. Department of Energy Commercial Reference Building Models of the National Building Stock', Golden, CO (United States), Feb. 2011. https://doi. org/10.2172/1009264.

[23] K. Otto, B. Eisenhower, Z. O'Neill, S. Yuan, I. Mezic, S. Narayanan, 'Prioritizing Building System Energy Failure Modes Using Whole Building Energy Simulation',

Madison, Aug. 2012. [Online]. Available: <https://www.researchgate.net/pu blication/236649496>.

[24] A. Shamshiri, K.R. Ryu, J.Y. Park, Text mining and natural language processing in construction, Autom. Constr. 158 (Feb. 2024) 105200, https://doi.org/10.1016/j. autcon.2023.105200.

[25] H.B. Gunay, W. Shen, C. Yang, Text-mining building maintenance work orders for component fault frequency, Build. Res. Inf. 47 (5) (2019) 518–533, https://doi. org/10.1080/09613218.2018.1459004.

[26] R. Bortolini, N. Forcada, Analysis of building maintenance requests using a text mining approach: building services evaluation, Build. Res. Inf. 48 (2) (2020) 207–217, https://doi.org/10.1080/09613218.2019.1609291.

[27] M. Marocco, I. Garofolo, Operational text-mining methods for enhancing building maintenance management, Build. Res. Inf. 49 (8) (2021) 893–911, https://doi.org/ 10.1080/09613218.2021.1953368.

[28] M. D'Orazio, E. Di Giuseppe, G. Bernardini, Automatic detection of maintenance requests: Comparison of Human Manual Annotation and Sentiment Analysis techniques, Autom. Constr. 134 (2022), https://doi.org/10.1016/j. autcon.2021.104068.

[29] S. Dutta, H.B. Gunay, S. Bucking, Benchmarking operational performance of buildings by text mining tenant surveys, Science and Technology for the Built Environment 27 (6) (2021) 741–755, https://doi.org/10.1080/

23744731.2020.1851545.

[30] P. Nojedehi, W. O'Brien, H.B. Gunay, Benchmarking and visualization of building portfolios by applying text analytics to maintenance work order logs, Science and Technology for the Built Environment 27 (6) (2021) 756–775, https://doi.org/ 10.1080/23744731.2021.1913957.

[31] Y. Yu, D. Woradechjumroen, D. Yu, A review of fault detection and diagnosis methodologies on air-handling units, Energ. Buildings (2014), https://doi.org/ 10.1016/j.enbuild.2014.06.042.

[32] K. Bruton, P. Raftery, B. Kennedy, M.M. Keane, D.T.J. O'Sullivan, Review of automated fault detection and diagnostic tools in air handling units, Energ. Effi. (2014), https://doi.org/10.1007/s12053-013-9238-2.

[33] S. Chitkara, 'Continuous Monitoring and Automated Fault Detection and Diagnosis of Large Air-Handling Units', Eindhoven, May 2022. Accessed: Jul. 07, 2024. [Online]. Available: <https://research.tue.nl/en/publications/continuous-monit

oring-and-automated-fault-detection-and-diagnosis>.

[34] A. Thamban, 'Fault detection and diagnosis of the low ΔT syndrome in cooling coils of chilled water systems', Eindhoven, Aug. 2022. Accessed: Jul. 07, 2024. [Online]. Available: <https://research.tue.nl/en/publications/fault-detection-and-diagnosi

s-of-the-low-t-syndrome-in-cooling-co>.

[35] M.J. Gestwick, J.A. Love, Trial application of ASHRAE 1051-RP: Calibration method for building energy simulation, J. Build. Perform. Simul. 7 (5) (2014) 346–359, https://doi.org/10.1080/19401493.2013.838698.

[36] ASHRAE, 'ASHRAE Guideline 14-2014, measurement of Energy, Demand, and Water Savings', Atlanta, Dec. 2014. [Online]. Available: <www.ashrae.org>.

[37] B. Gunay, W. Shen, C. Yang, Characterization of a Building's operation using automation data: A review and case study, Build. Environ. (2017), https://doi.org/ 10.1016/j.buildenv.2017.03.035.

[38] Orange Data Mining, 'Corpus to Network'.

[39] J.P. Levy, J.A. Bullinaria, M. Patel, Explorations in the derivation of word co- occurrence statistics, S. Pac. J. Psychol. 10 (1) (1999) 99–111.

[40] Orange Data Mining, 'Bag of Words'.

[41] Juran, 'Pareto Principle (80/20 Rule) & Pareto Analysis Guide'. Accessed: Jun. 03, 2024. [Online]. Available: <https://www.juran.com/blog/a-guide-to-the-pareto- principle-80-20-rule-pareto-analysis/>.

[42] U.S. Department of Energy, 'EnergyPlusTM Version 22.1.0 Documentation Engineering Reference', 2022.

[43] Y. Chen, E. Crowe, G. Lin, 'Lawrence Berkeley National Laboratory LBL Publications Title What's in a Name? Developing a Standardized Taxonomy for HVAC System Faults Publication Date', 2024. [Online]. Available: <https://escho larship.org/uc/item/351568bv>.