Energy Conversion and Management 76 (2013) 155–161

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

A case study and critical assessment in calculating power usage effectiveness for a data centre

Gemma A. Brady [a,*], Nikil Kapur [b], Jonathan L. Summers [b], Harvey M. Thompson [b]

[a] Energy Research Institute (ERI), School of Process, Environmental and Materials Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom

[b] Institute of Engineering Thermofluids, Surfaces & Interfaces (iETSI), School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

Article history:
Received 17 April 2013
Accepted 20 July 2013

Keywords:
Data center
Power
Efficiency
Metric
PUE

ABSTRACT

Metrics commonly used to assess the energy efficiency of data centres are analysed through performing and critiquing a case study calculation of energy efficiency. Specifically, the metric Power Usage Effectiveness (PUE), which has become a de facto standard within the data centre industry, will be assessed. This is achieved by using open source specifications for a data centre in Prineville, Oregon, USA provided by the Open Compute Project launched by the social networking company Facebook. The usefulness of the PUE metric to the IT industry is critically assessed and it is found that whilst it is important for encouraging lower energy consumption in data centres, it does not represent an unambiguous measure of energy efficiency.

1. Introduction

By providing the Information Technology (IT) backbone for banks, businesses, hospitals, universities, and many other important services including the internet, data centres have become an integral part of operations across the world. Such facilities house IT equipment providing the means to store, process and share data. In the 10 years leading to 2005 servers have been developed to operate at higher processing speeds resulting in the associated waste heat dissipated by a typical rack increasing from 1 kW to 12 kW [1]. A survey by the Uptime Institute in 2012 found that the average rack density to be slightly lower at 8.4 kW, although their highest surveyed rack was 24 kW [2]. Together with an increase in the number of servers due to societal demand, electricity consumption both for computing and ancillary building services in data centres worldwide has led to a higher level of CO_2 emissions.

In 2011, it was reported that data centres consume 1.1–1.5% of worldwide electricity [3] with estimates showing that up to 2% of global CO_2 emissions can be accounted for by the IT sector [4]. A number of studies conducted on energy consumption and efficiency in data centres have estimated that they consume 40 [5], 15 [6] or 10–30 times [7] more energy per square foot compared to commercial office space. Similar studies on 14 data centres by the Lawrence Berkeley National Laboratory (LBNL) found energy consumption was between 120 and 940 W/m^2 [8] whereas only 50–100 W/m^2 was consumed in a typical commercial office space [9]. Assessing the energy consumption of processes and facilities such as data centres is of key importance if green house gas emissions from electricity generation are to be reduced. In order to measure the energy efficiency of data centres, metrics are a useful tool. However these metrics need to be fit for purpose [10] as inappropriate ones can act as a barrier to increased energy efficiency [7].

2. The power usage effectiveness metric

A range of metrics are currently available to assess data centres, however there is one particular metric that has over time become a de facto industry standard. The *Power Usage Effectiveness* (PUE) metric introduced in 2006 [11] and promoted by the Green Grid (a non-profit organisation of IT professionals) in 2007 [12] has become the most commonly used metric for reporting the 'energy efficiency' of data centres [7,13,14]. The PUE is useful to present the proportion of energy which is actually used to operate the IT equipment with respect to the total power draw of a facility, and is defined in Eq. (1). A partial PUE (pPUE) [15] can be used to assess the energy use of individual systems (such as cooling) compared to the IT load.

$$\text{PUE} = \text{Total Facility Energy}/\text{IT Equipment Energy} \quad (1)$$

Although it is named the 'power' usage effectiveness, the metric actually measures the energy use of the data centre. Yuventi and Mehdizadeh suggest adjusting the name of PUE to Energy Usage

* Corresponding author. Address: Energy Building, School of Process Environmental and Materials Engineering, Faculty of Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom.

E-mail address: cen5gab@leeds.ac.uk (G.A. Brady).

0196-8904/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.enconman.2013.07.035

Nomenclature

m	moisture content (kg water vapour/kg dry air)
φ	relative humidity (%)
P_s	saturation vapour pressure (Pa)
P	air pressure (Pa)
η	efficiency
W	power consumption (kW h/annum)

Subscripts

IT	IT load
C	cooling load
L	electrical losses
Other	other losses
T	transformer

Abbreviations

UPS	uninterruptable power supply
CPU	central processing unit

Effectiveness or EUE to avoid any confusion between power and energy measurements [16]. Only an instantaneous value gives the power usage, however the PUE needs to be measured over a year, hence it measures energy. The PUE is used at the design stage of projects to present a data centres' potential 'energy efficiency' then used post construction to aid in energy costing and monitoring the use of power. However, during the design stage accurate measurement is not possible and hence advertised PUE values are purely estimation. The metric has however still become a marketing tool, with owners/designers using it to promote the potential 'efficiency' of their data centre.

A PUE of 1 would be an ideal number; however this is not theoretically possible as it will always take some form of energy consumption to support the IT equipment. There is currently not enough data to illustrate the PUE of data centres on a world-wide scale [7] however, some smaller studies have been completed. 70% of 115 respondents in one particular study were aware of their PUE and an overall average value of 1.69 was reported [17]. A similar study for 22 data centres found PUE values of between 1.33 and 3 with an average value of 2.04 [6]. However neither of these studies provides detailed information regarding the locality of each facility, the scale of the operations within them, or how the energy consumption assessment was carried out.

A more in depth analysis of energy consumption was provided in [1] which assesses the energy consumption of a small data centre in Linköping, Sweden. Unfortunately, the energy consumption assessment of this data centre was only conducted during the coldest month of the year. A short assessment period was also used in [10], with the power requirement for the IT equipment only being taken over 1 month. Clearly, it is important to record the PUE over a representative period so that a realistic annualised average can be obtained, as demonstrated hypothetically in [16] and discussed in [18]. Only when the relevant data is collected over a year can a true representation of total energy use be presented, as it includes any influences from fluctuations in weather and IT load demand. Another study measured the energy consumption of two data centres in Singapore at intervals of 10 min for the duration of a week, presenting detailed information regarding the energy consumption of the IT, HVAC, UPS and lighting systems [9]. Due to the local climate experiencing relatively constant temperatures, the data could be extrapolated to represent a whole year. Despite this ability to measure a shorter period of time due to constant temperatures, this may still not be long enough to measure any variations in the IT load.

Even though the PUE metric is widely adopted in the industry there are some inherent problems with its indiscriminate use, especially in the calculation, interpretation and reporting of results. One of the problems with the PUE is that a significant amount of data is required (multiple parameters need to be measured over an entire year), making a direct verification difficult without access to this data. It can also be difficult to produce a design PUE that is a correct estimation for real life data centre operation. The difficulties of calculating the PUE without direct access to detailed data are explored in this work. A more open access industry regarding energy consumption would allow more ideas to be shared and more savings to be made.

Limited studies exist in the literature critically assessing the PUE despite the inherent problems with this metric. The PUE is reviewed along with other common metrics in one study [19] and authors in [16] provide a detailed assessment of the usefulness of PUE as a sustainability metric. However the focus in the latter is on stressing that the PUE should be measured over a representative period, even though this is already the defined intention of the metric.

An analysis of the literature has highlighted a lack of extensive energy consumption calculations which run for an annual cycle. It also shows a lack of detail in some studies for calculation of the energy efficiency, an exception is [5] which provides good detail (including information about hardware and cooling systems) and collates data over an entire year. Whilst there is literature available on the method of calculating the PUE, none look into the sensitivity of PUE to certain parameters or attempt to repeat a PUE calculation using open source information. Any PUE values that are presented in the literature are subject to strict privacy measures due to the sensitive nature of the industry, and the tendency of companies to be somewhat elusive about their designs means that data centres often have their energy consumption kept hidden from the public domain [10]. This makes it difficult to assess or verify PUE reporting with confidence. If energy use metrics are to be improved, it is important that reporting is transparent. This problem also makes it hard to communicate ideas for reducing energy consumption within data centres.

We investigate the PUE metric in more detail providing a critical analysis of its use. In order to aid in this analysis, open source specifications for a particular data centre [20] are used to conduct a PUE calculation. This would be a similar method to that conducted during the design process of the data centre using the typical (limited) information available to designers. Due to this, the study also serves to illustrate the difficulties of determining the PUE before the data centre is live. The energy usage values estimated will then be used to investigate the affect different parameters have on the PUE metric. The results from the case study show the difficulty of repeating a PUE measurement without detailed information, with the sensitivity analysis demonstrating that some factors can affect the PUE value more significantly than others.

We now describe the case study PUE measurement that uses data published in the Open Compute Project [20]. This project provides information and specifications for a data centre based in Prineville, Oregon in the USA. The data centre is operated by the social networking company Facebook and as such provides the IT services to support their website and users. One of the objectives of the Open Compute Project, is to share information and ideas about how to save energy in data centres. Initiatives such as this are important for the industry as sharing ideas will help to drive down energy consumption.

3. A case study on the Prineville data centre, Oregon

The aim of this case study is to use limited open source information to calculate the PUE of a data centre. The collated data will then be used to investigate the sensitivity of the PUE to certain parameters. A secondary aim of this case study is to see how much detail has been shared in the Open Compute Project and whether more is required to enable true comparisons of energy consumption. We are not aiming to present a methodology for calculating the PUE in this work, although the model presented could be used during the design stage of data centres, especially those that employ direct air cooling. The following section presents the calculation of the PUE using the information available from the Open Compute Project. The main aim of the following case study is to identify issues with reporting the energy efficiency of a data centre and problems with using the PUE metric.

3.1. The energy consumption of the IT servers

Measurement of the PUE requires the energy consumption of the servers to be accounted for. Whilst the Open Compute specifications do not indicate how much power is required to serve the total IT load in the data hall, the power supplied to each custom built server (further details of which can be found on the Open Compute website [21]) is given as 450 W (this power will also supply the internal fans in the server units). It is also known that 30 servers are supported in each rack, the racks being arranged in groups of three, termed as 'triplet racks' [22]. This leads to a maximum rating of 40.5 kW per triplet rack. However, there is no indication in the specifications as to how many of these triplet racks or servers there are in the entire facility, information which would make comparing energy use easier. Due to this, the energy consumption of the servers and the PUE will be estimated for one triplet rack. This does not account for other services, for example switches or storage, and is the best estimate possible with this limited information.

A server's power draw varies over time and they rarely operate at 100% of their rated power. They will also rarely operate at a full 100% CPU utilization, sometimes operating at a value as low as 10% [23]. Despite the variation though, an idle server (very low to zero CPU utilization) will still use a significant proportion of its rated power [24]. This proportion can be as high as 60% [25]. Without any usage data, we make the conservative assumption that the servers discussed above draw a continual 60% of the rated power, this gives an annual energy consumption of 212,868 kWh/annum per triplet rack. Using a peak value of 60% is a significant assumption to make (although the sensitivity of this will be examined below), given that the servers will not be using this level of power continuously throughout the year; however this assumption needed to be made given the limited amount of information provided in the Open Compute documents. Without direct access to the power use of the IT equipment, it is difficult to make an accurate prediction. This highlights a crucial inaccuracy that can arise from estimating the energy consumption or PUE for a data centre still in the design phase. The sensitivity of the PUE value to the assumed IT load is analysed in Section 3.5.

3.2. Energy consumption of the air handling and cooling systems

Energy costs for cooling and distribution of the air to maintain integrity of the electrical components can be significant. Depending on site location, air conditioning can consume a significant fraction of the energy required to run the ancillary services. In the case of the Prineville data centre, cooling is achieved through air-side economisation, where filtered outside air is delivered directly to the servers, and a high pressure misting system provides evaporative cooling and humidification. The facility uses hot aisle containment, with the air being delivered to the cold aisles through a series of rooms which house filters, misting jets, and fans [22].

3.2.1. Air distribution – the power required to operate the fans

Air delivered to and from the servers will experience a series of pressure drops, for example due to filtration, distribution to the cold side through vents and down aisles, and (similarly) through the return path. This will depend heavily on the design of the data centre (e.g. flow paths, duct sizes, distribution tiles).

The other significant pressure drop is across the rack server itself. Manufacturers generally specify on-board fans to ensure a low net pressure drop from the front to the back; one of the costs of cooling high density electronics is the energy associated with accelerating air to high velocities to affect the necessary heat transfer from the surface of heat exchangers and components within the server. Since fans generally sit within the rack server, this implicitly includes part of the energy costs for air distribution within the IT load. A move away from the multitude of small (inefficient) on-board fans to a larger fan unit that pressurised the inlet side of the server would lead to a larger PUE – despite a total reduction in energy consumption. In terms of air delivery, the Open Compute document indicates that each server would require a maximum $0.028\ m^3\ s^{-1}$ of air flow equating to $2.55\ m^3\ s^{-1}$ per triplet rack [26]. Along with pressure drop values for the whole system, the design air flow rate can be used to determine the total power required to operate the air distribution system. Open compute state that the filters in use have a pressure drop of 67 Pa at an air flow of $2.54\ m\ s^{-1}$ [22]. If additional pressure drops due to fans and air distribution were provided then the total energy requirement could be calculated.

However, detailed information about pressure drops and also room layout are not available. Therefore it is not possible to calculate the power required to provide air distribution within the data centre. Energy consumption for air distribution is therefore incorporated into the estimated energy use of the evaporative cooling system in the following section.

3.2.2. The energy consumption of the evaporative cooling system

Electronic components have increased failure rates when operating above or below allowable temperature and humidity levels [27]. The ASHRAE TC 9.9 guidelines recommend a delivery temperature range of 18–27 °C and a humidity range of 5.5 °C dew point to 60% relative humidity and 15 °C dew point [28], although the maximum allowable envelope within ASHRAE is actually wider than this. Data centres which employ air-side economisation take advantage of a recent expansion of the recommended ASHRAE operating envelope, as they allow ambient outside air to be delivered directly to the servers for a greater proportion of the year. Whilst there is an associated increase in expected failure rates associated with operating outside the allowable envelope, ASHRAE provide factors which show that this increase is actually very low especially, especially for regions such as North America [28].

In order to calculate the energy consumption of the cooling system, weather data must be used for the local area. Data from the same weather station given in the Open Compute specifications was used for this case study (sourced from [29]). The data used is a Typical Meteorological Year data set for Redmond, Oregon. Fig. 1 (plotted using a Matlab code [30]) shows the weather data plotted on a psychrometric chart with the outside air properties for each hour of the year represented by a point on the graph.

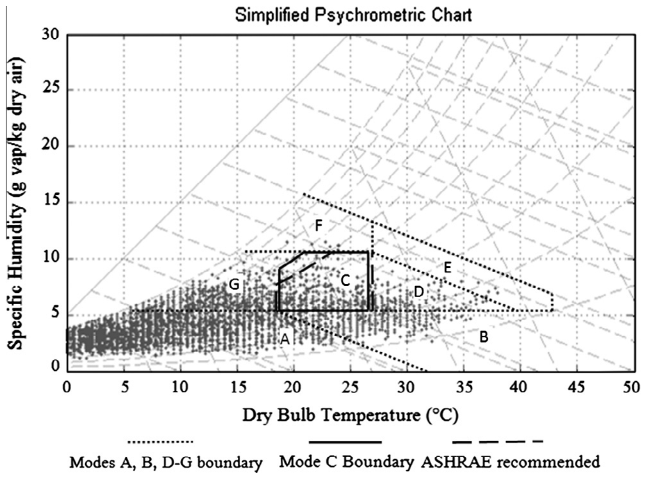

Fig. 1. Psychrometric chart with weather data for Redmond, Oregon and the cooling system modes A to G for the data centre as per information from (*Jay Park, 2011, Open Compute Project: Data Center v1.0*). Also shown is the ASHRAE recommended envelope.

Table 1
Description of each mode of operation for the cooling system.

Mode	Operation of cooling system
A	Evaporative cooling provides humidification, mixing of outdoor/return air to raise temperature
B	100% outdoor air, evaporative cooling provides humidification
C	100% outdoor air
D	100% outdoor air, cooled by evaporative cooling
E	100% outdoor air, cooled by evaporative cooling
F	Mixing of outdoor and return air
G	Mixing of outdoor and return air

The mechanical specifications for the Open Compute Project provide a description of the modes of operation for the cooling system (Table 1), and the psychrometric properties required for each one. Fig. 1 is divided into sections illustrating the cooling systems modes of operation as per the specifications. The chart has been used in conjunction with the weather data to calculate the proportion of the year that the misting system is required to provide evaporative cooling or humidification. Fig. 2 illustrates the results of this analysis.

The analysis shows that the misting system will need to be in operation for 85.8% of the year. In order to account for the pumping power, it is necessary to calculate the amount of water that needs to be delivered to the supply air through the misting system. The TMY weather data was once again used to calculate the moisture content of the outdoor air relative to the required moisture content of the supply air to the data hall. Calculation of the moisture content (m) was possible by using the psychrometric relationship between relative humidity (φ), air pressure (P), and saturation vapour pressure (P_s) as shown in Eq. (2). Explanations of this equation can be found in [31,32]:

$$m = (0.6219\varphi P_s)/(P - \varphi P_s) \quad (2)$$

Given that:

$$\text{For } 0\ ^\circ\text{C} < T < 63.0\ ^\circ\text{C}; \quad P_s = 610.78\exp[17.269T/(237.3 + T)] \quad (3)$$

$$\text{For } T < 0\ ^\circ\text{C}; \quad P_s = 610.78\exp.[21.874T/(265 + 0.9615T)] \quad (4)$$

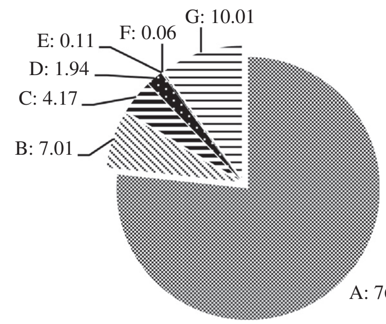

Fig. 2. Percentage of the year each system mode (A–G) is in operation for, based upon weather data.

Eqs. (3) and (4) calculate the saturation vapour pressure [32], having been modified for different temperature ranges based upon original work in [33]. The above set of equations are used to calculate the quantity of water that needs to be added to the outside air per hour given the ambient conditions and the required conditions of the air supplied to the servers. The TMY weather data was used to see how much cooling would be required for each hour of the year to bring the air to the minimum standards required by the design requirements for the data centre (boundary C in Fig. 1). Calculations conducted using the above equations estimate that the amount of water required during a typical year would be 148 m^3 per triplet rack (equivalent to 6.1 m^3/kW of cooling load given continuous 60% server power utilisation), with a maximum flow rate during the year of 0.054 m^3/h per triplet rack (0.7 l/kW h). Given the water flow, it would be possible to then calculate the energy required to pump this through the misting system. However no details about energy utilisation during pumping, spray generation or filtration are provided. This means that it was necessary to use power consumption figures from another data centre. A comparison has therefore been made to a data centre in Bedford, UK which uses evaporative cooling and requires 4 kW for the air distribution fans and high pressure pumps when the misting system is in operation [34].

Using the figures provided for the Bedford facility, and given that the water flow rate in Prineville is 3.8 times higher, it could be approximated that 119 W is required to operate the misting pumps and air distribution fans for each triplet rack. This gives a total predicted consumption of 893 kW h/annum per triplet rack. Since the data in [34] incorporates air conditioning and air distribution, the figure above is increased to incorporate the 14.2% of the year when the misting system is not in operation, giving a consumption of 1048 kW h/annum. This will be a slight over estimate in energy requirements, since the droplet misting is not required, however filtration will still be required to maintain a clean environment within the data centre.

3.3. Miscellaneous and other loads on the electrical system

3.3.1. Lighting. The data centre uses LEDs which are much more energy efficient than incandescent lamps. Power is delivered to them through network cables rather than electrical wires, a method known as Power over Ethernet. Material is saved through the reduced need for electrical wiring providing an additional environmental benefit. Using Power over Ethernet, energy consumption for the LED lamps can be included in the denominator of the PUE, reducing its value when compared to more traditional lighting arrangements. Whilst there is a reduction in the energy consumption due to the use of LED lamps, there may be an additional decrease in the PUE by including the lighting in the denominator. This is one of the ways that the PUE metric can be used incorrectly.

3.3.2. Electrical losses. The power distribution system in a typical data centre will incur losses of energy through inefficiencies in its equipment, mainly the transformers, Power Distribution Unit (PDU) and the Uninterruptible Power Supply units (UPSs). The Prineville facility makes use AC supply with a DC backup system to protect servers in the event of a power supply failure, which requires a continuous supply of DC power to the servers. With a 95% efficient power delivery network, this is another contributing factor towards high efficiency in the data centre. There is also a loss due to power transformation of 2% [35]. The energy losses in the power network are calculated here based upon the IT load, air distribution and evaporative cooling systems, totalling a loss of 15,854 kW h/annum per triplet rack.

3.4. Energy efficiency of the Prineville data centre

The values computed above are now used in Eq. (1) to produce a PUE value for a triplet rack as follows:

$$\mathrm{PUE} = (212,868 + 1048 + 15,854)/212,868 = 1.08 \tag{5}$$

Facebook reports that the Prineville data centre achieved a PUE of 1.08 (for the end of quarter three in 2011) [35]. The live PUE reporting tool provided for the data centre indicates that the PUE has fluctuated between 1.11 and 1.07 over the year leading to June 2013 [36]. The calculations above appear to enable this PUE figure to be reproduced accurately, even though loads such as plant room lighting and generator heaters have not been included. It is clear from the numbers in (5) that the energy usage of the cooling system is very small compared to the IT energy usage, this is due to the design of this particular data centre. Although assumptions were made during the calculation, this means that any inaccuracies in estimating the energy consumption of the cooling system are unlikely to cause large differences in the calculated PUE in this case. Differences in the cooling system energy usage are examined below in the sensitivity analysis.

3.5. Sensitivity analysis

Despite the assumptions that needed to be made, the above PUE analysis provides a good estimate of the published value from Open Compute. In this section a brief sensitivity analysis of the effect of these assumptions on the calculated PUE is carried out. The PUE is defined as the total energy use divided by the energy consumed by the IT equipment (W_{IT}). Total energy consumption is equal to the total amount of energy used by the equipment and infrastructure in the facility (W_T) plus the energy losses due to inefficiencies in the power delivery network (W_L), hence:

$$\mathrm{PUE} = (W_T + W_L)/W_{\mathrm{IT}} \tag{6}$$

In a typical data centre the total energy consumption, W_T, includes the energy used by the IT equipment and supporting infrastructure. The following sensitivity analysis considers the effects that the electrical losses, IT load and the humidification system at Prineville have on the PUE. The analysis consisted of three scenarios to see which parameter causes the largest changes in PUE. The power requirement for the humidification system and the efficiency of the power distribution network were increased or decreased by 20% and 1% respectively. Then the IT equipment was investigated by decreasing the power draw of the servers to 20% and increasing it to 80%. Within each scenario, only one parameter was changed. When it came to decreasing the utilisation of the IT equipment, the cooling and air distribution power requirements were not scaled down to account for the lower IT load, as this does not always happen in data centres [37].

Fig. 3 illustrates the percentage changes in the PUE value depending on which parameter has been altered. It demonstrates that an alteration in the energy use of a particular system causes a corresponding change in the PUE value. Changes in the energy consumption of the cooling system have minimal effect on the PUE value, whereas a much larger difference is noted when changes in the energy consumption of the IT load and efficiency of the power supply network are made. It is clear from the graph that the greatest impact on the PUE came from changing these latter values. For the power supply efficiency, the changes are even more important when considering that this was only altered by ±1%, leading to a corresponding +1% change in PUE. By comparison, the PUE changed minimally (±0.1%) when the power required for humidification was altered by ±20%. The significance of the power distribution efficiency (η_E) and the IT load can be explained by the following expressions, given their relationship to the total energy use in the facility:

$$W_L = (W_T/W_{\eta E}) - W_T \tag{7}$$

Therefore:

$$\mathrm{PUE} = [W_T + (W_T/\eta_E) - W_T]/W_{\mathrm{IT}} \tag{8}$$

$$\mathrm{PUE} = W_T/(\eta_E W_{\mathrm{IT}}) \tag{9}$$

Eq. (9) highlights the fact that whatever energy savings are made in the data centre, the power distribution efficiency and IT load will still have a significant effect on the overall PUE value.

The sensitivity analysis also demonstrates that the PUE should be used with caution due to the effect of IT load changes. For example, if the IT load increases through improved utilisation, the PUE will reduce in spite of the increased overall power consumption in the data centre. This highlights the fact that a reduced PUE can predict that a data centre is operating more efficiently even though its overall energy consumption has increased, and vice versa. This is one of the PUE's main limitations as an energy efficiency metric.

4. Critical assessment of the PUE metric

A number of significant assumptions were required to complete the PUE analysis using the Open Compute Project, together with (at times) convoluted ways of estimating power requirements. A more realistic PUE could have been calculated if information about the number of racks and the mix of ancillary IT equipment, together with engineering specifications of the air handling units was available. Despite this, the estimated PUE agreed with that published through the Open Compute Project. The main aim of this study has been to assess some of the issues surrounding the use of PUE, its findings are now summarised.

4.1. Benefits

Measuring the energy efficiency of a data centre is clearly very important if carbon emissions from the IT sector are to be reduced, and if companies are to reduce their electricity consumption. The

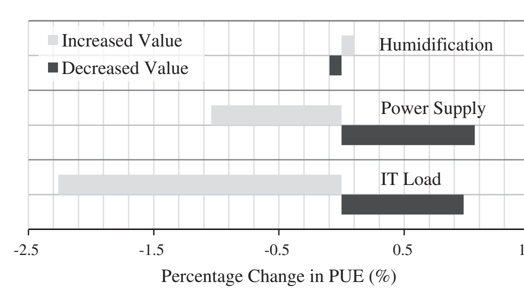

Fig. 3. Percentage changes in PUE as a result of altering different parameters within the metric: (i) humidification by ±20%, (ii) η_E by ±1%, (iii) IT load by −40% and +20%.

PUE metric is useful to the individual data centre in assessing overall energy consumption since large values are clear indicators of system inefficiency. It is also useful for recording annual variations in usage effectiveness as it should be monitored continuously over a year. The metric has also gone some way to creating competition, driving efficiencies up as advertised PUE values become lower.

4.2. Limitations

Through the study of PUE, it is clear that there are some issues involved with the metric. Measurement of the PUE involves accounting for the energy consumption of a number of different systems. It has been found that unless direct access to data and information is provided, detailed measurement of the PUE is a difficult task. Whilst the Open Compute Project is open source, there was not sufficient information to conduct a detailed re-calculation of the PUE and hence the calculation could only be done based upon one triplet rack.

It is crucial that an accurate IT load is used for the PUE, and that it is not based upon the rated power use of the equipment [38]. Accuracy in the IT load is one of the major factors affecting the measurement of the PUE metric, as utilisation of the servers has an important effect on IT energy consumption and hence the overall PUE value. The sensitivity analysis showed that the PUE will tend to increase if the IT load decreases and the cooling system is not scaled back to account for this change. Careful use of virtualization where servers are progressively loaded to capacity could allow some servers to be switched off [39]. This will reduce energy consumption (although unless the cooling system power is reduced the PUE will increase).

Although the PUE is commonly treated as an energy efficiency metric, it is actually an effectiveness metric. It does not accurately reflect the overall energy efficiency of a data centre as it does not include the efficiency of power use in the IT equipment. A data centre with a low PUE but with low server utilisation can be less efficient overall than a facility with a higher PUE value and a higher server utilisation.

Problems can also arise with the reporting of the PUE metric. Ideally, the PUE should be reported along with information regarding what period it is measured over, and where the power measurements are taken. This would limit the problem that has arisen with the PUE value of different facilities being compared. This was never the original intention of the metric – differences in locality (and hence climate) will inherently affect the ultimate PUE achievable at a given location due to the requirements for cooling of the servers. The comparison of a facility where free cooling is available with one that is not, says little about the inherent efficiency of the equipment. A comparison of data centres that includes climate information and data analysis alongside the PUE would give a more comparable ranking of energy efficiency.

At the design stage of a data centre, an estimated PUE value will normally be used as a marketing tool to potential owners. However, once the data centre is operational it is normally only able to achieve this PUE value when it is at full IT capacity. This means that comparing a PUE value of data centres is somewhat meaningless unless it is known whether it is operating at full capacity or not. The PUE metric is more useful for data centres that are operating at full IT capacity, as this should mean that the supporting infrastructure is also operating at its designed capacity. The reuse of waste heat either within the data centre or by an external consumer (such as heating in an office) is a valuable method of reducing overall energy consumption. However, the PUE itself does not take into account any reuse of energy, for this purpose the Green Grid created the ERE or the Energy Reuse Effectiveness [40]. As the PUE only assesses the actual consumption of energy within the data centre, it is not designed to incorporate any reuse of waste heat from the servers. Technology that enables the reuse of heat within the data centre itself such as the absorption chillers suggested by Haywood et al. [41] is not able to have its full benefits reflected in the PUE. Therefore in situations such as this, careful use of the correct metric is needed to enable the technologies full benefits to be realised, Haywood et al. suggest the use of an alternative metric to the PUE.

5. Conclusion

Metrics are essential if energy efficiency and energy consumption are to be assessed in a data centre; however they must be suitable and fit for purpose. A case study has been conducted in order to analyse the PUE metric which has become a de facto industry standard. This case study has demonstrated the type of detailed engineering data that is needed in order for meaningful PUE values to be calculated. Even when open source specifications are given around IT choice, this level of detail is not necessarily available. The assumptions made in this work enabled good agreement of the published PUE for a data centre to be completed. However the data centre in this study is a special case and has a unique design, leading to low energy consumption by the cooling system. Due to this, the energy use of the cooling system has little effect on the PUE value. The case study also illustrates the ease of simplifying the PUE calculation to achieve a good value, something which may occur during the design process.

The sensitivity analysis shows that once a data centre has made significant energy savings through reducing cooling system power requirements (as in the case of the data centre assessed in this work), the efficiency of the power supply network and the IT load will have the largest effect on the PUE value. The analysis also highlights problems with the relationship between changing IT loads and the PUE value, demonstrating that the PUE metric must be used with caution. This is due to the fact that energy saving measures such as virtualization can actually increase the PUE, falsely implying a less energy-efficient operation.

Even though the PUE is not an energy efficiency metric, it has over time become a marketing tool to present the overall efficiency of a facility. Despite the fact that PUE values cannot be directly compared, its use has helped to create an industry where data centres have become more competitive in their energy use and efficiency. The metric has helped to set benchmarks for energy consumption relative to the IT load in the data centre. However, the metric does not show a true representation of energy efficiency in a data centre, due to it not including the efficiency of the servers in their required operation. The data centre industry needs a metric which incorporates energy efficiency of all the equipment and infrastructure including the server units, but also a metric which is useful when comparing one facility to another. Incorporating IT equipment operational efficiencies and also climate/weather information may allow for a fuller picture to be painted about energy efficiency, allowing more direct comparisons between facilities.

Significant gains have been made in increasing the efficiency of data-centres through careful choice of location and ancillary equipment. PUE was never intended to bench-mark data-centres however the (understandable) desire to rank facilities has lead to some distortion of the PUE calculation process. Little discussion has focussed on the efficiencies of the IT processes; the PUE metric does not necessarily drive such improvements. An analogy between a manufacturing environment and a data centre where raw materials (unprocessed data/requests for information) are processed (computed) before dispatch (through networks) would suggest that a careful examination of the unit cost of operation will lever efficiency gains. This will allow data-centre operators to balance responsiveness and resilience in a holistic sense.

Acknowledgements

Gemma Brady would like to thank the EPSRC for the funding support provided through the DTC Energy: Technologies for a Low Carbon Future (Grant EP/G036608/1).

References

[1] Karlsson JF, Moshfegh B. Investigation of indoor climate and power usage in a data center. Energy Build 2005;37:1075–83.
[2] Institute U. Uptime Institute Annual Report: Data Center Density Preliminary Results 2013. In: Uptime Institute Symposium. Santa Clara, CA; 2013.
[3] Koomey JG. Growth in data center electricity use 2005 to 2010. Analytical Press; 2011.
[4] Gartner. Green IT: the new industry shock-wave. In: Symposium/ITXPO San Diego, CA; 2007.
[5] Lu T, Lü X, Remes M, Viljanen M. Investigation of air management and energy performance in a data center in Finland: case study. Energy Build 2011;43:3360–72.
[6] Greenberg S, Mills E, Tschudi B, Rumsey P, Myatt B. Best practices for data centers: lessons learned from benchmarking 22 data centers. In: Proceedings of the 2006 ACEEE summer study on energy efficiency in buildings. Pacific Grove, CA; 2006.
[7] Daim T, Justice J, Krampits M, Letts M, Subramanian G. Data center metrics. Manage Environ Qual 2009;20:712–31.
[8] LBNL. Data centre website of Lawrence Berkeley National Laboratory; 2003 <http://hightech.lbl.gov/datacenters> [accessed 03.13].
[9] Sun HS, Lee SE. Case study of data centers' energy performance. Energy Build 2006;38:522–33.
[10] Mitchell-Jackson J, Koomey JG, Nordman B, Blazek M. Data center power requirements: measurements from Silicon Valley. Energy 2003;28:837–50.
[11] Malone C, Belady C. Metrics to characterize data center & IT equipment energy use. In: 2006 digital power forum. Richardson, TX, USA; 2006.
[12] The Green Grid. Green grid metrics: describing datacenter power efficiency: technical committee white paper. White Paper #1; 2007 <http://www.thegreengrid.org/~/media/WhitePapers/Green_Grid_Metrics_WP.pdf?lang=en>.
[13] Kant K. Data center evolution: a tutorial on state of the art, issues, and challenges. Comput Networks 2009;53:2939–65.
[14] Ruth S. Reducing ICT-related carbon emissions: an exemplar for global energy policy. IETE Tech Rev 2011;28:207–11. <http://tr.ietejournals.org/text.asp?2011/28/3/207/81229>.
[15] Grid TG. PUE: a comprehensive examination of the metric. White Paper #49; 2012.
[16] Yuventi J, Mehdizadeh R. A critical analysis of power usage effectiveness and its use in communicating data center energy consumption. Energy Build 2013;64:90–4. <http://www.sciencedirect.com/science/article/pii/S037877881300251X>.
[17] The Green Grid. Survey results: data center economizer use. White Paper #41; 2011 <http://www.thegreengrid.org>.
[18] Avelar V, Azevedo D, French A. PUE: a comprehensive examination of the metric. White Paper #49: Green Grid; 2012.
[19] Patterson M. Energy efficiency metrics. In: Joshi Y, Kumar P, editors. Energy efficient thermal management of data centers. Springer Verlag; 2012.
[20] Open Compute Project. Hacking conventional computing infrastructure. <http://www.opencompute.org> [accessed 04.13].
[21] Project OC. Motherboard and server design. <http://www.opencompute.org/projects/motherboard-design/> [accessed 06.13].
[22] Park J. Open compute project: data center v1.0; 2011 <http://www.opencompute.org>.
[23] Greenberg A, Hamilton J, Maltz DA, Patel P. The cost of a cloud: research problems in data center networks. ACM SIGCOMM Comput Commun Rev 2009:39.
[24] Mohan Raj VK, Shriram R. A study on server sleep state transition to reduce power consumption in a virtualized server cluster environment. In: fourth international conference on Communication systems and networks (COMSNETS), 2012. p. 1–6.
[25] Barroso LA, Hölzle U. The case for energy-proportional computing. IEEE Comput 2007:33–7.
[26] Furuta S. Open compute project: server chassis and triplet hardware v1.0; 2011. <http://www.opencompute.org>.
[27] Ohring M. Reliability and failure of electronic materials and devices. Boston, MA: Academic Press; 1998.
[28] American Society of Heating Refrigeration and Air-Conditioning Engineers. In: 2011 thermal guidelines for data processing environments – expanded data center classes and usage guidance. ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment; 2011.
[29] National Renewable Energy Laboratory. National solar radiation data base: typical meteorological year 3. Data Set 726835 Redmond Roberts Field; 2008.
[30] Ngai T. Psychrometrics with Matlab. 2011 <http://www.tedngai.net/teaching/psychrometrics-with-matlab.html> [accessed 07.12.11].
[31] McQuiston FC. In: McQuiston Faye C, Parker Jerald D, Spitler Jeffrey D, editors. Heating, ventilating, and air conditioning: analysis and design. Hoboken, NJ: John Wiley & Sons; 2005. International ed..
[32] Singh AK, Singh H, Singh SP, Sawhney RL. Numerical calculation of psychrometric properties on a calculator. Build Environ 2002;37:415–9.
[33] Dilley AC. On the computer calculation of vapor pressure and specific humidity gradients from psychrometric data. J Appl Meteorol 1968:7.
[34] Modern Building Services. 2.5 MW data centre is kept cool by evaporative cooling; 2010 <http://www.modbs.co.uk/news/fullstory.php/aid/8384/2.5_MW_data_centre_is_kept__cool_by_evaporative_cooling_.html> [accessed 11.11.11].
[35] Open Compute Project. Energy efficiency; 2011 <http://opencompute.org/about/energy-efficiency/> [accessed 12.12.11].
[36] Facebook. Prineville, OR Data Center; 2013 <https://www.facebook.com/prinevilleDataCenter/app_399244020173259> [accessed 06.13].
[37] Malone C, Belady CL. Optimizing data centre TCO: efficiency metrics and an infrastructure cost model. ASHRAE Trans 2008;v114 part 1:44–50.
[38] Itoh S, Kodama Y, Shimizu T, Sekiguchi S, Nakamura H, Mori N. Power consumption and efficiency of cooling in a data center. In: 2010 11th IEEE/ACM international conference on grid computing (GRID 2010), 25-29 Oct 2010. Los Alamitos, CA, USA: IEEE Computer Society; 2010. p. 305-12.
[39] Schulz G. The green and virtual data center. Florida, USA: CRC Press; 2009.
[40] Grid TG. ERE: a metric for measuring the benefit of reuse energy from a data center. White Paper #29; 2010.
[41] Haywood A, Sherbeck J, Phelan P, Varsamopoulos G, Gupta SKS. Thermodynamic feasibility of harvesting data center waste heat to drive an absorption chiller. Energy Convers Manage 2012;58:26–34. <http://www.sciencedirect.com/science/article/pii/S0196890411003712>.