



Analysis of performance metrics for data center efficiency

– should the Power Utilization Effectiveness PUE still be used as the main indicator? (Part 1)



TOM VAN DE VOORT
Eindhoven University of Technology
Department of Architecture, Building, and Planning
the Netherlands
t.v.d.voort@student.tue.nl



VOJTECH ZAVREL
Eindhoven University of Technology
Department of Architecture, Building, and Planning
the Netherlands
v.zavrel@tue.nl



IGNACIO TORRENS GALDIZ
Eindhoven University of Technology
Department of Architecture, Building, and Planning
the Netherlands
j.i.torrens@tue.nl



JAN HENSEN
Eindhoven University of Technology
Department of Architecture, Building, and Planning
the Netherlands
j.l.m.hensen@tue.nl

To halt the ever-increasing energy consumption by data centers it is important to use performance indicators which accurately represent this performance. The strengths and limitations of PUE as the key performance indicator are analyzed and suggestions are made to complement any limitations.

Data centers were responsible for 1.5% of global energy consumption in 2010 and this figure is only expected to double soon. Data centers are becoming more energy efficient, a trend lead by the introduction of PUE (Power Utilization Effectiveness) as a performance metric. PUE's simplicity and focus on infrastructure efficiency was quickly adopted by the industry, but now the question is raised if PUE is still able to lead the quest for improved energy efficiency. PUE does not show performance regarding IT efficiency, water usage, heat recovery, on-site energy generation or carbon impact. This can lead to misuse of PUE by focusing on just improving PUE values instead of real energy use. Improving data center performance assessment is proposed in this paper by broadening the scope beyond PUE.

Key Words: PUE, Performance metrics, Data Center, Energy Efficiency, Indicators

A data center is a building which houses IT hardware, like computational units, network infrastructure and data storage, next to supporting equipment, like cooling and power supply. What all

these different types of equipment have in common is that they are all high-energy density systems. This results in great amounts of energy being used, but also major energy savings potential by improving these systems.

Context

The data center industry is growing rapidly and the overall industry is expected to have an annual growth rate of over 10% until 2019 (Technavio, 2015). This is caused by the increase in number of chip driven appliances from 3 billion devices in 2010 to 15 billion devices in 2015 and it is expected to increase to up to 50 billion devices by 2020 (Modoff *et al.*, 2014).

Because of the growth of the data center industry, its energy consumption is rapidly increasing as well. Data center energy consumption accounted for between 1.1% and 1.5% of global energy consumption and up to 2.2% of US energy consumption in 2010 (Kooimey, 2011). This meant a 56% increase over the period between 2005 and 2010 after doubling between 2000 and 2005 (*idem*, 2011). The slowing of this trend has partially been caused by increasing energy prices leading to increased operational cost. Giving more incentive to adopt energy efficiency strategies. Another important reason is the economic crisis in 2008. Despite this, energy consumption by data centers is still predicted to double between 2010 and 2020 (Whitney *et al.*, 2014), thus requiring more focus on energy efficiency measures to halt this trend.

Research and Markets (2015) proposes a 30% annual growth rate for green data centers compared to 10% for the whole industry. This predicted demand for energy efficient data centers shows a way forward, but the question is how to accomplish and monitor such progress.

Performance metrics

Over the last years, different performance metrics have been introduced to measure and compare performance and efficiency of data centers. These metrics can be used to assess individual pillars (cooling, IT, power supply) of the data center or the data center as a whole. This can relate to total energy use, water use or carbon emissions, as well as subsystem efficiency like temperature distribution (Wang *et al.*, 2011).

PUE: Power Utilization Effectiveness

The most widely used performance metric is PUE, which shows the ratio between total facility power use and IT equipment power use (Averal *et al.*, 2012):

$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}; 1 \leq PUE$$

Therefore, the optimal value for PUE is 1.0, the maximum value is infinity. PUE has been developed to give data collection standards 'to determine the effective-

ness of any changes made within a given data center' (*idem*, 2012). Beyond its intended use, PUE has been adopted by the industry to make comparisons between data centers. The limited scope PUE offers can make it unreliable for comparison as some strategies can improve PUE values without reducing energy consumption. There is also a lack of strict measurement and reporting guidelines, only recommendations exist. This leads to publishing PUE values based on designed nominal values instead of part-load values measured during operation. (Donnelly, 2015). A survey by Uptime Institute (2014) found that 'a large majority (75%) of participants said the data center industry needs a new energy efficiency metric'. Which is part of the aim of this study.

The analysis presented in this paper will start by discussing the merits and shortcomings of PUE and continue by presenting other metrics to complement these shortcomings and try to find improved ways for accurately assessing data center energy performance.

Research methodology

To find a solution to the problem described above, the following research question has been formulated:

'Does the broadly accepted PUE metric reflect the real energy performance of a data center?'

This question is answered by performing a literature review on PUE and other available performance indicators.

The first part of this literature review focuses on both the merits of the PUE metric and its limitations. The misuse of the metric resulting from these limitations is also discussed as this helps to illustrate the reason behind the need of complementary performance metrics.

To solve the issues raised in the first part of the literature review, the second part consists of a review of existing metrics which can be used to complement PUE, or in some cases could even replace PUE. An overview of relevant metrics and their intended purpose is provided.

PUE analysis

PUE merits

The total efficiency of a data center comes down to how much useful work is produced per unit of energy. But as different data centers perform different tasks and are often relying on external input, useful work is difficult to determine. Therefore, PUE gained popularity as it shows the efficiency not by quantifying useful work, but by showing the ratio of energy available for useful work and the part that is lost to overhead, also referred

to as the infrastructure. This led to an industry wide adoption of PUE as the main performance metric after PUE's introduction. As in the data center industry energy consumption is one of the major expenses, what all data centers have in common, despite their different specializations, is the requirement to reduce their infrastructure energy consumption to increase efficiency.

When PUE was introduced in 2007 it provided new guidelines for measuring and reporting the internal energy flows in data centers. Industry average PUE values found after PUE's introduction lay between 2.5 and 3.0 in various studies (Foster, 2013). By using the framework provided by PUE average values have decreased to around 1.7 in the last major industry survey by uptime industries (Stansberry, 2015). In this way, PUE has led the first major industry shift towards energy efficient data centers.

For state-of-the-art large-scale internet data centers the PUE value has always been significantly lower and is close to values of 1.1 now (Google, 2016). Which means further improvement within the boundaries PUE provides is difficult. This underlines the need for other metrics to broaden the scope of energy efficiency assessment beyond PUE to further lower data center energy consumption.

PUE limits and misuse

As said, PUE has been used for comparison since its introduction. As the green grid (2012) states 'the metric

is best applied for looking at trends in an individual facility over time and measuring the effects of different design and operational decisions within a specific facility'. Despite the recommendation of applying PUE for internal use it's understandable that it started to be used for comparison. If a facility reports very low PUE values other facilities will be interested in the ways to achieve this efficiency. This also led to infrastructure designers 'rating' their system with achievable PUE values, but as no strict guidelines apply to the origins of these values it often remains unclear for which conditions they were calculated and if they can be achieved in real life. PUE is supposed to be a tool to decrease the energy consumption of data centers, but decreasing the PUE value has become the goal itself. This leads to strategies where PUE doesn't necessarily reflect real energy performance.

As it can be taken as a fact that PUE will be used for comparison, its reporting parameters should be better regulated. At this moment, there is a lot of flexibility in choosing the measurement point for a data center's energy use (appendix A). As PUE was introduced for internal use it can be decided within a data center which level of monitoring is chosen. For comparison and marketing purposes it is obvious that you would like to choose the best-case scenario. Regulating this reporting parameters can greatly increase the reliability of the PUE metric.

Guidance as to which measurement points and intervals are required and recommended for each PUE measurement level.

Where do I measure?		Level 1 (L1)	Level 2 (L2)	Level 3 (L3)
How often do I measure?		Basic	Intermediate	Advanced
IT Equipment Energy	Required	UPS outputs	PDU outputs	IT equipment input
Total Facility Energy	Required	Utility inputs	Utility inputs	Utility inputs
	Additional recommended measurements*		UPS inputs/outputs	PDU outputs UPS inputs/outputs Mechanical inputs
Measurement Intervals	Required	Monthly	Daily	15 minutes
	Additional recommended measurements*	Weekly	Hourly	15 minutes or less

* Recommended measurements are in addition to the required measurements. The additional measurement points are recommended to provide further insight into the energy efficiency of the infrastructure.

The scope of PUE is limited to energy consumption and as stated by The Green Grid (2012) ‘PUE awards no credits or percentage points for on-site energy generation, waste heat recovery, etcetera. While important, these are not the focus of the PUE metric’. Also, the energy source being used isn’t monitored by PUE. Electricity generated by PV-panels is treated the same as electricity from a coal plant. By including the ecological impact of the energy source the total energy impact can be better assessed.

Other forms of resource consumption fall beyond the scope of PUE, like water consumption by evaporative cooling. Especially when treated water is used for this purpose a significant energy impact exists. Broadening the scope of performance assessment to include these effects will increase the complexity, but will help to promote the circular use of resources and the use of low impact energy sources.

Maybe the most important issue with using PUE as the guiding performance metric is its disregard for IT equipment efficiency. As the computational power per watt increases per Moore’s law (Moore, 1965) the useful work produced per watt can double every two years, therefore renewing IT equipment might be one of the best energy efficiency strategies. ‘A typical data center’s PUE is likely to vary with the levels of its IT load’ (Green Grid, 2012). And as illustrated in **Figure 1**, PUE values are better during periods of high relative IT load. **Figure 2** illustrates how the average IT load can drop when more efficient IT equipment is installed, causing a degradation in PUE values. This is obviously not a desirable effect for accuracy of performance evaluation using PUE. When the cooling temperature set point is increased the PUE value doesn’t accurately reflect real performance. This leads to a decrease of energy consumption by the cooling system, but an increase of IT equipment energy use as the server fans speed up. Also, the electric resistance of the IT equipment increases together with IT energy consumption. It is obvious that the PUE value improves, but total energy consumption might be unchanged or could even increase (Hartfield, 2011). This effect is illustrated in **Figure 3**.

Performance metrics complementing PUE

As made clear in the previous section, to provide a complete assessment of energy efficiency for the data center industry through performance metrics, the scope should be widened from PUE alone. On the other hand, it is important to track the efficiency of separate

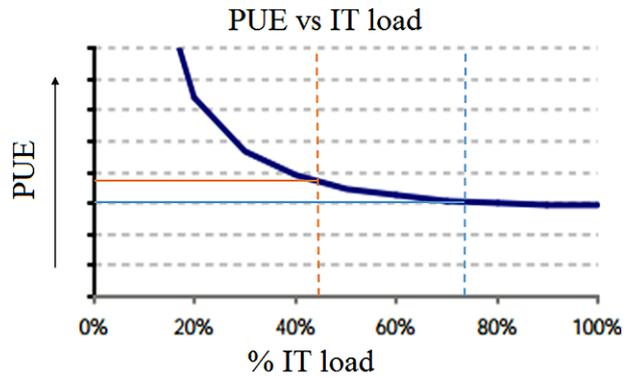


Figure 1. Relationship between PUE and IT Load with example from Figure 2 (adapted from Bisci 2009).

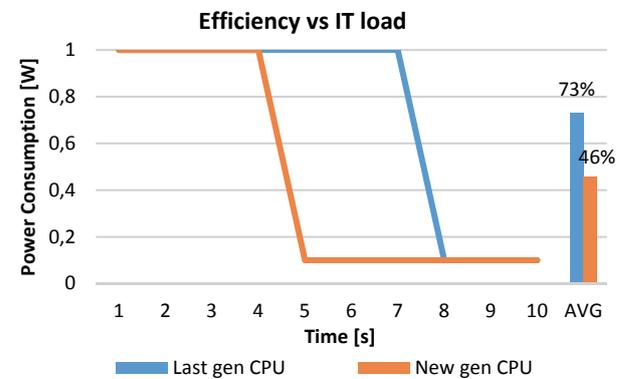


Figure 2. Relationship between efficiency and IT Load (adapted from Wasson 2015).

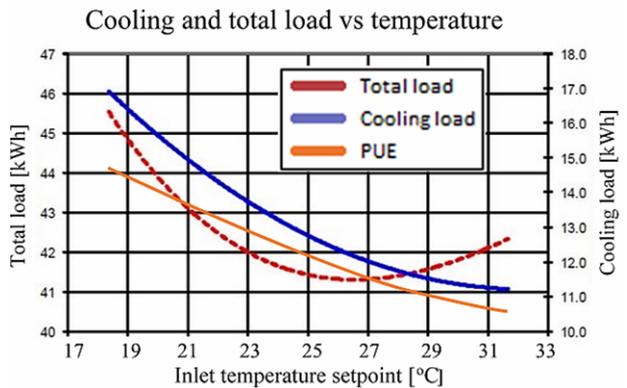


Figure 3. Typical relationship between temperature set point, cooling load, total load and PUE (adapted from Hartfield 2011).

parts of the data center as well. PUE performs very well if its limits are respected. Therefore, it is proposed to use complementing metrics to PUE addressing the previously presented issues.

Energy source impact: CUE

To give insight into the primary energy impact of a data center and related carbon emissions the Carbon Usage Effectiveness (CUE) metric has been selected to evaluate this aspect (Belady, 2010). From the same developers of PUE, CUE multiplies the total facility energy with its Carbon Emission Factor (CEF), being the carbon emitted per unit of energy. It is defined as:

$$CUE = \frac{CEF * Total\ energy}{IT\ energy} \left[\frac{kgCO_2}{kWh} \right]$$

$$0 \leq CUE$$

This adds information about the data center’s ecological footprint. If the data center has multiple energy sources, like a combination of grid-sourced electricity and on-site renewable sources, the partial contribution of both should be considered. Adopting the CUE metric will incite the industry to choose low impact energy sources, like on-site renewables.

On-site renewables: OEF & OEM

The CUE metric already reflects the positive impact on-site renewables can have on the total energy impact of a data center, but doesn’t provide enough insight on the effectiveness of these on-site renewables. To evaluate the energy (mis)matching the On-site Energy Fraction (OEF) and On-site Energy Matching metrics have been chosen. They are defined as:

$$OEF = \frac{\int_{t_1}^{t_2} Min[R(t); L(t)]dt}{\int_{t_1}^{t_2} L(t)dt}; 0 \leq OEF \leq 1$$

$$OEM = \frac{\int_{t_1}^{t_2} Min[R(t); L(t)]dt}{\int_{t_1}^{t_2} R(t)dt}; 0 \leq OEM \leq 1$$

where R(t) is the on-site generated renewable power and L(t) is the load power at time ‘t’. And ‘dt’ is the time-step of the calculation (Cao, Hasan and Sirén 2013).

Ideally, the on-site renewable generation is equal to the facility power, this is where the lines in **Figure 4** intersect. Area I is the amount of useable renewable energy, Area II is the surplus generation distributed

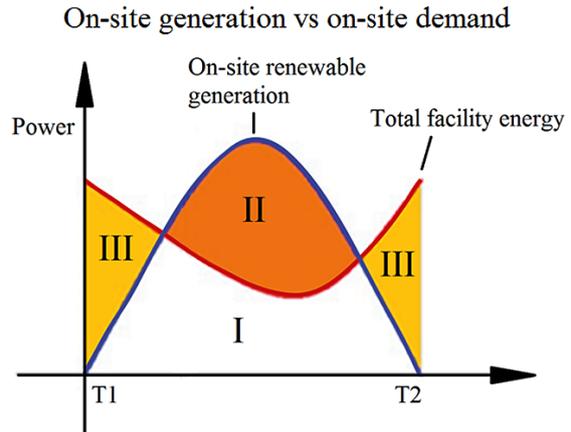


Figure 4. On-site generation vs. demand (adapted from Cao et al. 2013).

back to the grid (when OEM < 1) and area III is the energy required from the grid (when OEF < 1).

Information obtained from the OEF and OEM metrics can be used to track and improve the energy matching. This can be done by adapting generation to the expected demand or adapting demand to supply, i.e. by saving some of the non-essential workload for periods with high on-site energy availability. But also by applying energy storage to conserve surplus generation.

Energy reuse: ERF

Data centers always have a heat surplus resulting from the conversion of electrical energy into heat within the IT equipment. This heat surplus can be reused in different ways depending on local circumstances, like heat demand near the data center. Though it might be difficult to quantify the amount of energy that is efficiently being reused, it does provide opportunities for improving energy efficiency. The metric to track the amount of energy reuse is the Energy Reuse Factor (ERF), defined as (Patterson, 2010):

$$ERF = \frac{Reuse\ energy}{Total\ energy}; 0 \leq ERF \leq 1$$

Some data centers have already taken measures to efficiently reuse waste heat by providing it to greenhouses or residential and commercial buildings. The best results have been achieved when this is done through an aquifer thermal storage system which helps to mitigate the effect of seasonal demand.

Water usage: WUE

Though PUE doesn't include water use in the total energy consumption, it is estimated that '4% of U.S. electricity demand is for the movement and treatment of water' (EPRI, 2002). To keep track of the impact this has on the ecology the Water Usage Effectiveness (WUE) is available (Green Grid, 2011). It is defined as:

$$WUE = \frac{\text{Water usage}}{\text{IT energy}} \left[\frac{L}{kWh} \right]; 0 \leq WUE$$

Alternatively, if information concerning the embodied energy of the water source is available it's also conceivable to add this embodied energy from the total water usage to the total facility energy use.

IT efficiency: ITEE & ITEU

The IT efficiency is a very important factor contributing to the total data center facility energy use, but it's also a very complicated contribution. Different data centers have different purposes like storage, calculation and networking, or a combination. This makes a comparison of efficiency difficult. For every type of function the efficiency of all the installed equipment can be compared to a standardized alternative. The average value that is found results in the IT Equipment Efficiency (ITEE) metric (Green IT council, 2012), defined as:

$$ITEE = \frac{W_{DC, \text{rated}} [\text{units of useful work}]}{P_{DC, \text{rated}} [W]}$$

$$0 \leq ITEE \leq 1$$

With $W_{DC, \text{rated}}$ being the capacity of the IT equipment multiplied by the standardized capacity per watt and $P_{DC, \text{rated}}$ is the rated power of the IT equipment. The capacity is subdivided in three categories: servers [GTOPS], storage [Gbyte] and networking [Gbps].

Also, important to monitor is the average IT load, as total energy efficiency is better for high IT utilization. This can be done using the IT Equipment Utilization (ITEU) metric (Green IT council, 2012), defined as:

$$ITEU = \frac{E_{IT}}{E_{IT, \text{rated}}}; 0 \leq ITEU \leq 1$$

Some data centers only provide the infrastructure and rent out floor space to customers, meaning the owners have no influence over the efficiency of the IT equipment installed, in this case ITEE and ITEU shouldn't be used to assess the data center's efficiency.

Discussion, conclusion & further research

The literature review has provided sufficient information to answer the research question:

'Does the broadly accepted PUE metric reflect the real energy performance of a data center?'

The merits of PUE for improving data center energy efficiency are clear, but the industry has come to a point where further improvements can only be found by assessing energy performance in a broader sense. It is concluded that the scope of the PUE metric is insufficient to reflect the real energy performance of a data center. Subjects that PUE doesn't touch upon, but should be included in energy performance assessment, are water usage, on-site renewable energy generation, energy recovery, IT equipment efficiency and carbon footprint. For these topics, respectively the WUE, OEF/OEM, ERF, ITEE & ITEU and CUE metrics can be used.

The literature review showed a large energy savings potential by using up-to-date IT equipment, as the IT equipment efficiency doubles each 2 years on average. A very important conclusion is that, at least in some cases, PUE values can be positively influenced by increasing the energy use of IT equipment as it does not track the actual meaningful work that's being done by the data center. This can also be achieved by shifting cooling loads from the HVAC system to the server fans. These practices can be prevented by using complementary metrics as proposed in this paper.

Also, the importance of tracking water usage was shown as, for example, 4% of the U.S. national energy consumption is connected to water treatment and transport. This becomes increasingly important because of the increased use of evaporative cooling systems.

Further research should show to what extent the issues raised in this paper influence the discrepancy between PUE values and real energy performance. This can aid data center designers in their decision-making process on energy efficiency measures and help the data center industry to take another step in reducing its, still increasing, energy impact. ■

Literature

Avelar, V., Azevedo, D., French, A., (2012), PUE™: *A Comprehensive Examination of the Metric*, The Green Grid.

Avram, M.G., (2014), Advantages and Challenges of Adopting Cloud Computing from an Enterprise Perspective, *Procedia Technology*, Volume 12, pp. 529-534.

Bouley, D. (2009), Understanding PUE, Data Center Science Center APC by Schneider Electric, accessed on April 1st 2016 https://www.bicsi.org/uploadedFiles/BICSI_Website/Global_Community/Presentations/Caribbean/APC%20Understanding%20PUE%20Denis.pdf.

Cao, S., Hasan, A., Siréna, K., (2013) "Onsite energy matching indices for buildings with energy conversion, storage and hybrid grid connections." *Energy and Buildings* 64, pp. 423–438.

Donnelly, C., (2015), found on October 15th 2015: <http://www.computerweekly.com/feature/Datacentre-PUE-Whos-keeping-score>.

EPRI, (2002), electricity consumption for water supply & treatment - the next half century, *Water & Sustainability* (Volume 4): U.S.; No. 1006787; Electric Power Research Institute: Palo Alto, California.

Foster, P., (2013), European data centers are less energy efficient than you might think, last accessed on February 19th 2016, <http://www.thegreenitreview.com/2013/05/european-data-centres-are-less-energy.html>.

Green IT promotion council, (2012), New Data Center Energy Efficiency Evaluation Index DPPE (Datacenter Performance per Energy) Measurement Guidelines, p. 37.

Google, (2016), Efficiency: How we do it, last accessed on February 16th 2016, <https://www.google.com/about/datacenters/efficiency/internal/>.

Belady, C., (2010), Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric, *The green grid*.

Guo, C. *et al.*, (2009), BCube: a high performance, server-centric network architecture for modular data centers, *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pp. 63-74.

Hartfield, L., (2011), Server room temperature myth busting – energy savings, disk failure and temperature, Accessed on February 28th 2016, <http://www.carbondescent.org.uk/blog/?p=164>.

Koomey, J.G., (2011), Growth in data center electricity use 2005 to 2010, *Analytical Press*, August 1, 2011.

Leopold, G., (2014), Data center spending up, energy efforts hitting wall, last accessed on October 12th 2015: <http://www.enterprisetech.com/2014/06/04/datacenter-spending-energy-efforts-hitting-wall/>.

Modoff, B., Bhagavath, V., Clifton, K., (2014), Not quite the Jetsons yet, but places to look *Internet of Things (IOT) is a big idea but promise and reality are a distance apart*, F.I.T.T. for investors, Deutsche Bank Securities, 6 May 2014.

Moore, G. E., (1965), Cramping more components onto integrated circuits, *Electronics*.

Patterson, M., (2010), ERE: A metric for measuring the benefit of reuse energy from a data center, p 9.

Research and Markets, (2015), Green Data Center Market - Global Industry Analysis, Size, Share, Growth, Trends, and Forecast 2014 – 2022, http://www.researchandmarkets.com/research/594p54/green_data_center.

Stansberry, M., (2015), 2014 data center industry survey, [accessed on 16-02-2016], <https://journal.uptimeinstitute.com/2014-data-center-industry-survey/>.

Uptime Institute (2014), found on October 15th 2015: <https://journal.uptimeinstitute.com/2014-data-center-industry-survey/>

Technavio, (2015), Global Data Center Market 2015-2019, accessed on October 15th 2015 <http://www.technavio.com/report/data-center-market>.

Wang, L., Khan, S.U., (2011), Review of performance metrics for green data centers: *a taxonomy study*, *J Supercomput*, Volume 63, pp. 639-656.

Wasson, S., (2015), Intel's Core i7-6700K 'Skylake' processor reviewed, last accessed on April 1st 2016, <http://techreport.com/review/28751/intel-core-i7-6700k-skylake-processor-reviewed/5>.

Whitney, J., Delforge, P., (2014), America's Data Centers Are Wasting Huge Amounts of Energy, *Critical Action Needed to Save Billions of Dollars and Kilowatts*, NRDC, August 2014. APPENDIX A - Measurement guidelines PUE.