The Water Footprint of Data Center Workloads: A Review of Key Determinants

Nuoa Lei
nuoalei@lbl.gov

Lawrence Berkeley National Laboratory, Berkeley, CA

Jun Lu
University of Illinois, Chicago

Arman Shehabi
ashehabi@lbl.gov

Lawrence Berkeley National Laboratory, Berkeley, CA

Eric Masanet
emasanet@ucsb.edu

University of California, Santa Barbara; Lawrence Berkeley National Laboratory, Berkeley, CA

Research Article

Keywords: Data centers, Water, Water footprint, Water-energy nexus, Water usage effectiveness

Posted Date: March 26th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4159702/v1

Additional Declarations: The authors declare no competing interests.

The Water Footprint of Data Center Workloads: A Review of Key Determinants

Nuoa Lei, Jun Lu, Arman Shehabi, Eric Masanet

Abstract

The importance of data center water use is growing. Operators are beginning to track and report WUE-site and WUE-source, but these metrics exclude potential efficiencies at the workload-hosting level. In this study, we review the determinants of overall water footprint, which we define as liters per workload ($\mathrm{liters/ssj\_ops}$) to allow for more direct comparisons across data center types. Using statistical analysis of recent server performance data, we first illustrate the relationship between server efficiency, utilization level, refresh rates, and management of inactive (i.e., redundant or comatose) servers within data centers as determinants of direct server energy use per workload. Next, we review the relationship between PUE and WUE-site using climate and technology-specific simulations for major data center cooling archetypes and their influence on facility-level energy and water demand per workload. Finally, we review the relationship between electrical grid water consumption factors (WCFs), WUE-source, and the overall water footprints per workload. Our analysis shows that there is no single recipe for minimizing data center water footprints; rather, we show that such minima can be achieved through different combinations of our reviewed determinants. Sensitivity analysis reveals that WCFs, server utilization level, and server efficiency are generally the strongest predictors of per-workload water footprints. Our review and analysis fill important knowledge gaps surrounding the determinants of water footprints and their practical minima under different site-specific data center constraints.

Keywords

Data centers; Water; Water footprint; Water-energy nexus; Water usage effectiveness.

Introduction

The increasing demand for digital services has led to an increased reliance on data centers, which are vital to the digital economy and provide the necessary functions of data processing, communication, and storage[1]. However, the operation of data centers requires significant amounts of electricity and water[2,3]. While data center electricity use has been the subject of much debate[3–7], data center water consumption is a less studied topic that is becoming increasingly important in the context of water scarcity and freshwater conservation[2,8,9]. On the one hand, data centers can consume a significant amount of water directly for cooling and space conditioning[8]. Google's data centers in Dalles consumed 355 million gallons of water in 2021, representing 29% of the city's total water consumption, which has raised concerns about local water stress, particularly in light of the recent drought conditions in this region[10]. On the other hand, data centers consume water indirectly through the water used for electricity generation[1,2,9]. This type of water use can be significant at the local or national level depending on the power generation technologies, power mixes, and electricity usage of the data centers. For example, some large data centers have a power capacity of 100 megawatts[11] and the generation of 1 kilowatt-hour of thermoelectricity or hydropower could consume 0.18-2.0 liters or 0.67-1194 liters of water[12], respectively. This demonstrates a significant amount of data center indirect water use and further emphasizes the importance of considering data center water sustainability going forward.

To align with the growing importance of water conservation and sustainability in data centers, operators are beginning to track, manage, and report their water usage effectiveness (WUE) at the site level (WUE-site) and at the source level (WUE-source)[13,14]. WUE-site measures the efficiency of onsite water consumption at a data center facility, expressed as the ratio of the data center's total onsite water use to its IT equipment electricity use, while WUE-source quantifies water use efficiency at the source level by considering the total water use (including on-site and off-site water) in relation to IT equipment electricity use[15]. Although WUE-site and WUE-source are popular sustainability metrics within the data center industry due to their simplicity, they are incomplete metrics that do not consider hardware efficiency or computing performance[16], which makes it difficult to compare the water footprint per digital service among data centers with different workload efficiencies. There is also a lack of systematic analysis on the determinants of data center water footprints in the research community, which makes it difficult for designers, engineers, operators, investors, and policymakers to effectively understand and address data center water usage issues. While the impact of cooling system and economizer selection on WUE-site is well understood[17–19], the influence of data center operating efficiency and climate conditions on WUE-site has only recently been investigated[8]. Additionally, the workload efficiency of servers in data centers is affected by various factors such as server utilization level, power efficiency, refresh rate, and inactive server percentage[1,3], which also contribute to the water footprint per digital service but are often overlooked by data center sustainability practitioners. Thus, a comprehensive understanding of the water footprint of data centers is necessary, which involves not only considering the determinants of WUE-site and WUE-source, but also the factors that affect the server workload efficiency.

In this study, we aim to fill the aforementioned knowledge gaps through a comprehensive analysis of the overall data center water footprint per workload. To make a fair comparison and gain a deeper understanding of the water footprint of data centers at the workload hosting level, we integrated the workload throughput score (i.e., ssj_ops) from the SPECpower_ssj 2008 benchmark into our analysis[20]. This allowed us to create a new metric for overall water footprint, expressed in liters per workload ($\text{liters}/\text{ssj\_ops}$), which addresses the limitations of the WUE-site and WUE-source metrics[16]. We followed up with a teardown analysis of the defined overall data center water footprint per workload, focusing on the analysis of the determinants of the water footprint and their relationships through the examination of real-world situations. In the teardown analysis, we examined three key aspects of the data center water footprint per workload: (1) the relationship between server workload efficiency and factors such as server utilization level, power efficiency, refresh cycles, and inactive server percentage; (2) the effect of cooling technology, climate, and infrastructure efficiency on PUE, WUE-site, and their relationship, and how these interact with server workload efficiency to further impact the relationship between facility energy and water use per workload; and (3) the determinants of WUE-source and how they interact with server workload efficiency to further influence the overall data center water footprint per workload. Our analysis confirmed that considering workload efficiency is crucial in comparing water footprints between data centers, and the measures identified in our analysis can be effective for optimizing data center water footprints. This is an important step towards achieving water sustainability in data centers.

Methods

Data center water footprint

The overall data center water footprint per workload ($liters/ssj\_ops$) is defined as Eq. (1).

$$WF_{overall} = WF_{site} + EF_{site} \times WCF \qquad (1)$$

where WF_{site} is the facility-level water use per workload ($liters/ssj\_ops$, Eq. (2)), EF_{site} is the facility-level energy use per workload (kWh/ssj_ops, Eq. (3)), and WCF is the electrical grid water consumption factor, defined as the amount of freshwater consumed (in liters) per kilowatt-hour of electricity generated[12] ($liters/kWh$).

$$WF_{site} = WUE_{site} \times e_{server} \qquad (2)$$

where WUE_{site} is the water usage effectiveness at the data center facility level ($liters/kWh$), e_{server} is the server energy use per workload (i.e., sever workload efficiency) (kWh/ssj_ops, Eq. (5)).

$$EF_{site} = PUE \times e_{server} \qquad (3)$$

where PUE is the power usage effectiveness, defined as the ratio of a data center's total electricity use to its IT equipment electricity use[21] (kWh/kWh).

Substituting Eq. (2) and (3) into Eq. (1) gives Eq. (4):

$$WF_{overall} = (WUE_{site} + PUE \times WCF) \times e_{server} = WUE_{source} \times e_{server} \quad (4)$$

where WUE_{source} is the source-based water usage effectiveness ($liters/kWh$).

Direct server energy use per workload

We focused on servers as the main component of IT loads because they are mainly responsible for workload hosting. The server energy use per workload at the data center level was modeled as a function of server utilization level, vintage distribution (or refresh cycle), server efficiency, and the fraction of inactive servers (including comatose servers and redundant servers operating at idle state), expressed as Eq. (5).

$$e_{server} = 3.6 \times \frac{\tilde{P}}{\widetilde{Th}} = 3.6 \times \frac{\varphi \bar{P}_i + (1-\varphi)\bar{P}_a}{(1-\varphi)\bar{P}_a \overline{Perf}} \quad (5)$$

where $\tilde{P}$ is the weighted-average power use of severs in data centers (W), and $\widetilde{Th}$ is the weighted-average server throughput (ssj_ops/s). $\bar{P}_i$ is the average power use of idle servers (W, Eq. (6)), $\bar{P}_a$ is the average power use of active servers (W, Eq. (7)), $\overline{Perf}$ is the average server power-performance (ssj_ops/W, Eq. (8)), and φ is the fraction of inactive servers in data centers.

$$\bar{P}_i = \sum_y \emptyset_y \times P(u = 0, y) \quad (6)$$

$$\bar{P}_a = \sum_y \emptyset_y \times P(u, y) \quad (7)$$

$$\overline{Perf} = \sum_y \emptyset_y \times Perf(u, y) \quad (8)$$

where $\emptyset_y$ is the server vintage distribution, represented by the percentage of installed servers of year y in data centers. $P(\cdot)$ and $Perf(\cdot)$ are respectively the server power and power-performance model, which are formulated as functions of the server utilization level (u) and the server installed year (y) using the local linear kernel regression approach based on the SPECpower_ssj 2008 benchmark database[20]. Due to variability in server efficiency each year y, we analyzed $P(\cdot)$ and $Perf(\cdot)$ at different power efficiency quantiles for the pool of servers reported by the SPECpower_ssj 2008 database.

Climate and technology-specific PUE and WUE-site

The PUE and WUE-site values of data centers are simultaneously influenced by cooling system technologies, climate conditions, and operating efficiency practices[8,22]. We conducted an in-depth review of the simulation methodology and results for the PUE and WUE-site reported by Lei et al[8], complemented by additional simulations of the cooling system and climate scenarios, aiming at clarifying the interrelationships between PUE and WUE-site values and further illustrate how their determinants affect data center water footprint. To thoroughly quantify the impact of the various determinants, we covered typical data center cooling technologies and operating conditions[8,22,23], as well as major types of global climate zones specified by American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)[24]. In total, we considered eight cooling technologies including (1) air-cooled chiller with no economizer, (2) airside economizer using air-cooled chiller as supplement cooling, (3) airside economizer using water-cooled chiller as supplement cooling, (4) airside economizer and direct adiabatic cooling using air-cooled chiller as supplement cooling, (5) airside economizer and direct adiabatic cooling using water-cooled chiller as supplement cooling, (6) direct expansion system, (7) water-cooled chiller with no economizer, (8) waterside economizer using water-cooled chiller as supplement cooling. Additionally, we considered data center operations under nineteen different climate zones including extremely hot humid (0A), extremely hot dry (0B), very hot humid (1A), very hot dry (1B), hot humid (2A), hot dry (2B), warm humid (3A), warm dry (3B), warm marine (3C), mixed humid (4A), mixed dry (4B), mixed marine (4C), cool humid (5A), cool dry (5B), cool marine (5C), cool humid (6A), cold dry (6B), very cold (7), subarctic/arctic (8). Particularly, we took different data center operating efficiency practices into considerations[7,8,22], which are represented by different quantiles of PUE and WUE-site simulation values for data centers implementing a given cooling technology in a given climate zone.

Results

Determinants of direct server energy use per workload

To gain a comprehensive understanding of the water footprint of data center workloads, we performed a teardown analysis, starting at analyzing the factors that contribute to the server electricity use per data center workload. Fig. 1 shows the impact of various factors on server electricity use per workload, obtained using Eq. (5) and assuming that the newest servers in the data center were purchased in 2020. The result shows that the lowest energy use per workload is generally associated with high server utilization levels, efficient servers, fast refresh cycles, and low fractions of inactive servers, but none of these factors alone is usually sufficient to justify low direct energy use per workload, as indicated by how different combinations of these factors result in the same color (i.e., the same server electricity per workload) in Fig. 1. Server energy use per workload is mainly affected by the server utilization level, where the electricity use per workload decreases with the increase of the utilization level, but the decline rate also decreases when the server utilization level exceeds 40% (Supplementary information (SI) Fig. S1 and S2). Using high-efficient servers (servers whose power efficiencies locate at higher quantiles each year) is another strategy to reduce direct energy use per workload in data centers. Efficient servers consume less power at idle state, and typically have larger dynamic ranges and better power proportionality[25,26] (SI Fig. S4), leading to lower energy use per workload. Optimizing the server refresh cycle reduces energy use per workload in a manner similar to using energy efficient servers because servers become more energy efficient overtime (SI Fig. S1 and S4). Likewise, removing inactive servers from the data center can reduce energy use per workload (SI Fig. S5), since servers consume a certain amount of power but are not performing any computing tasks at idle, albeit that newer generation of servers have significantly lower idle power.

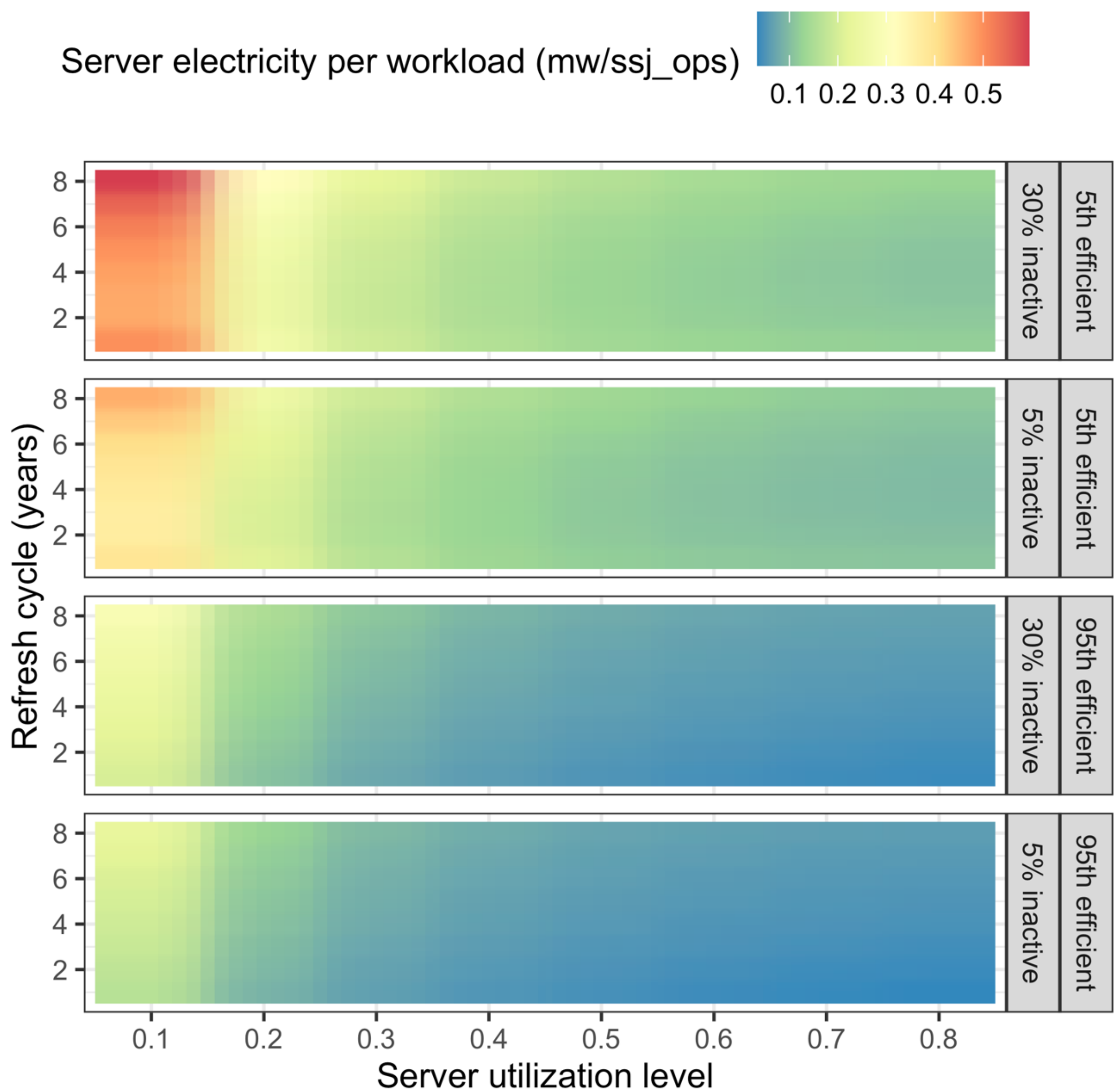

Fig. 1. Determinates of server electricity use per workload.
5% inactive ~ 5% percentage of inactive servers; 30% inactive ~ 30% percentage of inactive servers; 5^{th} efficient ~ server power efficiency at 5^{th} quantile; 95^{th} efficient ~ server power efficiency at 95^{th} quantile.

Determinants of facility-level water use per workload

We further analyzed facility-level water use per workload as it established a baseline for overall data center water footprint per workload (Eq. (1)). Fig. 2 shows the relationship between WUE-site, server energy use per workload, and facility-level water use per workload. For a given server electricity use per workload, facility water use per workload increases with the increase of WUE-site, and server energy use per workload determines the rate of this increase. Notably, a low facility water footprint per workload is not necessarily a result of a low WUE-site, which is indicated by points A and B in Fig. 2, in which both points have the same facility water use per workload but have different values of WUE-site and server energy use per workload, demonstrating the necessity of considering both WUE-site and server workload efficiency in the design of water-efficient infrastructure.

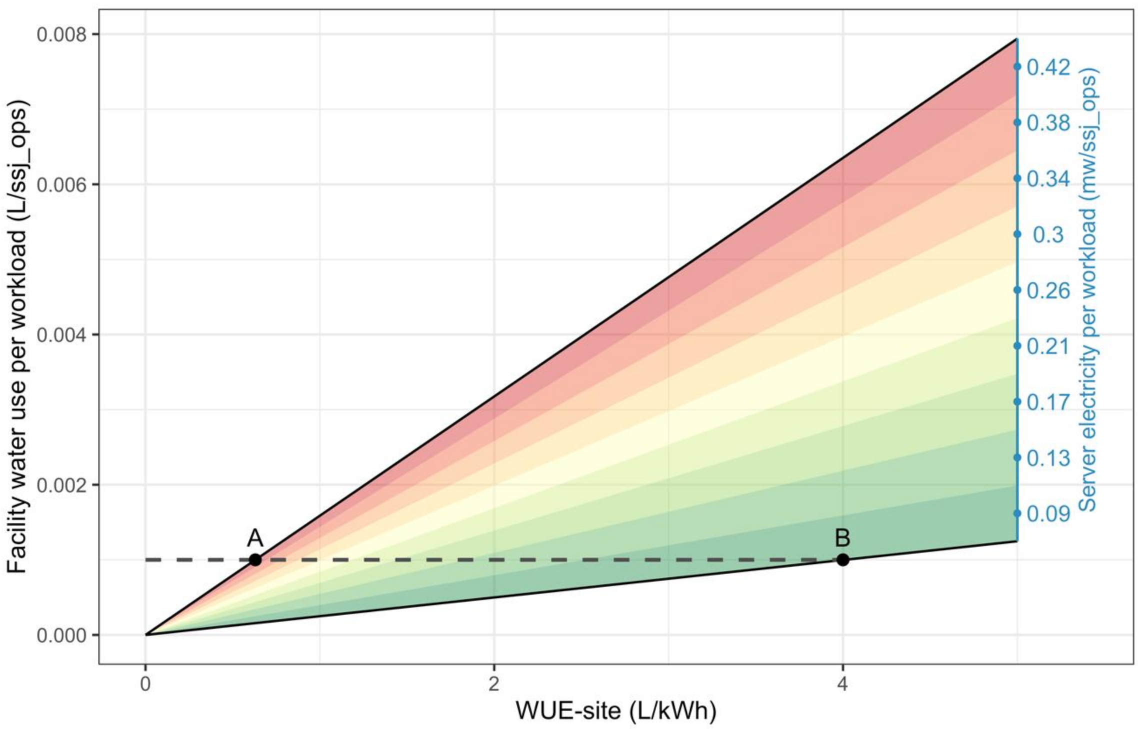

Fig. 2: Determinants of facility-level water use per workload.

Relationship between PUE and WUE-site values

PUE and WUE-site values are two widely used sustainability metrics to quantify the energy and water use efficiency of data center operations[15,21]. They are intricately related and depend on various factors including climate, cooling technology, and operational efficiency practices[7,8,22]. However, they are sporadically reported by operators and are poorly understood by sustainability analysts[2]. Given their importance in determining the overall data center water footprint (Eq. (4)), we conducted a comprehensive and first-of-its-kind review of their relationship to provide readers with a clearer view of PUE and WUE-site. Fig 3 shows the PUE and WUE-site values clustered by eight different cooling technologies, where the PUE and WUE-site variation within each cluster is driven by the effects of climate and data center infrastructure operating efficiency (SI Fig. S7). It can be clearly observed that PUE and WUE-site values vary significantly by cooling technologies. Data centers that employ water-cooled chillers (as primary or supplemental cooling) have relatively higher WUE-site values due to large water consumed at cooling towers, while data centers using direct-expansion systems or air-cooled chillers (as primary or supplemental cooling) have relatively lower WUE-site values, but typically have higher PUE values due to lower chiller coefficient of performance (COP)[8]. Additionally, data centers that employ economizers or adiabatic cooling systems can reduce both PUE and WUE-site values, with the greatest energy and water savings achieved for data centers using airside economizers and direct adiabatic cooling. Within each cooling technology cluster, we noticed a non-linear and positive relationship between PUE and WUE-site, where favorable climates and high operational efficiencies result in low PUE and WUE-site values, and hot and/or humid climates and low operational efficiencies lead to high PUE and WUE-site values (SI Fig. S7). Finally, we noted that similar PUE and WUE-site values can be achieved with different combinations of cooling technology, climate, and efficiency, represented by overlaps between clusters, which highlights the importance of system-wide optimization that simultaneously considering technology, climate, and efficiency in PUE and WUE-site reduction.

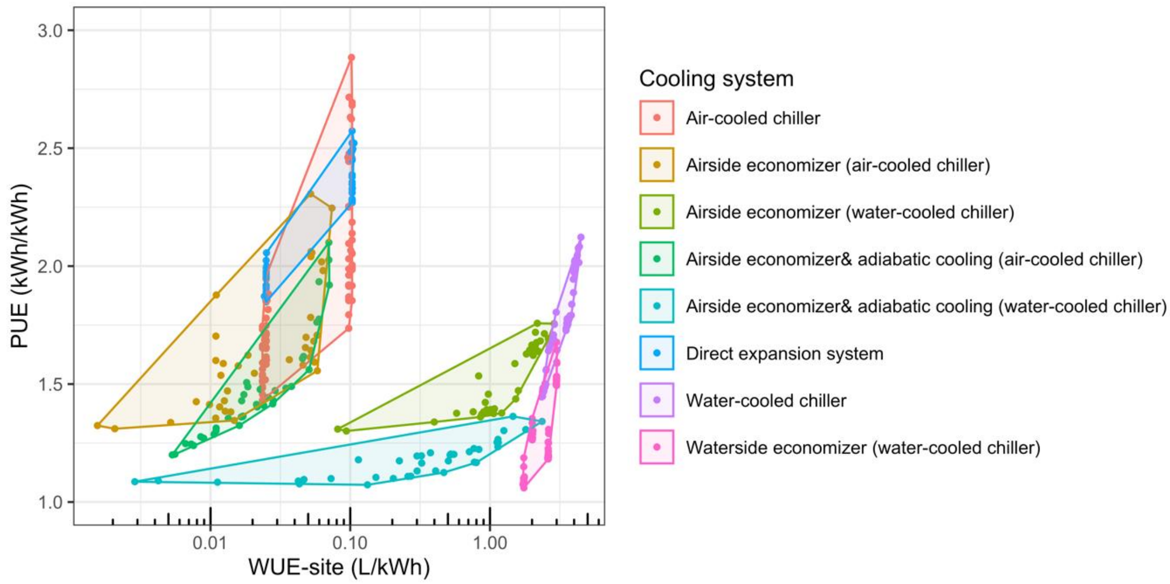

Fig. 3: Relationship between PUE, WUE-site, and cooling technology (X values are logarithmically spaced).
See SI Fig. S6 (where X values are linearly spaced); Scatters within a given cooling technology consist of PUE and WUE-site values of different climate zones and data center operating efficiency practices (SI Fig. S7); air-/water-cooled chillers in the brackets stand for supplemental cooling sources; PUE and WUE-site are yearly average values.

Relationship between facility-level energy and water use per workload

Not only PUE and WUE-site are complexly related, facility-level energy and water use per workload also exhibit a delicate relationship due to the multiplicative effect of server energy use per workload (Eq. (2) and (3)). For a clear illustration, we first visualize the relationship between facility energy and water use per workload for data centers using airside economizers and adiabatic cooling and using water-cooled chillers as auxiliary cooling systems, which result in a region in Fig. 4 (a) that is bounded left to right by the product of workload efficiency and possible WUE-site values, and top to bottom by the product of server workload efficiency and possible PUE values.

Fig. 4 (b) depicts the relationship between facility-level energy and water use per workload for different data center cooling systems. Our analysis showed that the lowest data center facility-level energy and water use per workload are generally associated with the highest server workload efficiencies, but the achievable minima (bottom left points) vary significantly by cooling system archetype. The results also indicate a tradeoff between facility-level energy and water use per workload when selecting data center cooling technologies. That is, data centers with air-cooled chillers and direct expansion systems typically use less water but more energy per workload, whereas data centers using water-cooled chillers consume less energy but more water per workload. The implementation of economizers or adiabatic cooling in data centers can simultaneously reduce facility-level energy and water use per workload, and adiabatic cooling, in particular, has been found to significantly reduce energy per workload in these settings. However, different combinations of determinants can lead to same energy and/or water use per workload (e.g., in Fig 4. (b), the points marked A have the same facility-level water use per workload, and the points marked B have the same facility-level energy use per workload), which demonstrates that energy and water use per workload cannot be simply assessed or compared using a single determinant and emphasizes that there is no single way to reduce energy and/or water use per workload in data centers.

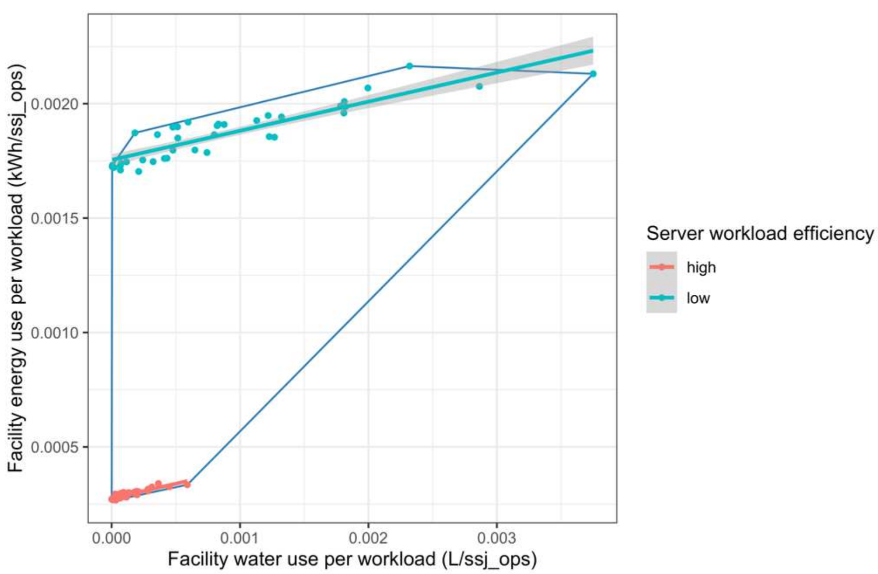

Fig. 4: (a) Relationship between facility energy and water use per workload in data centers with airside economizer and adiabatic cooling (water-cooled chiller) systems (Low and high server workload efficiency represents minimum and maximum server electricity per workload in Fig. 1).

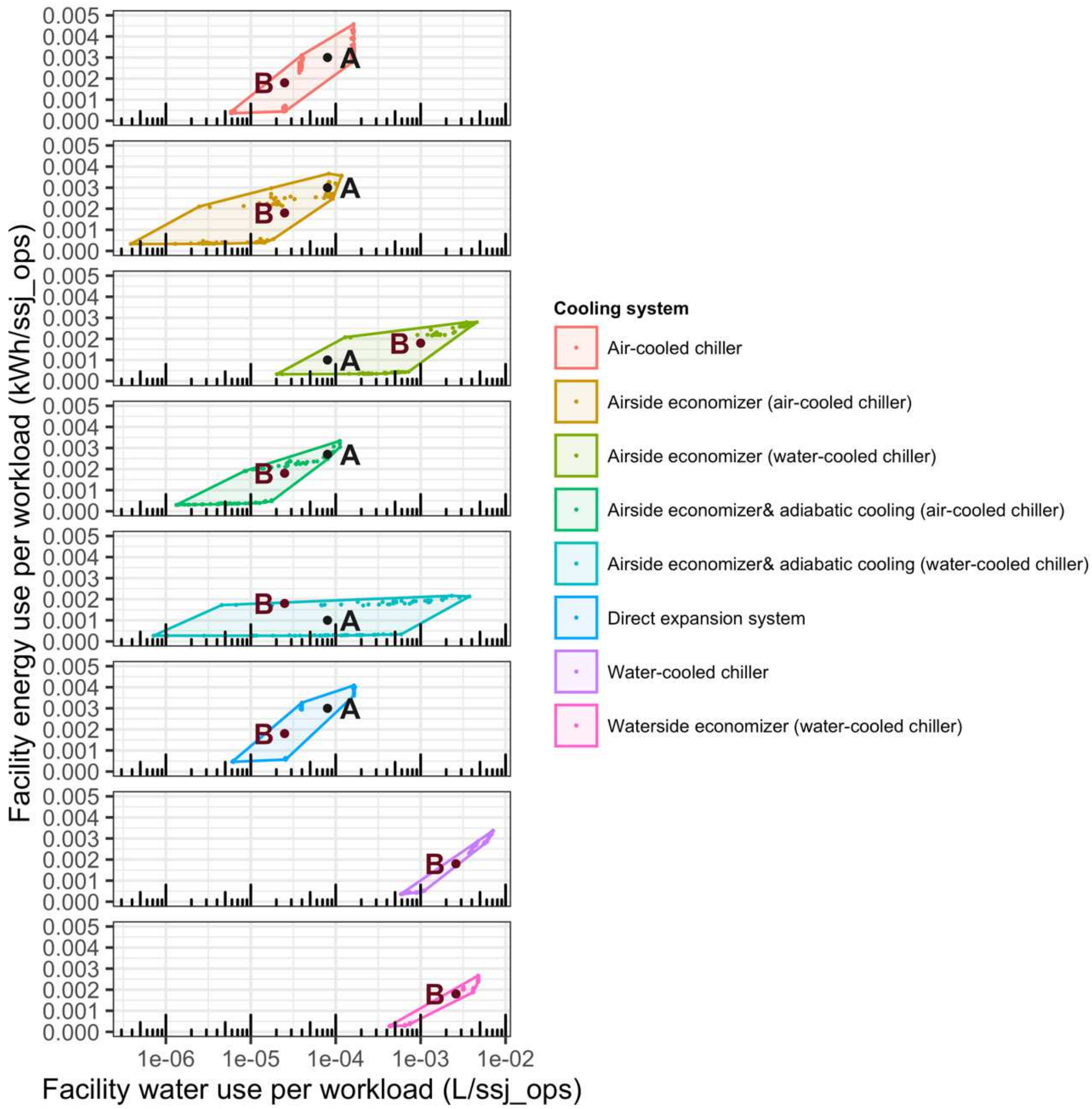

Fig. 4: (b) Relationship between facility energy and water use per workload by data center cooling system (Points A have the same facility water use per workload; Points B have the same facility energy use per workload).

Determinates of WUE-source

WUE-source takes into account the on-site water consumption for space-conditioning and/or cooling tower operations, and the off-site water consumption at power generation sites. It is currently used as a proxy to quantify the water footprint of a data center[9,15] and is a key component in assessing the overall data center water footprint per workload (Eq. (4)). To clearly illustrate determinates of WUE-source, we first compared the relationship between WUE-source and WUE-site for two cooling systems (airside economizer and adiabatic cooling (water-cooled chiller) and direct expansion systems) and two power sources (wind and hydro), which results in two regions in Fig. 5 (a). Regions for each cooling technology are bounded from left to right by the range of WUE-site values, where the region width corresponding to the airside economizer and adiabatic cooling (water-cooled chiller) is larger than that corresponding to the direct expansion system due to greater variability in WUE-site values (Fig. 3). On the other hand, for a given a cooling technology and a given WUE-site value, the distance between the lower and upper bound of the region is determined by the product of PUE value and WCF, where the lower bound and upper bound is respectively associated with the power source of the lowest WCF (wind) and the highest WCF (hydro). And for data centers that rely on other/mixed power sources (Fig. 5 (b)), their WUE-source and WUE-site values generally form a line between the upper and lower boundaries (SI Fig. S8). Importantly, large gaps between the lower and upper boundaries of these regions highlights the huge influence of WCF in determining WUE-source, where data centers with high WUE-site values can achieve low WUE-source values by utilizing wind or solar power, while data centers with low WUE-site values can have high WUE-source values due to the use of hydro or fossil fuel based electricity.

Fig. 5 (c) depicts the relationship between WUE-site and WUE-source for different data center cooling technologies. Due to differences in PUE and WUE values for different cooling technologies, the area and shape of the regions vary widely in Fig. 5 (c). And depending on the product of the PUE value and WCF, data centers with the same WUE-site value can have very different WUE-source values. The same WUE-source value can be achieved by combining various determinants (e.g., WUE-source values between 2.5 and 12.5 $liters/kWh$ can be achieved for all data center cooling technologies we considered), which highlights two main approaches for data center WUE-source minimization to address local water issues and water conservation concerns. First, WUE-source can be reduced by improving data center operational efficiency, which can reduce both PUE and WUE-site, but the level of this reduction is limited by cooling technology implementation and data center climate conditions. Second, WUE-source can be reduced through choice of power sources with low WCFs, which is a more general approach that is not constrained by the cooling technology implementation, but rather the availability of electric grid options.

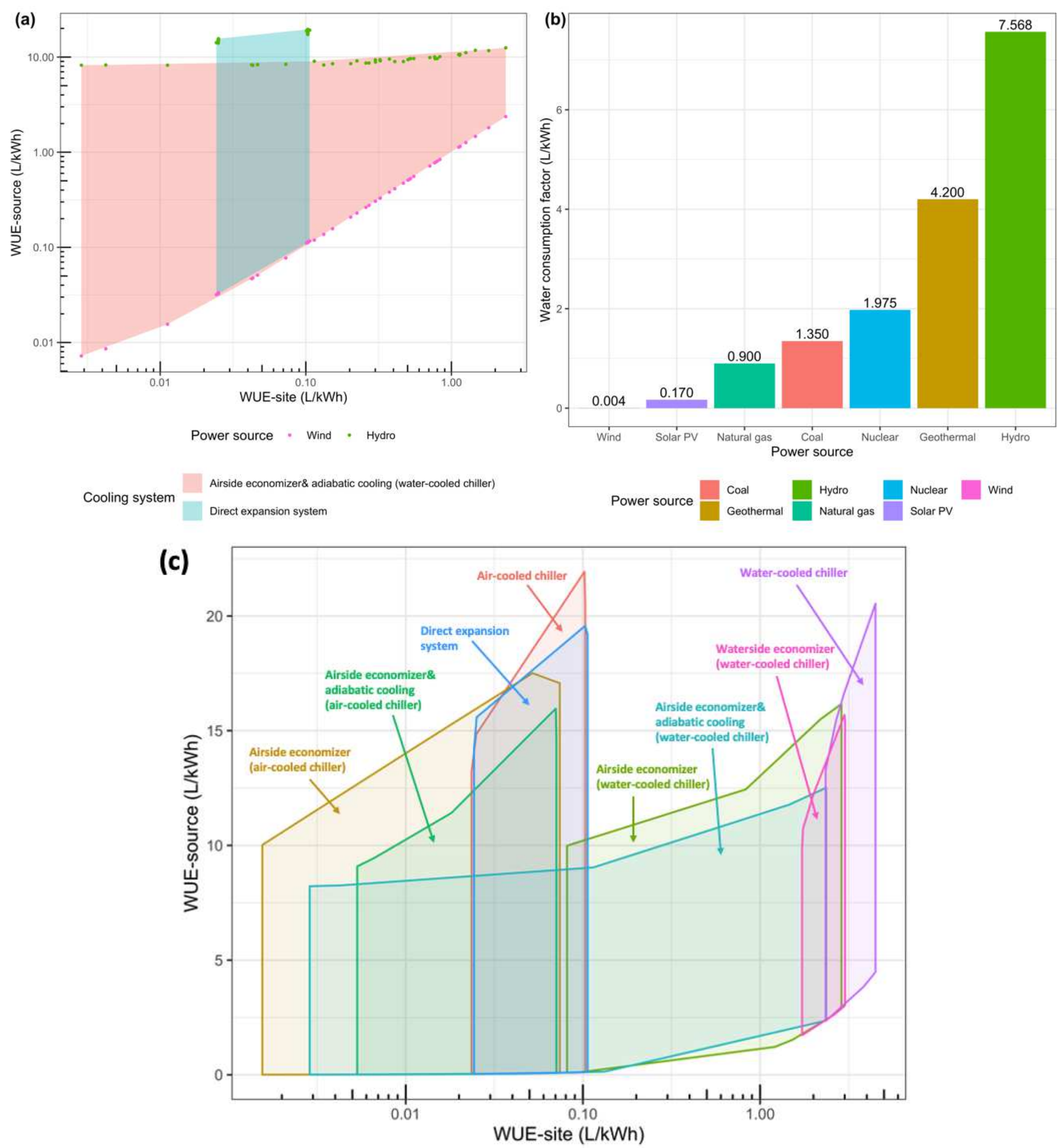

Fig. 5: (a) Relationship between WUE-source and WUE-site in data centers with airside economizer and adiabatic cooling (water-cooled chiller) or direct expansion systems (see SI Fig. S8, where X values are linearly spaced); (b) Water consumption factor by power source; (c) Relationship between WUE-source and WUE-site by cooling technology.

Determinates of overall water footprint per workload

The overall water footprint per workload considers data center water use efficiency at the workload hosting level, which is influenced by factors that affecting direct server electricity use per workload and WUE-source (Eq. (4)). For simplicity, we analyzed data centers that rely on wind or hydro power, corresponding to the power sources with the lowest and highest WCF in Fig 5 (b) respectively. Fig. 6 (a) shows the relationship between overall and facility-level water use per workload for data centers that using airside economizers and using air-cooled chillers as supplemental cooling equipment. Notably, the upper and lower bounds of the regions in Fig 6 (a) respectively correspond to low and high server workload efficiencies, and the relative distance between them represent the relative importance of server workload efficiency to the overall water footprint per workload in data centers. Specifically, the relative distance of the lower and upper bound in Fig 6 (a) is larger for data centers that use hydro power than those use wind power, indicating the significance of reducing the overall water footprint by improving server workload efficiency when a data center is constrained by power sources with high WCFs. Conversely, the relative distance between the lower and upper bound in Fig 6 (a) is small for data centers that use wind power, highlighting the importance of reducing the overall water footprint by reducing facility-level water use per workload in data centers that rely on low WCF power source.

Fig 6 (b) further shows the relationship between the facility-level and the overall water use per workload for data centers with different cooling systems. Data centers using wind power typically have a much lower water footprint than those using hydropower, unless the data center has a significant on-site water consumption (e.g., the data center relies on water-cooled chillers or waterside economizers for cooling), which reconfirms the effectiveness of using low WCF electricity to reduce the water footprint of data centers. Additionally, the facility-level and overall water use per workload become more positively correlated as data centers transition from hydro to wind power, suggesting that more action should be taken to reduce facility-level water footprint (either through cooling system selection, PUE or WUE-site reduction, or server workload efficiency optimization) when low-WCF power is already available. Nevertheless, we would like to emphasize that different combinations of determinates (WCF, cooling system type, and efficiency, etc.) can lead to similar overall water footprints per workload, so water footprints between data centers cannot be simply compared or assessed. To shed light on important and effective strategies for optimizing the overall water footprint per workload in data centers, we performed a robust global sensitivity analysis in the following section.

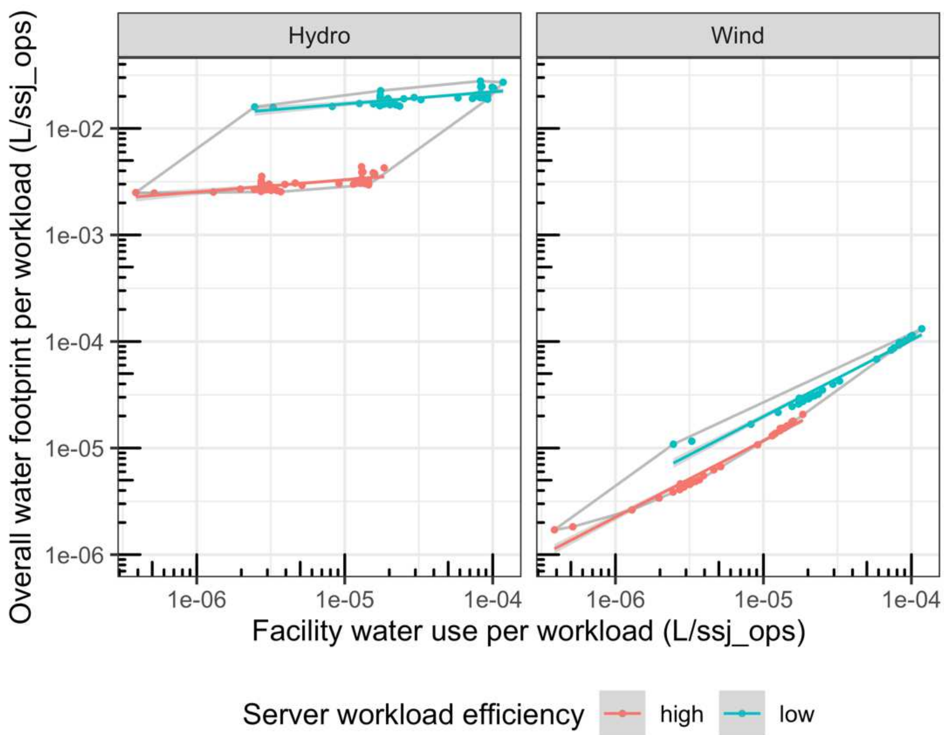

Fig. 6 (a): Relationship between data center facility-level and overall water use per workload for data centers with airside economizers (air-cooled chillers) and rely on hydro/wind power. Low and high server workload efficiency represents minimum and maximum server electricity per workload in Fig. 1.

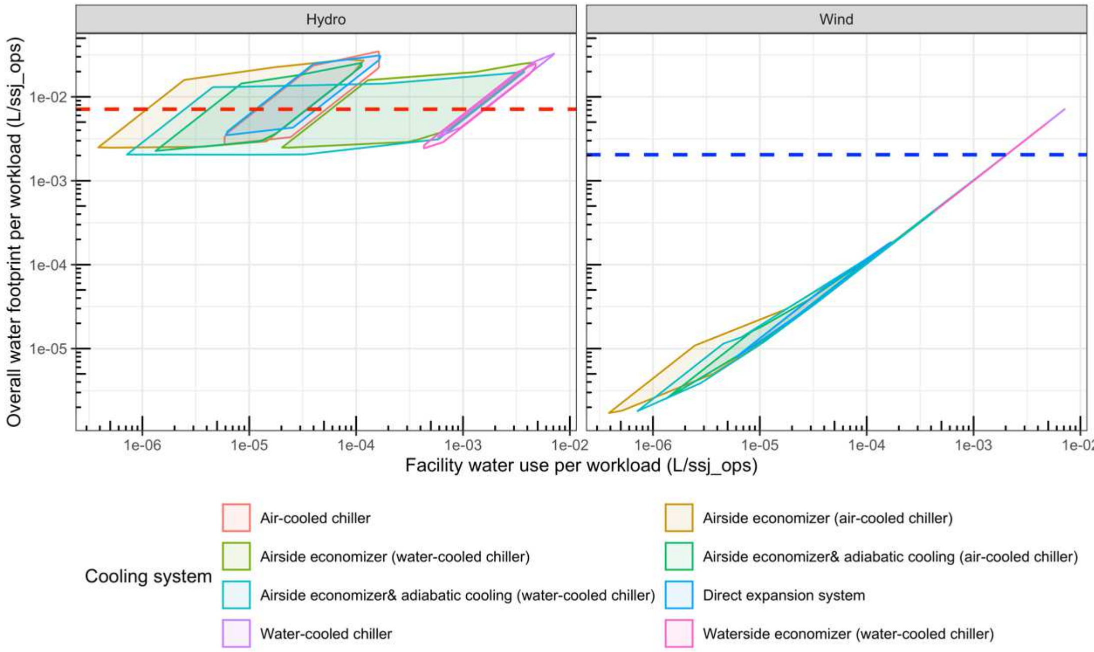

Fig. 6 (b): Relationship between data center facility-level and overall water use per workload by cooling technology for data centers that rely on hydro/wind power (see SI Fig. S9 for data centers that rely on other power sources). Blue dashed line ~ minimum overall water footprint per workload for data centers using hydro power; Red dashed line ~ maximum overall water footprint per workload for data centers using wind power.

Sensitivity analysis

Fig 7 shows the sensitivity results based on Sobol's total effect index[27], which ranks the relative importance of the determinants of the overall water footprint per workload via variance decomposition. The WCF is the largest contributor to the variance in the result of the water footprint per workload, suggesting that powering the data center with low-WCF electricity such as wind or solar photovoltaics (PV) can be effective in reducing the data center water footprint. Server utilization level and server efficiency quantile are the second and third largest contributors to variance in the water footprint result, as they have large impact on direct server electricity use per workload, which signifies that the data center water footprint can also be reduced through server workload management (e.g., consolidation and virtualization) or installing energy-efficient servers (e.g., Energy Star certified servers). Cooling system type, data center infrastructure operating efficiency, and climatic conditions are important factors in determining data center PUE and WUE-site, making them nontrivial factors in data center water footprint optimization. Server fresh cycle and inactive server percentage also affect the water footprint per workload through affecting server workload efficiency, further confirming the usefulness of server management through equipment refurbishment or removal of needless equipment.

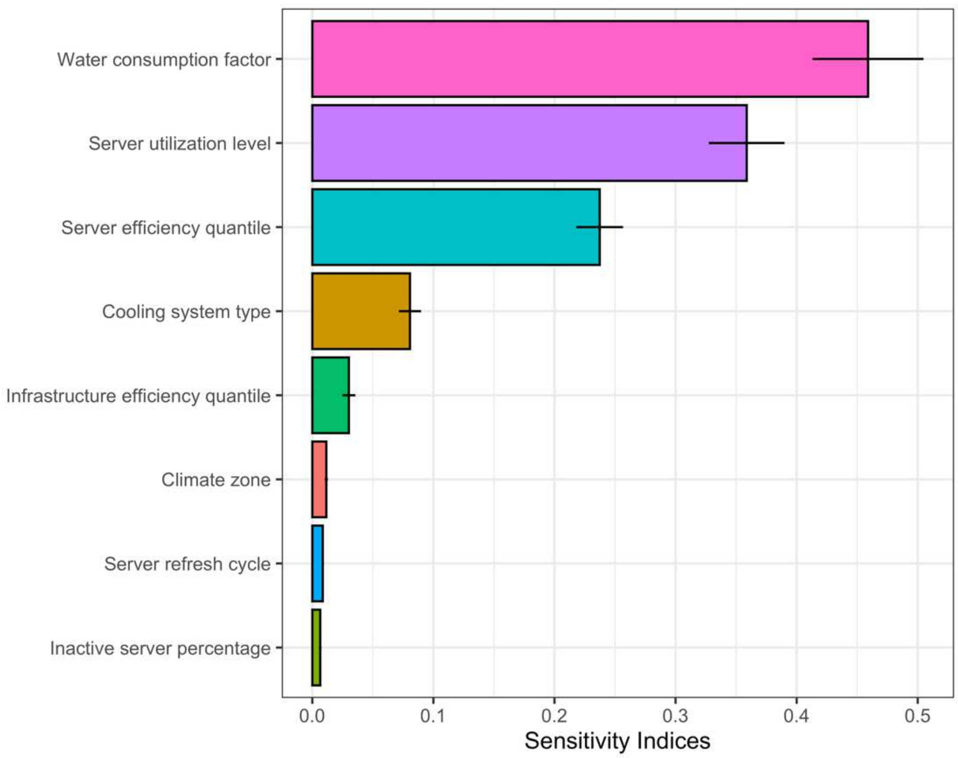

Fig. 7: Importance ranking (based on Sobol's total-effect index) for the determinates of the overall water footprint per workload (Error bar ~ 95% confidence interval of sensitivity indices).
Server efficiency quantile ~ measures the server power efficiency for a given year; infrastructure efficiency quantile ~ measures the operating efficiency of the data center infrastructure (i.e., PUE and WUE-site) for a given cooling technology under a given climate zone.

Discussion

Increasing digitalization and freshwater restrictions have brought data center water conservation into the spotlight. Our factor analysis illuminates strategies that can be used to optimize the data center water footprint per workload, which is becoming increasingly important as policymakers and operators begin to focus on the direct and indirect water consumption of digital services. Our findings could help policymakers design common measures to promote water conservation in data centers. For example, promoting renewable energy (wind or solar PV) or computing equipment with the Energy Star label can effectively reduce water footprints at the national level. Our findings also highlight that, despite various technical measures, there is no “silver bullet” for addressing water use in all types of data centers, so strategies should be developed on a case-by-case basis. For instance, if the data center is geographically constrained by the electric grid or climate, increasing server workload efficiency is the most effective strategy to reduce water footprint. Furthermore, judicious selection of cooling technologies during the design phase of a data center has great potential to reduce a data center’s water footprint, which can lessen competition for local water resources from data center on-site water consumption in the meantime. Collectively, our analysis provides detailed evidence of the potential to reduce the water footprint of data centers through renewable energy utilization, server workload efficiency optimization, and data center infrastructure efficiency optimization, which can guide policymakers and operators in designing specific polices or measures to advance the water sustainability in data centers.

Our analysis also illustrates the value of evaluating water sustainability at the workload hosting level in data centers, which brings to light the necessity of considering the computing power efficiency of IT equipment. This is an aspect that is frequently overlooked by sustainability analysts and policymakers, and our analysis enables impartial comparisons of energy and water consumption among data centers by proposing several workload-based indicators, including direct server energy use per workload, facility-level water use per workload, and overall water footprint per workload. For example, a cloud data center that utilizes airside economizers and adiabatic cooling, in addition to using water-cooled chillers for supplementary cooling, may consume more water on-site per kilowatt-hour of electricity used compared to a traditional data center with a direct-expansion system. However, such a comparison may result in a biased view that the traditional data center is more water-efficient by ignoring the number of digital services that can be delivered for the same amount of on-site and off-site water use. Our analysis fills the gap in the current understanding of water footprints in data centers by providing examples of different scenarios that result in the same overall water footprint per workload. Our method can be useful for sustainability analysts and policymakers who want to make fair comparisons of water footprints between data centers.

While the primary focus of this study is the water footprint of data centers, the conservation strategies we suggest also have the potential to impact other aspects of data center operations. Data center workload management through virtualization allows multiple applications to run on a single server, potentially reducing the required IT capacity and saving initial investment costs through proper capacity planning. Improving the efficiency of IT equipment and infrastructure can reduce the electricity and water usage per data service, leading to cost savings (i.e., electric and water bills) in data center operations. Increasing energy efficiency and using renewable energy sources like wind or solar power can also reduce carbon emissions and contribute to climate change mitigation efforts, which are crucial in the current global warming context. To summarize, reducing the water footprint of data centers has many co-benefits such as reduced resource usage, reduced electric and water costs, and reduced carbon emissions. This provides strong evidence for policy incentives and operational measures to further improve the sustainability of data centers.

There are several limitations to consider in this analysis. Because of limited information on server computational efficiency, we rely on the SPECpower_ssj 2008 benchmark database, which may not accurately reflect the distribution of server computational efficiency in practical data center operations over time. There are concerns that even the least efficient servers in the SPECpower_ssj 2008 benchmark database may still be more efficient than the typical volume servers on the market. This could result in an underestimation of the upper bounds of water footprints in our analysis, leading to a higher probability that different data centers may have the same water footprint due to different combinations of water use determinants (e.g., cooling system, power source) that we analyzed. Our study primarily examines servers as the main component of IT loads, as they are the primary hosts for workloads. However, future research could consider a broader range of IT loads, including graphics processing units (GPUs), storage, and network loads. Additionally, the water footprint of data centers with other types of cooling systems (e.g., liquid cooling archetypes) has not been thoroughly studied. While liquid cooling may not have a major impact on reducing on-site water consumption without significant technological advancements in data center humidification and cooling tower operations, the potential for indirect water reductions through energy savings is a topic that deserves further examination. It is important to further emphasize that the workload-level water footprint does not take into account absolute water use, which is crucial for understanding the potential effect that data centers can have on local water resources. We recommend that decision makers consider using this metric in combination with total site water use in the future.

Conclusion

In this study, we performed a teardown analysis to understand the overall water footprint per workload in data centers. Our analysis identified the key factors that impact the water footprint of data centers and examined their relationships using real data and real-world scenarios, which is essential for effectively pursuing data center water sustainability in practical, real-world settings.

We conducted a thorough analysis of how data center server workload efficiency is affected by utilization level, refresh cycle, power efficiency, and inactive server fraction. We also examined the relationship between PUE and WUE-site with cooling technologies, climate zones, and facility operating efficiencies, providing further insights into the relationship between facility-level energy and water use per workload. Additionally, we examined the impact of the water consumption factor (WCF) of different power sources on WUE-source and the overall water footprint per workload. Finally, we carried out a robust global sensitivity analysis to quantify the relative importance of these parameters on the overall water footprint of data centers. Our findings indicate that the overall water footprint per workload can generally be reduced through increased server workload efficiency, increased data center site-level energy and water use efficiency (i.e., PUE and WUE-site), and increased renewable energy use.

Our analysis addresses important knowledge gaps in quantifying the overall water footprint per workload, which is essential for making unbiased comparisons of water efficiency between different data centers and serves as a foundation for understanding the water use of digital services. The determinants analysis that we conducted in this research emphasizes the various measures that can be exploited and tailored by policymakers and operators to achieve data center water sustainability goals, which is becoming increasingly important in the context of conserving water resources and mitigating water stress.

Acknowledgements

The authors express gratitude for the support from Leslie and Mac McQuown, which contributed to part of the research conducted at Northwestern University.

The authors gratefully acknowledge support from the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Industrial Efficiency and Decarbonization Office. Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy.

Supplementary information

Supplementary information related to this article can be accessed in the online version.

Data and code availability

The data and code used in this research are available on request from the corresponding author.

References

1. Shehabi, A. *et al.* United States Data Center Energy Usage Report. *Berkeley Lab* 65 (2016).
2. Mytton, D. Data centre water consumption. *npj Clean Water* vol. 4 1–6 (2021).

3. Masanet, E., Shehabi, A., Lei, N., Smith, S. & Koomey, J. Recalibrating global data center energy-use estimates. *Science (80-.).* **367**, 984–986 (2020).

4. Andrae, A. S. G. & Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges 2015, Vol. 6, Pages 117-157* **6**, 117–157 (2015).

5. Jones, N. How to stop data centres from gobbling up the world's electricity. *Nature* **561**, 163–166 (2018).

6. Belkhir, L. & Elmeligi, A. Assessing ICT global emissions footprint: Trends to 2040 & recommendations. *J. Clean. Prod.* **177**, 448–463 (2018).

7. Lei, N. & Masanet, E. R. GLOBAL DATA CENTER ENERGY DEMAND AND STRATEGIES TO CONSERVE ENERGY. *Data Cent. Handb. Plan, Des. Build, Oper. a Smart Data Cent.* 15–26 (2021) doi:10.1002/9781119597537.CH2.

8. Lei, N. & Masanet, E. Climate- and technology-specific PUE and WUE estimations for U.S. data centers using a hybrid statistical and thermodynamics-based approach. *Resour. Conserv. Recycl.* **182**, 106323 (2022).

9. Siddik, M. A. B., Shehabi, A. & Marston, L. T. The environmental footprint of data centers in the United States. *Environ. Res. Lett.* (2021) doi:10.1088/1748-9326/abfba1.

10. Oregon Tech. Google's water use is soaring in The Dalles, records show, with two more data centers to come. *oregonlive* https://www.oregonlive.com/silicon-forest/2022/12/googles-water-use-is-soaring-in-the-dalles-records-show-with-two-more-data-centers-to-come.html (2022).

11. Masanet, E. & Lei, N. How Much Energy Do Data Centers Really Use? *Aspen Glob. Chang. Inst.* 2021 (2020).

12. Lee, U., Han, J., Elgowainy, A. & Wang, M. Regional water consumption for hydro and thermal electricity generation in the United States. *Appl. Energy* **210**, 661–672 (2018).

13. Facebook. Open sourcing PUE/WUE dashboards. *Facebook Engineering* https://engineering.fb.com/2014/03/14/data-center-engineering/open-sourcing-pue-wue-dashboards/ (2014).

14. Scaleway. Environmental leadership. https://www.scaleway.com/en/environmental-leadership/ (2021).

15. Patterson, M., Azevedo, D., Belady, C. & Pouchet, J. Water usage effectiveness (WUE): a Green Grid data center sustainability metric. *White Pap.* **35**, (2011).

16. Horner, N. & Azevedo, I. Power usage effectiveness in data centers: overloaded and underachieving. *Electr. J.* **29**, 61–69 (2016).

17. Sharma, R., Shah, A., Bash, C., Christian, T. & Patel, C. Water efficiency management in datacenters: Metrics and methodology. in *2009 IEEE International Symposium on Sustainable Systems and Technology, ISSST '09 in Cooperation with 2009 IEEE International*

Symposium on Technology and Society, ISTAS (2009). doi:10.1109/ISSST.2009.5156773.

18. Gozcu, O., Ozada, B., Carfi, M. U. & Erden, H. S. Worldwide energy analysis of major free cooling methods for data centers. in *Proceedings of the 16th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, ITherm 2017* (2017). doi:10.1109/ITHERM.2017.7992592.
19. Sharma, R., Shah, A. J., Bash, C., Christian, T. & Patel, C. Water Efficiency Management in Datacenters (Part I): Introducing a water usage metric based on available energy consumption. *TR HP Labs* (2008).
20. Standard Performance Evaluation Corporation. SPECpower_ssj® 2008. https://www.spec.org/power_ssj2008/ (2022).
21. Jaureguialzo, E. PUE: The Green Grid metric for evaluating the energy efficiency in DC (Data Center). Measurement method using the power demand. in *International Telecommunications Energy Conference (Proceedings)* (2011). doi:10.1109/INTLEC.2011.6099718.
22. Lei, N. & Masanet, E. Statistical analysis for predicting location-specific data center PUE and its improvement potential. *Energy* **201**, 117556 (2020).
23. Evans, T. The different types of air conditioning equipment for IT environments. *APC White Pap.* 24 (2004).
24. ASHRAE. Climatic data for building design standards. *ASHRAE Stand.* (2013).
25. Ryckbosch, F., Polfliet, S. & Eeckhout, L. Trends in server energy proportionality. *Computer (Long. Beach. Calif).* **44**, 69–72 (2011).
26. Barroso, L. A. & Hölzle, U. The case for energy-proportional computing. *Computer (Long. Beach. Calif).* **40**, 33–37 (2007).
27. Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55**, 271–280 (2001).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformation.docx