Title

Shedding light on U.S. small and midsize data centers: Exploring insights from the CBECS survey

Journal

Energy and Buildings, 339

ISSN

0378-7788

Authors

Lei, Nuoa

Ganeshalingam, Mohan

Masanet, Eric

et al.

Publication Date

2025-07-01

DOI

10.1016/j.enbuild.2025.115734

Peer reviewed

Shedding Light on U.S. Small and Midsize Data Centers: Exploring Insights from the CBECS Survey

Nuoa Lei[1], Mohan Ganeshalingam[1], Eric Masanet[1,2,3], Sarah Smith[1], Arman Shehabi[1]

[1] Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[2] Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA, USA
[3] Department of Mechanical Engineering, University of California, Santa Barbara, USA

Abstract

As digital demand accelerates, the energy and environmental footprint of data centers faces increasing scrutiny. While hyperscale cloud facilities have driven efficiency gains, small and midsize U.S. data centers remain a critical yet underexamined segment with significant untapped potential for energy savings. This study leverages data from the Commercial Buildings Energy Consumption Survey (CBECS) to analyze trends in server stocks, computing customers, cooling system adoption and efficiency, and geospatial distribution from 2012 to 2018. Findings reveal a sharp decline in small and midsize data centers, from 1.764 million to 1.398 million, with server counts dropping from 5.177 million to 4.262 million—aligning with the broader shift toward cloud computing. More than 40% of servers in small data centers and 55% in midsize data centers are housed in office buildings, and over half of all servers are concentrated in climate zones 5A (cold), 3A (mixed-humid), and 4A (mixed-humid), with the highest densities in metropolitan hubs. While direct expansion units remain the dominant cooling system, a clear transition toward more energy-efficient solutions, particularly air economizers, is evident. By integrating server and cooling system distributions, we estimate Power Usage Effectiveness (PUE) and Water Usage Effectiveness (WUE) for U.S. data centers by size and year. Results show that midsize data centers are more energy-efficient but more water-intensive due to the widespread use of water-cooled chillers. These findings highlight the trade-offs in cooling system selection and provide a critical foundation for policies aimed at enhancing efficiency in an evolving data center landscape.

Keywords

Data Centers, Servers, Cooling, Energy Policy, Information and Communication Technology (ICT)

1. Introduction

Data centers, which serve as the backbone of digital services and the digital economy, are highly energy-intensive infrastructures estimated to consume about 1% of the world's electricity (Masanet et al., 2020). As the demand for data center services continues to grow, heightened attention is being devoted to mitigating their energy and environmental impacts (Lei and Masanet, 2021; Rozite Vida, 2023; Shehabi et al., 2024). In response, the data center industry is actively exploring strategic measures to enhance energy efficiency and sustainability (Kamiya and Kvarnström, 2019; Masanet and Lei, 2020). From 2010 to 2018, global data center workloads and compute instances witnessed a more than sixfold increase, while the corresponding energy consumption of global data centers increased only marginally owing to significant advancements in computing efficiency and energy management (Masanet et al., 2020).

Many of these advancements can be credited to improvements within large cloud data centers (Barroso et al., 2013). These facilities leverage highly efficient computing devices, employ ultra-efficient cooling systems, are situated in climate-favorable locations, and implement superior energy efficiency practices (Lei et al., 2023; Lei and Masanet, 2022, 2020). In contrast, their small and midsize counterparts are often constrained by investment and design budgets, geographical locations, and operational and management expertise, resulting in significantly lower efficiencies in terms of resource usage (Lei and Masanet, 2022; Ni and Bai, 2017). These small, decentralized, and heterogeneous facilities, often embedded within commercial buildings and in close proximity to their customers, pose challenges for energy analysts and policymakers in formulating effective policies to address their energy and sustainability issues. Therefore, a nationwide overview of their status and recent trends, encompassing server stock, cooling system types, geospatial locations, and infrastructure efficiency metrics such as power usage effectiveness (PUE) and water usage effectiveness (WUE), can serve as a crucial foundation for crafting informed policies and incentives to promote their energy efficiency and sustainability.

At the same time, there is a global trend of shifting compute instances to large cloud data centers (CISCO, 2018). In the U.S., this trend has prevented significant rises in the national data center energy use, as indicated by estimates of 69 billion kWh and 79 billion kWh in 2012 and 2018, respectively (Shehabi et al., 2016). However, despite the growing volume of compute instances handled by large cloud data centers, the prevalence of small and midsize data centers, representing 81% and 70% of total U.S. data center electricity use in 2012 and 2018 (Shehabi et al., 2016), respectively, suggests significant untapped energy-saving potential. Undoubtedly, harnessing these energy-saving potentials necessitates effective strategies from data center operators and policymakers, who can derive substantial benefits from a detailed understanding of the determinants that influence the energy use and environmental footprint of small and midsize data centers.

However, there is a scarcity of research and datasets offering a comprehensive overview of the energy and environmental impacts of small and midsize U.S. data centers. Existing literature either provides a high-level estimate of their national energy use (Dreibholz et al., 2007; Koomey, 2008; Masanet et al., 2020; Shehabi et al., 2024, 2016), proposes general strategies to mitigate energy and environmental footprint in the data center industry (Koomey, 2011; Lei and Masanet, 2021; Masanet et al., 2013; Mytton, 2021; Ni and Bai, 2017), overly concentrates on large cloud data centers (Lei et al., 2023; Lei and Masanet, 2020; Mytton, 2020), or singularly focuses on individual data centers (Choo et al., 2014; Jin and Bai, 2019; Mi et al., 2023; Sharma et al., 2017), presenting significant challenges for effective policymaking tailored to small and midsize data centers.

This research aims to address the existing knowledge gaps by providing a comprehensive understanding of the energy use and environmental impact of U.S. small and midsize data centers at the national level. Specifically, our study offers comprehensive insights into the server stocks, computing customers, cooling systems distribution and efficiency, and geospatial distribution of servers for U.S. small and midsize data centers, providing detailed snapshots and highlighting their trends from 2012 to 2018. In addition, it provides insights into the digital customers of small and midsize data centers, which can inform strategies to encourage increased penetration of cloud workload migration for energy savings and environmental impact mitigation.

2. Materials and Methods

2.1. Commercial Buildings Energy Consumption Survey

This study employs the Commercial Buildings Energy Consumption Survey (CBECS) to characterize data center server and cooling systems in the United States (Energy Information Administration, 2018, 2012). CBECS is a national survey conducted by the U.S. Energy Information Administration (EIA) to document the energy usage and energy-related characteristics of commercial buildings, which are defined as buildings where more than half of the floorspace is utilized for non-residential, industrial, or agricultural purposes.

The U.S. EIA has released microdata from ten CBECS surveys since 1979. This study analyzes the two most recent surveys (2012 and 2018) to discern recent trends in data centers, considering their robust and systematic sampling strategies, minor variations in survey methodology and content, as well as the increasing significance of U.S. data center energy consumption in light of the recent surge in the digital economy. The microdata for the 2012 and 2018 CBECS comprises 6720 and 6436 records, respectively, representing sampled commercial buildings, with each sample assigned a weight indicating the number of buildings in the U.S. commercial building stock it represents, collectively forming a reliable basis for characterizing U.S. data centers[1] at the national level.

The 2012 and 2018 CBECS document information on server[2] units, cooling systems, principle building activity (PBA)[3], and building locations at the census division level, which provides insights into server stock, information technology (IT) system cooling resource use and efficiency, computing service customers, and the geographical distribution of servers in the United States. To our best knowledge, CBECS is the only publicly-available data source offering such detailed information on servers and cooling systems within the U.S. commercial buildings, making it a valuable resource for understanding the energy and sustainability aspects of the U.S. data center industry.

The following sections introduce the data center classification system and detail the methodology used to enhance the spatial resolution of U.S. data centers and servers. This methodology is primarily based on our previously peer-reviewed report on U.S. data centers (Ganeshalingam et al., 2017), with extensions to incorporate PUE and WUE estimations for U.S. data centers.

2.2. Data Center Categorization

Table 1 compiles the classification metrics employed in this analysis, incorporating the framework established by Mohan (Ganeshalingam et al., 2017) with updated values from the International Data Corporation (IDC) (IDC, 2014a, 2014b). According to these metrics, we employed a systematic scheme (Appendix A) to classify data centers into different space types, leveraging their attributes such as reported server/rack counts, computer area square footage, gross building square footage, and more.

For every building housing servers in the CBECS, we categorize it as either a small, midsize, or large data center, aiming to enhance our understanding of U.S. data center characteristics across different space types. Notably, the EIA employed a categorical representation (coded as '9995' in CBECS) for server counts when a building housed 500 or more servers, leading to around 26 and 85[4] records (equivalent to approximately 1843 and 2373 buildings) in the microdata of the 2012 and 2018 CBECS, respectively (Energy Information Administration, 2018, 2012). These identified buildings likely represent only a subset

---

[1] This study defines a data center as a building or a floorspace within a building that accommodates computer servers.

[2] CBECS defines a server as computing equipment that performs functions such as data storage and processing.

[3] PBA refers to the primary business, commerce, or function conducted within each building, as defined by CBECS for the purpose of classifying buildings.

[4] This value includes records where the derived server counts, calculated from reported rack counts, equal or exceed 500.

of the U.S. large or hyperscale data centers (i.e., service providers), thereby narrowing the focus of this study to small and midsize U.S. data centers.

Table 1. Data center classification and characteristics.

Data Center Space Type	IDC Data Center Environment	Typical Number of Servers	Average Number of Servers per Rack	Space Dimensions (square feet)
Small Data Center	Telecommunications Edge	1-25	0.75	≤1000
	Enterprise Edge			
Midsize Data Center	Internal	26-499	3	1001-20,000
Large Data Center (beyond scope)	Service Provider	≥500	10	>20,000

2.3. Geographical Distribution of small and midsize U.S. Data Centers

To safeguard the data privacy of survey respondents, the EIA excluded geographic details from the CBECS microdata, limiting our ability to quantify server distribution to census division and PBA levels. To achieve enhanced spatial resolutions and conduct a detailed assessment of the geographical distribution of U.S. data centers, we employed the approach established by Mohan (Ganeshalingam et al., 2017), leveraging the robust linear association between server counts and the number of workers within buildings that classified as small or midsize data centers, to approximate server counts by U.S. counties.

In this process, we utilized the Occupation Employment Statistics (OES) from the U.S. Bureau of Labor Statistics (BLS) (U.S. Bureau of Labor Statistics, 2018, 2012), which provides information on the number of employees across different occupation categories and census divisions. The BLS publishes OES data annually, and for temporal consistency, this study utilizes the 2012 and 2018 OES data in conjunction with the respective 2012 and 2018 CBECS microdata. In addition, we established mappings from BLS occupations to CBECS PBAs by assessing the likelihood of an occupation to be associated with a specific PBA (Appendix B). As a result, the number of servers at the U.S. county level can be expressed by Eq. (1), where census division level server counts are first distributed across counties and allocated to buildings with different principle building activities, based on the aforementioned linear association between server counts and workforce size, as well as the occupational-to-building mapping. These estimates are then aggregated back to the census division level by integrating across all buildings with different principle building activities. Having server estimates at the county level further enables us to aggregate server counts at the climate zone level in the later analysis using geospatial maps that define IECC climate zone boundaries (Baechler et al., 2015). This facilitates a consistent and structured mapping of servers across different climate zones, providing additional insights through more granular regional analyses.

$$S_c = \sum_{pba} S_{c,pba} \qquad \text{(Eq. 1)}$$

in which,

$$S_{c,pba} = S_{cd,pba} \times \frac{W_{c,occ\rightarrow pba}}{W_{cd,occ\rightarrow pba}} \qquad \text{(Eq. 2)}$$

where S and W represent the server counts and the number of employees. The subscripts c, cd, pba, and occ denote the county, census division, principle building activity, and occupation, respectively. The notation $occ \rightarrow pba$ represents the mapping from BLS occupations to CBECS PBAs. For further details on these calculations, readers are directed to Mohan (Ganeshalingam et al., 2017) and the accompanying code in the online supplementary materials (OSM).

2.4. Power and water usage effectiveness of small and midsize U.S. Data Centers

PUE and WUE are key metrics for data center efficiency, with PUE measuring the ratio of total power consumption to IT power consumption and WUE representing water use (liters) per kilowatt-hour of IT electricity consumption. The average PUE and WUE for U.S. data centers of a given size (small vs. midsize) and year (2012 vs. 2018) were calculated using a weighted-average approach in this research, where the weights corresponded to server counts at the county level. This calculation process, described by Eqs. (3) and (4), incorporates: (1) PUE and WUE values categorized by data center size, climate zone, and cooling system type, (2) the distribution of cooling system types segmented by year, data center size, and census division, and (3) the server counts across U.S. counties, differentiated by data center size and year.

$$UE_{si,y} = \frac{\sum \overline{UE}_{si,y,cd,cz} \times S_{y,si,c\rightarrow(cz,cd)}}{S_{si,y,c}} \qquad \text{(Eq. 3)}$$

in which,

$$\overline{UE}_{si,y,cd,cz} = \sum UE_{si,cz,cs\prime\rightarrow cs} \times \emptyset_{si,y,cd,cs} \qquad \text{(Eq. 4)}$$

where UE represents the PUE or WUE values of U.S. data centers, while $\emptyset$ denotes the market share of different cooling systems as a percentage. The subscripts si, y, cz refer to data center size, year, and climate zone, respectively. The notation $c \rightarrow (cz, cd)$ represents the mapping of U.S. counties to their corresponding climate zones and census divisions, while $cs' \rightarrow cs$ denotes the mapping of cooling systems with available PUE and WUE estimates (Lei et al., 2023; Lei and Masanet, 2022, 2020) to the cooling system classifications used in this study, both follow a similar classification approach with slight variations in system granularity. Specifically, the cooling system mapping process functions as a data imputation step, with the detailed imputation steps documented in the code available through the OSM.

3. Results and Discussions

3.1. Overview of Trends in Data Centers and Servers Across Different Space Types

Fig. 1. illustrates the observed changes in small and midsize U.S. data centers from 2012 to 2018. While some degree of sampling variability may exist, the CBECS survey employs stratified sampling with scientific weight adjustments to account for different building types, including underrepresented ones (Energy Information Administration (EIA), 2018), ensuring statistically significant national-level estimates of data center and server counts. Given this robust survey methodology, it is confidently concluded that the decline in both the total number of data centers and the servers within them (Fig. 1. (a) and (b)) reflects the industry-wide shift toward cloud computing over this period (Masanet et al., 2020). Specifically, this trend is evident in our CBECS analysis, which shows that the total count of servers within small and midsize data centers dropped from 5.177 million in 2012 to 4.262 million in 2018. In contrast, Shehabi et al. estimated an overall increase in the total number of U.S. servers, rising from 13.581 million in 2012 to 16.832 million in 2018 (Shehabi et al., 2016). These data collectively indicate a corresponding increase in the number of servers within large data centers, from 8.404 million to 12.570 million, illustrating a significant shift of computing workloads towards large cloud data centers in recent years. On the other hand, the average server count per data center showed a distinct trend in the CBECS from 2012 to 2018 (Fig. 1. (c)). While the

average count remained consistent within small data centers, it notably increased from 69 to 100 units per midsize data center, potentially driven by the growing demand for computing resources in this category.

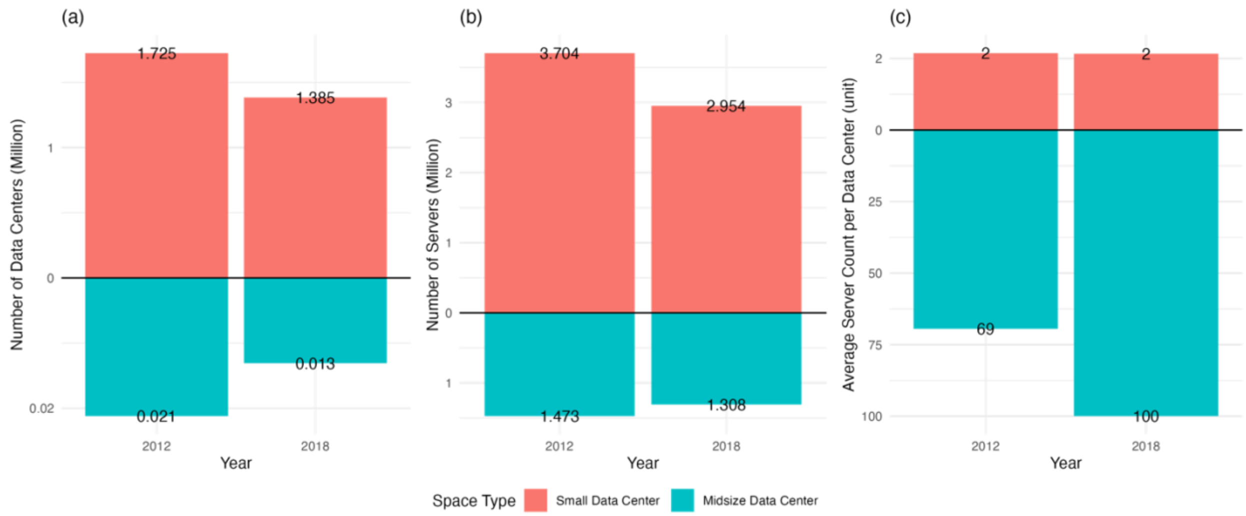

Fig. 1. Changes in U.S. Small and Midsize Data Centers from 2012 to 2018: (a) Number of Data Centers, (b) Number of Servers, (c) Average Server Count per Data Center.

A comprehensive overview of the installed server counts in U.S. small and midsize data centers is presented in Fig. 2. The left panel illustrates that, for both 2012 and 2018, almost half of the servers in small data centers were deployed in settings with fewer than four servers. Particularly, in 2018, about a quarter of the total servers in small data centers were located in places with just one server, a modest decrease from the approximately 31% observed in 2012. For midsize data centers, roughly half of the servers were situated in environments with fewer than 100 servers in 2012. However, this figure rose to 150 servers in 2018, reinforcing the notable surge in server density within midsize data centers, in tandem with the growing trend of digitalization.

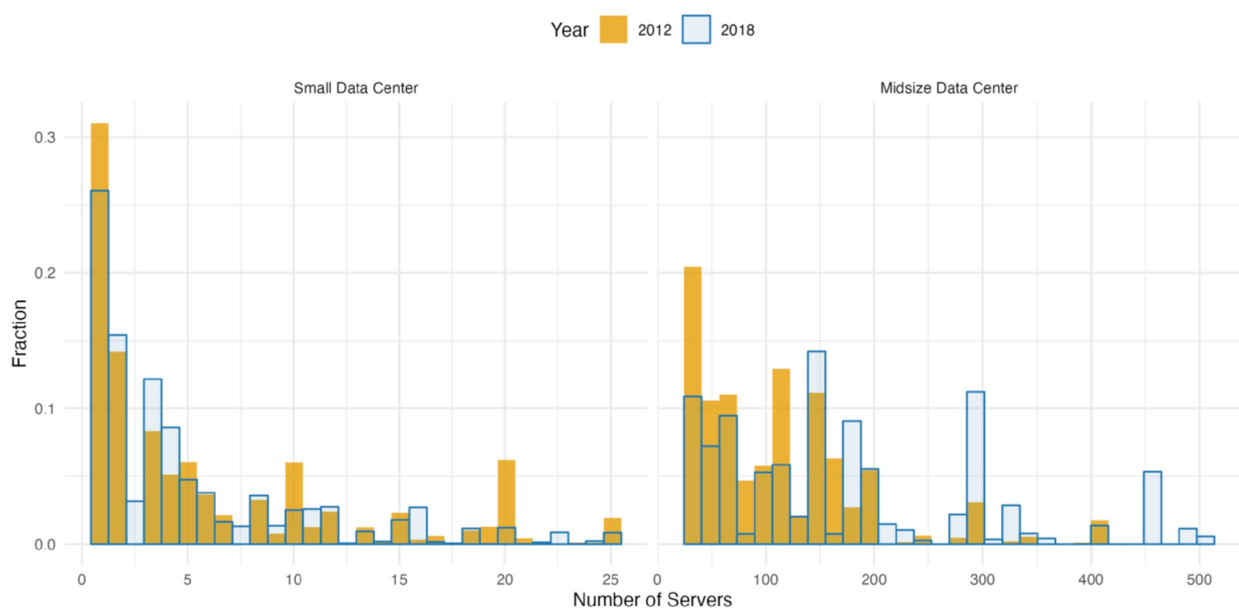

Fig. 2. Distribution of Server Counts Across Data Center Types and CBECS Survey Years (the horizontal axis represents the number of servers per data center, while the vertical axis indicates the total number of

servers in data centers with this specific server count relative to the total server count within U.S. small and midsize data centers).

3.2. Server Distribution by Principle Building Activity

Fig. 3. depicts the distribution of servers across PBAs as defined by CBECS, shedding light on the various use cases and digital-service customers served by U.S. small and midsize data centers. While servers are distributed across diverse building types, it is evident that a substantial majority, accounting for over 40% and 55% of total servers in small and midsize data centers, respectively, are concentrated in office buildings. Remaining servers in small data centers are primarily linked to retail, education, medical, and warehouse purposes, while those in midsize data centers are predominantly associated with medical, education, and public assembly uses. Furthermore, while the decrease in server counts was distributed relatively evenly from 2012 to 2018 within small data centers, the decline in server counts within midsize data centers is primarily attributed to decreases in the medical and warehouse sectors, representing about 161 thousand and 70 thousand servers, respectively. These changes may be attributed to factors such as the widespread adoption of cloud-based enterprise applications within the medical and warehouse sectors, including the implementation of telehealth and inventory management systems. Moreover, despite a decline in servers linked to office buildings over the examined period, they continue to constitute a substantial portion of servers within small and midsize data centers. This indicates considerable potential for computing efficiency improvements and energy savings through cloud workload migration, especially with incentives for the use of cloud-based enterprise applications for office work.

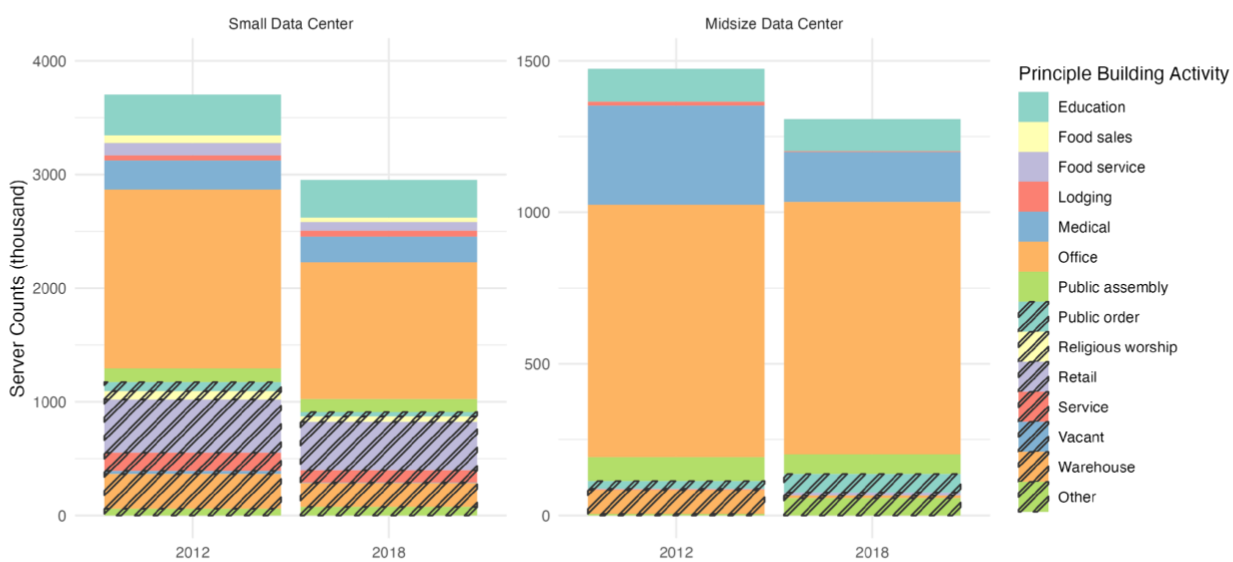

Fig 3. Distribution of Servers by CBECS PBA in Different Data Center Space Types and CBECS Survey Years.

3.3. Distribution of Data Center Cooling Systems

Fig. 4. presents the distribution of cooling system usage in U.S. small and midsize data centers, where the percentages represent the proportion of buildings using each cooling system type within a given year, revealing several key trends in the cooling system market. Firstly, there is a significant variation in the implementation of cooling systems across different data center space types, characterized by the prevalent use of direct expansion systems in small data centers and the significant adoption of air economizers in

midsize data centers. Additionally, direct expansion systems emerge as the predominant refrigeration systems in both small and midsize data centers, primarily owing to their cost-effectiveness in terms of capital, installation, and maintenance, especially when compared with mechanical chillers (Evans, 2012, 2010). Furthermore, a small percentage of small data centers employ air economizer systems, constituting less than 17% of the cooling systems, revealing a significant potential to enhance energy efficiency by utilizing favorable outside air for free cooling, a practice more prevalent in midsize data centers. Lastly, both small and midsize data centers exhibited a consistent trend of embracing more energy-efficient cooling systems from 2012 to 2018, highlighted by the increased utilization of air economizers over this period. Specifically, we observed a noteworthy increase in the adoption of evaporative cooling within midsize data centers, rising from 0.2% to 5.2%. This system is acknowledged for its notable energy and water efficiency among contemporary data center cooling system implementations (Lei and Masanet, 2022, 2020). While the prevalence of evaporative cooling remains relatively low, the increase in its adoption signals a proactive effort by data centers to enhance energy efficiency and environmental sustainability.

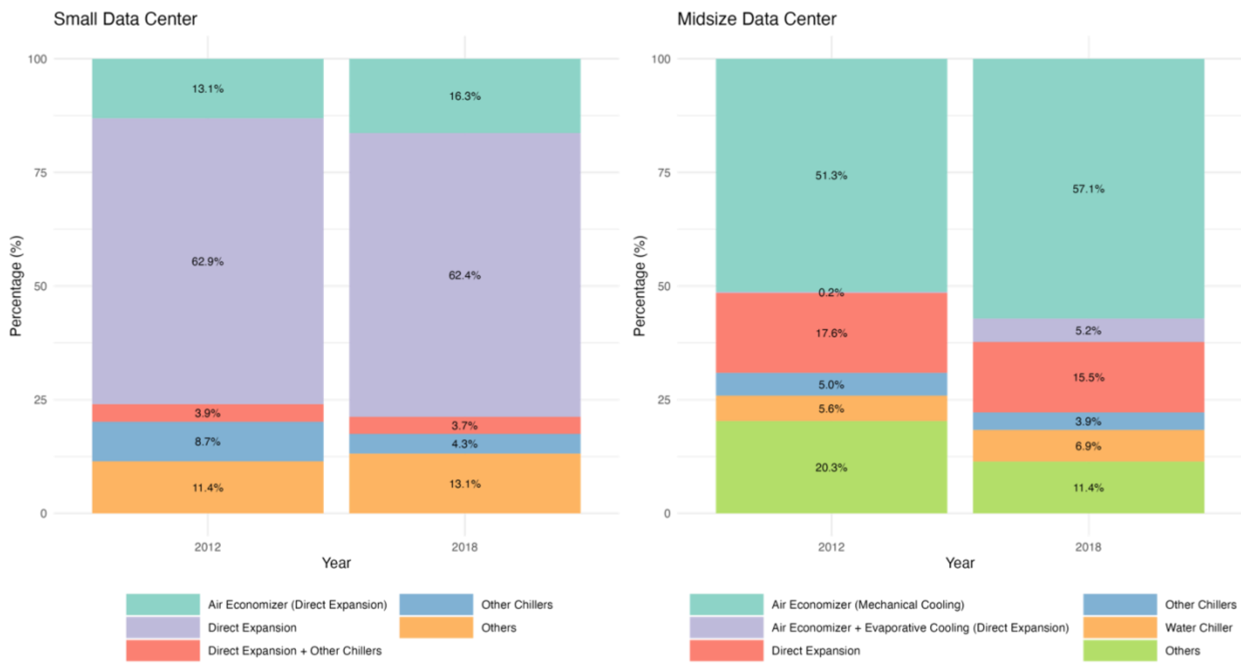

Fig. 4. Distribution of Cooling Systems by Data Center Space Type and CBECS Survey Years. Note: the cooling system within the bracket signifies the supplemental cooling system employed when a data center cannot solely depend on air economizers for cooling; the '+' sign indicates that a data center relies on multiple cooling systems; in the right panel, multiple cooling systems were grouped under the 'mechanical cooling' category, and a more detailed breakdown is presented in the OSM; the corresponding absolute changes in cooling system counts and details on the market share derivation process are available through OSM.

3.4. Geospatial Distribution of Servers in the U.S.

This section provides a geospatial analysis of U.S. servers in small and midsize data centers, derived by integrating the CBECS and OES datasets using methodology outlined in section 2.3. Particularly, we normalized server counts by the area of U.S. counties to control for the effect of geographic size.

Fig. 5. (a) illustrates the geospatial distribution of server density in 2012 for U.S. small and midsize data centers. For simplicity, the visualization of server density in 2018, which produces similar results, is not

presented here. Server density in 2012 demonstrates significant variations across diverse U.S. geographic areas. In terms of census divisions, the Middle Atlantic, New England, and Pacific regions exhibit the highest server density, while the West North-Central, Mountain, and West South-Central regions show the lowest. At the state level, the District of Columbia, New Jersey, New York, and Massachusetts boast the highest server density, contrasting with Wyoming, Montana, North Dakota, and South Dakota, which have the lowest. Drilling down to the county level, New York County, San Francisco County, and the District of Columbia lead in server density, whereas Kootenai County (Idaho), Kent County (Maryland), and Gates County (North Carolina) display the lowest server density. Moreover, metropolitan areas demonstrate significantly higher server density compared to non-metropolitan areas, aligning with the increased population density and job concentrations that heavily depend on digital services in urban areas.

Fig. 5. (b) displays the net change in the server density across the U.S. from 2012 to 2018. Overall, changes in server density are generally modest, with most areas experiencing changes of less than 1 unit per square kilometer, and locations with changes greater than 1 unit per square kilometer are sporadically scattered. Additionally, the majority of U.S. regions witnessed a decline in server density, except for concentrated areas in the East North Central and West North Central regions, where an increase was observed.

Two main factors collectively contribute to the observed geospatial changes in server density. On one hand, the growing demand for digital services is propelling an increase in server density. For example, between 2012 and 2018, both the East North Central and West North Central regions experienced employment growth, with approximately 8.63 million and 1.08 million additional jobs, respectively. Many of these positions involve tasks that rely on servers, resulting in an overall uptick in server density. On the other hand, while the rising demand for digitalization fuels the deployment of servers in data centers, the widespread adoption of cloud-based enterprise applications such as cloud data platforms (e.g., Snowflake, Databricks) and productivity tools (e.g., Microsoft 365, Google Workspace) is driving a shift of computing workloads to large cloud data centers. This transition likely played a counteractive role, contributing to a reduction in servers within small and midsize data centers across most geographical areas in the U.S.

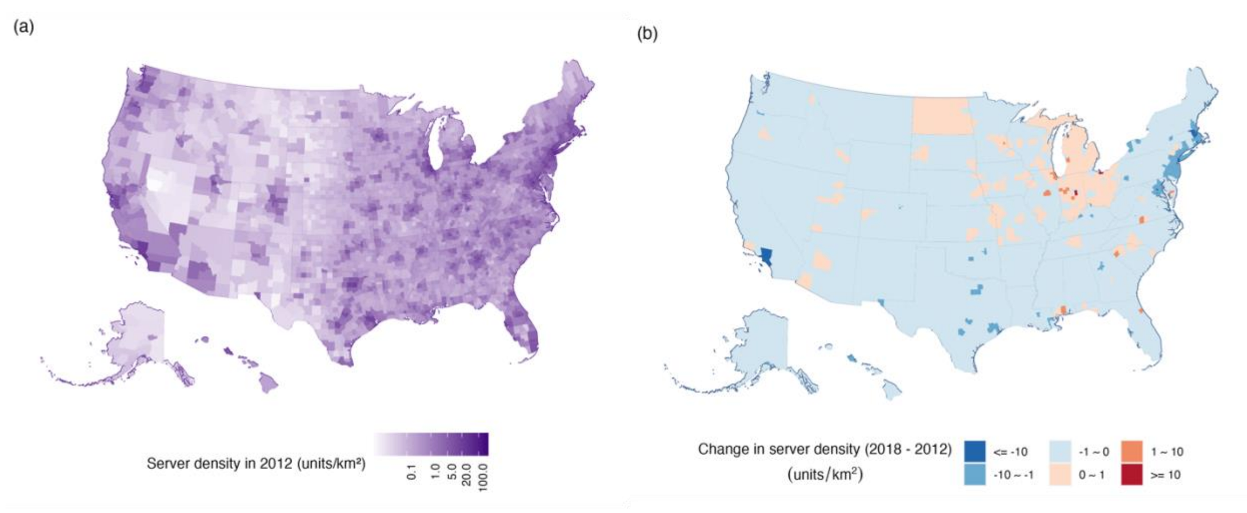

Fig. 5. Geospatial Distribution of Servers in U.S. Small and Midsize Data Centers: (1) Server Density Across Counties in 2012; (2) Changes in Server Density Across Counties from 2012 to 2018 (detailed geospatial distribution data by year and data center space type can be found in the OSM).

Compared to large data centers, small and midsize data centers often demonstrate lower resource use efficiency, particularly concerning cooling system energy and water consumption (Lei and Masanet, 2022). This is primarily attributed to constraints in siting, cooling system selection, and efficiency practices. To investigate the impact of the geospatial distribution of servers on data center resource consumption, this section further provides an overview of server counts in small and midsize data centers across U.S. climate zones, as illustrated in Fig. 6. The International Energy Conservation Code (IECC) categorizes U.S. territories into eight temperature-oriented zones (ranging from 1 for the hottest zone to 8 for the coldest zone) with three moisture subcategories (A for moist, B for dry, and C for marine) (Baechler et al., 2015), and data centers in cooler areas generally exhibit higher efficiencies in their cooling system energy and water use (Lei et al., 2023; Lei and Masanet, 2022).

In both 2012 and 2018, more than 20% of servers in U.S. small and midsize data centers were situated in zone 5A (cold), with over 50% distributed across zones 5A (cold), 3A (mixed-humid), and 4A (mixed-humid). These climate zones generally exhibit modest resource use efficiencies compared to data centers using the same cooling systems in other zones. About 15% of servers are in zones 1A (hot-humid), 2A (hot-humid), and 2B (mixed-dry), representing climate zones where data centers tend to have relatively low resource use efficiencies. Conversely, less than 7% of servers are located in zones 6A (cold), 7 (very cold), and 8 (subarctic), where data centers typically demonstrate higher resource use efficiencies. These colder zones are often preferred sitting positions for large cloud data centers, as they can extensively utilize free cooling throughout the year to reduce operational expenses associated with the cooling system (Bradbury, 2016; Lei and Masanet, 2020). This further reinforces the potential benefits of cloud-shifting, taking advantage of free cooling benefits not easily achievable for small and midsize data centers due to sitting constraints.

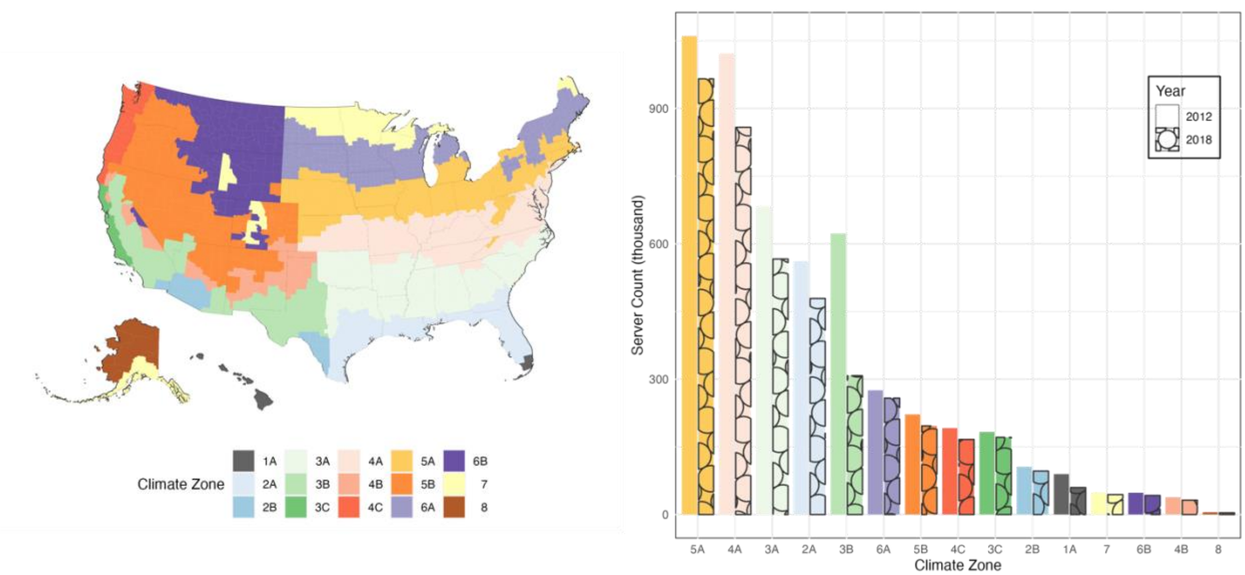

Fig. 6. Geospatial Distribution of Servers in Small and Midsize Data Centers Across U.S. Climate Zones and CBECS Survey Years.

3.5. PUE and WUE of Small and Midsize Data Centers

Fig. 7. presents the aggregated average PUE and WUE values for small and midsize data centers in the U.S. in 2012 and 2018, highlighting their significant variation by size and year. The uncertainties in the aggregated values arise from variations in PUE and WUE within specific cooling system types, driven by

differences in equipment efficiency and operational settings across U.S. data centers (Lei et al., 2023; Lei and Masanet, 2022, 2020).

Compared to small data centers, midsize data centers have notably lower PUE values due to more efficient cooling equipment and better operational practices. However, while midsize data centers generally achieve higher water use efficiency (lower WUE) for a given cooling system compared to small data centers, their overall aggregated WUE across the U.S. is higher. This is primarily due to the greater market share of water chiller systems (typically integrated with cooling towers that consume substantial amounts of water) in midsize data centers, which enhance energy efficiency but require higher water usage. In contrast, most small data centers rely more on direct expansion systems, which are less energy-efficient but require minimal water, resulting in a lower average WUE. This reflects a trade-off among capital investment, space requirements, energy efficiency, and water consumption in cooling system selection across different data center sizes.

A notable change in the aggregated average PUE and WUE values was observed in U.S. data centers from 2012 to 2018 for both small and midsize data centers, primarily driven by shifts in cooling system market share. For small data centers, PUE values remained relatively stable over this period, ranging from 1.82–2.29 in 2012 to 1.82–2.28 in 2018. However, WUE values declined from 0.26–0.42 in 2012 to 0.21–0.35 in 2018. For midsize data centers, PUE decreased from 1.46–1.83 in 2012 to 1.43–1.76 in 2018, reflecting a notable improvement in energy efficiency, primarily due to the increased adoption of cooling systems with economizers and a higher share of evaporative cooling. Conversely, WUE increased slightly from 0.52–0.86 in 2012 to 0.56–0.96 in 2018, reflecting a trade-off favoring energy efficiency over water efficiency in cooling system selection within this data center size category.

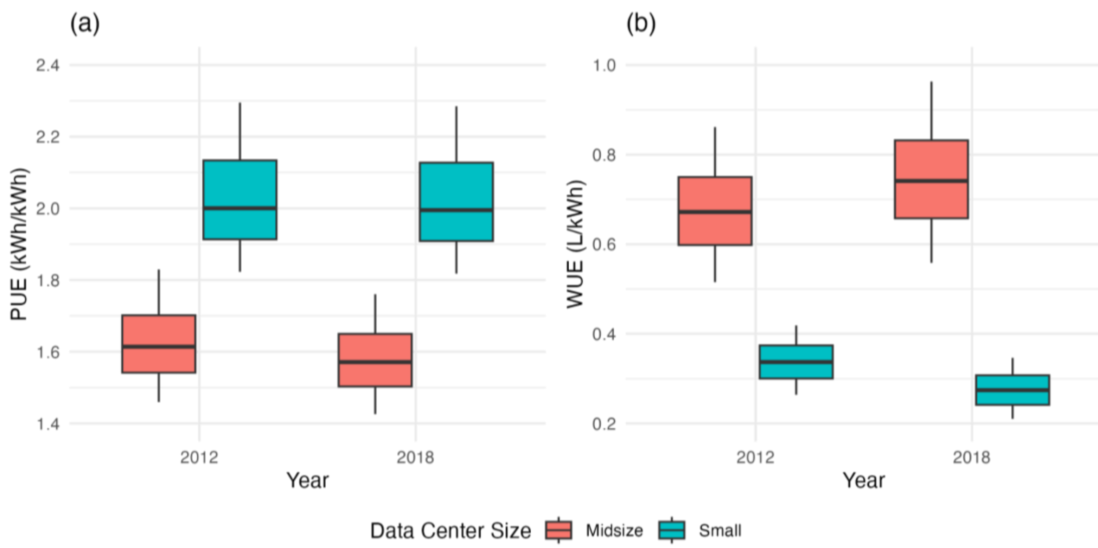

Fig. 7. Estimated PUE and WUE of Small and Midsize U.S. Data Centers (2012 - 2018). Note: the five points in the box plot represent the 5th, 25th, median, 75th, and 95th percentiles of the PUE and WUE results.

3.6. Comparison with Existing Literature

Studies on regional data center information (e.g., server stock, PUE, and WUE) remain limited and vary significantly in geographical coverage, data center classification, and focus year, making direct one-to-one comparisons challenging. To facilitate cross-validation, Table 2 summarizes studies with comparable scopes. The LBNL studies (Shehabi et al., 2024, 2016) focus on the U.S., while Masanet et al. (Masanet et al., 2020) provides regional granularity at the North American level. Additionally, all studies differ in data center classification methods (i.e., data center types) due to variations in data sources and availability. To ensure consistency, we mapped their original data center classifications to those used in this study based on the classification criteria in Table 1. Furthermore, the focus year for each study is indicated in brackets within Table 2 to clarify differences in time scope.

All four studies incorporated server stock data in their analyses. Like other bottom-up studies on data center electricity demand, the two LBNL studies (Shehabi et al., 2024, 2016) relied on IDC market data as a core input. However, their reports only provided aggregate figures at the national level (serving as an upper bound for validation) without disclosing server stock estimates at the granularity of different data center types. This highlights the need for national survey data in future data center electricity demand research, particularly when detailed commercial market data is inaccessible due to cost or confidentiality constraints. Among these studies, Masanet et al. (Masanet et al., 2020) was the only bottom-up study to publicly release comprehensive server stock estimates. While its data was aggregated at the North American level, it provides another upper bound for validating CBECS-derived estimates. However, beyond its broader geographical scope, differences in data center classification methodologies between Masanet et al. and this study complicate direct comparisons, as their classifications do not align perfectly one-to-one. Although beyond the scope of this study, we emphasize the growing importance of considering hyperscale data centers and artificial intelligence (AI) infrastructure in future analyses. As cloud computing expands, data storage demands increase, and AI-driven applications proliferate, server stock composition continues to evolve. This trend is underscored by the latest LBNL study (Shehabi et al., 2024), which finds that U.S. data center electricity consumption is accelerating, driven largely by the rapid deployment of Graphics Processing Unit (GPU)-accelerated servers for AI workloads since approximately 2017.

PUE and WUE have both been key inputs in bottom-up data center electricity and water demand models (Lei, 2022). While reporting years vary slightly across studies, each provides PUE and WUE estimates, with only one exception for WUE, offering a strong foundation for cross-validation. The estimated PUE and WUE ranges align broadly with prior findings, and variations across different data center types are well captured. Notably, both this study's PUE and WUE estimates and those from the latest LBNL report (Shehabi et al., 2024) account for cooling system distribution. Both studies highlight a high prevalence of direct expansion systems in small data centers, while midsize data centers have a greater share of direct expansion systems and airside economizer-based systems. These differences in cooling system adoption drive variations in PUE and WUE across data center sizes. However, while LBNL's cooling system distribution is based on market data, this study relies on CBECS survey data. Differences in data sources, reporting years, and classification methodologies likely explain the minor discrepancies in PUE and WUE estimates observed between the studies. Finally, we note that although beyond the scope of this study, it is essential to consider the implications of PUE and WUE for AI-driven workloads moving forward. The latest LBNL report highlights a major source of uncertainty in PUE and WUE estimates for AI-based data centers, primarily due to the lack of publicly available information on their cooling system implementations. Given its growing significance, we recommend that future iterations of CBECS take the initiative in collecting data on AI-specific cooling systems, including their PUE and WUE, to enhance understanding and forecasting in this rapidly evolving sector.

Table 2. Comparison of data center studies: server stock, PUE, and WUE.

Category	Data Center Type	This study	LBNL (Shehabi et al., 2016)	LBNL (Shehabi et al., 2024)	Masanet (Masanet et al., 2020)
Server Stock (Millions)	Small	3.704 (2012) 2.954 (2018)	13.581 (2012) 16.832 (2018)	14 (2014) 19.4 (2020)	11.866 (2010) 6.403 (2018)

	Midsize	1.473 (2012) 1.308 (2018)			0.449 (2010) 4.848 (2018)
	Hyperscale	N/A			1.677 (2010) 7.384 (2018)
	Artificial Intelligence (AI)			1.6 (2020)	N/A
PUE (kWh/kWh)	Small	1.82-2.29 (2012) 1.82-2.28 (2018)	2-2.5 (2014)	1.78-2.12 (2023)	2.23 (2010) 2.06 (2018)
	Midsize	1.46-1.83 (2012) 1.43-1.76 (2018)	1.7-1.9 (2014)	1.55-1.88 (2023)	1.78 (2010) 1.64 (2018)
	Hyperscale	N/A	1.2 (2014)	1.16-1.28 (2023)	1.22 (2010) 1.18 (2018)
	Artificial Intelligence (AI)		N/A	1.09-1.19 (2023)	N/A
WUE (L/kWh)	Small	0.26-0.42 (2012) 0.21-0.35 (2018)	1.8 (2014)	0.25-0.4 (2023)	N/A
	Midsize	0.52-0.86 (2012) 0.56-0.96 (2018)		0.55-0.8 (2023)	
	Hyperscale	N/A		0.20-0.46 (2023)	
	Artificial Intelligence (AI)		N/A	0.52-0.73 (2023)	

Note: N/A indicates values that are unavailable or not reported due to differences in study scope; numbers in brackets indicate the corresponding year for the reported values; some LBNL data presented in figures is not included in the table due to the challenges and limitations associated with extracting precise values from graphical representations; Masanet et al. reports at the North American level, covering the U.S., Canada, and Mexico, rather than focusing solely on the U.S.

4. Conclusions and future work

This study explored the characteristics of U.S. small and midsize data centers by integrating data from the CBECS and OES datasets, along with prior estimates of PUE and WUE by cooling system type and climate zone. The consistent CBECS survey methodology and content over recent years enable a reliable assessment of trends in U.S. data centers, and the robust data center categorization framework, driven by IDC data, facilitates comprehensive exploration of data center characteristics across different space types. As a result, this study provides a detailed investigation of the server stocks, computing customers, cooling system adoption and efficiency, and the geospatial distribution of servers in small and midsize data centers in the U.S., capturing their changes from 2012 to 2018.

The findings indicate that from 2012 to 2018, the average server density in small data centers remained relatively stable, while midsize data centers experienced a substantial increase of approximately 44%, driven by the growing demand for digital services. Remarkably, despite the ongoing digitalization trend, the total number of servers in U.S. small and midsize data centers declined by approximately 0.915 million during this period. This decline is attributed to the ongoing shift towards cloud computing, with a substantial portion of computations being migrated to large cloud data centers.

The results further indicate that most servers in U.S. small and midsize data centers are affiliated with office buildings, followed by those in medical, education, retail, and warehouse buildings. This sheds light on the primary customers of digital services provided by small and midsize data centers. Interestingly, servers associated with medical and warehouse facilities experienced the most notable decline during the period from 2012 to 2018, attributed to the higher penetration of cloud-based digital services in these sectors.

Our study provides the first comprehensive national-level analysis of cooling system usage in small and midsize U.S. data centers, revealing significant variations across different space types and recent shifts in system preferences. This analysis serves as the foundation for estimating average PUE and WUE, demonstrating that midsize data centers achieve lower PUE due to more efficient cooling equipment and operational practices, yet exhibit higher WUE due to the greater prevalence of water chiller systems. These findings provide essential insights for policymakers aiming to enhance energy and water efficiency in data center sustainability strategies. For example, our results highlight significant electricity-saving potential, particularly through incentivizing the adoption of free cooling (air economizer) or evaporative cooling systems in small and midsize data centers.

In the future, the data that this study produced can be coupled with server hardware power use to estimate local direct electricity and water use. These results can also leverage assumed water consumption factors (WCFs) and air pollutant emission factors for each grid region (Siddik et al., 2021, 2020) to estimate the water consumed, and the carbon and air pollutant emissions generated, by small and midsize U.S. data centers due to electric power generation. Finally, the server stock data can be used to estimate both upstream (e.g., server component manufacturing) and downstream (i.e., e-waste) impacts associated with U.S. small and midsize data centers.

This study contributes to the limited literature on the geospatial distribution of servers in U.S. small and midsize data centers and presents national-level estimates of average PUE and WUE that incorporate cooling system distribution, potentially improving the quantification of national-level data center energy use, water footprint, and carbon emissions (Koomey, 2011; Lei et al., 2023, 2021; Mytton, 2021, 2020; Shehabi et al., 2016; Siddik et al., 2021). It establishes the foundation for incorporating locational variations in climate-determined cooling system energy and water use, along with grid-dependent factors like water consumption and carbon emission associated with electricity generation and consumption (Lee et al., 2018; Siddik et al., 2020). These insights are crucial for a more comprehensive understanding of the environmental impact of data centers. It also highlights the energy and sustainability implications of AI penetration and cloud shifting, underscoring the need for greater scrutiny of the environmental impact of AI workloads, which remains poorly understood. Additionally, it provides valuable insights for policymakers evaluating the relocation of local computing workloads to the cloud, where favorable climate conditions can enhance data center cooling efficiency and abundant renewable energy sources can support growing energy demands.

The scope of this study was limited to small and midsize data centers in the U.S. Future research can enhance this analysis by aggregating data related to server stock, customers, cooling systems, and geospatial distribution of large cloud and AI data centers. It is important to acknowledge that the assumption of the linear association between server counts and the number of workers within buildings may not be applicable when approximating server geospatial distribution for large cloud and AI data centers, and subsequent work should concentrate on exploring alternative methods to investigate this aspect. Future research can build upon the insights provided in this study to refine estimates of energy consumption and the environmental footprint of data centers. This can be achieved by incorporating granular data on power draw of IT devices (including but not limited to servers), IT device management practices, electricity water consumption factor, and the carbon embodied in electricity (Lei and Masanet, 2021; Masanet et al., 2020; Masanet and Lei, 2020; Shehabi et al., 2016). Last but not least, a national survey on data center efficiency and management practices across different space types, including large cloud and AI data centers, and specifically focusing on IT devices and cooling systems, could serve as a valuable supplement to further enhance and extend this analysis.

Acknowledgements

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 is supported by the Office of Science of the United States Department of Energy and operated under contract grant no. DE-AC02-05CH11231

Online Supplementary Materials

Supplementary data and code related to this article can be accessed via:
https://github.com/nuoaleon/Shedding-Light-on-U.S.-Small-and-Midsize-Data-Centers-Exploring-Insights-from-the-CBECS-Survey

Appendix A. Data Center Classification Scheme

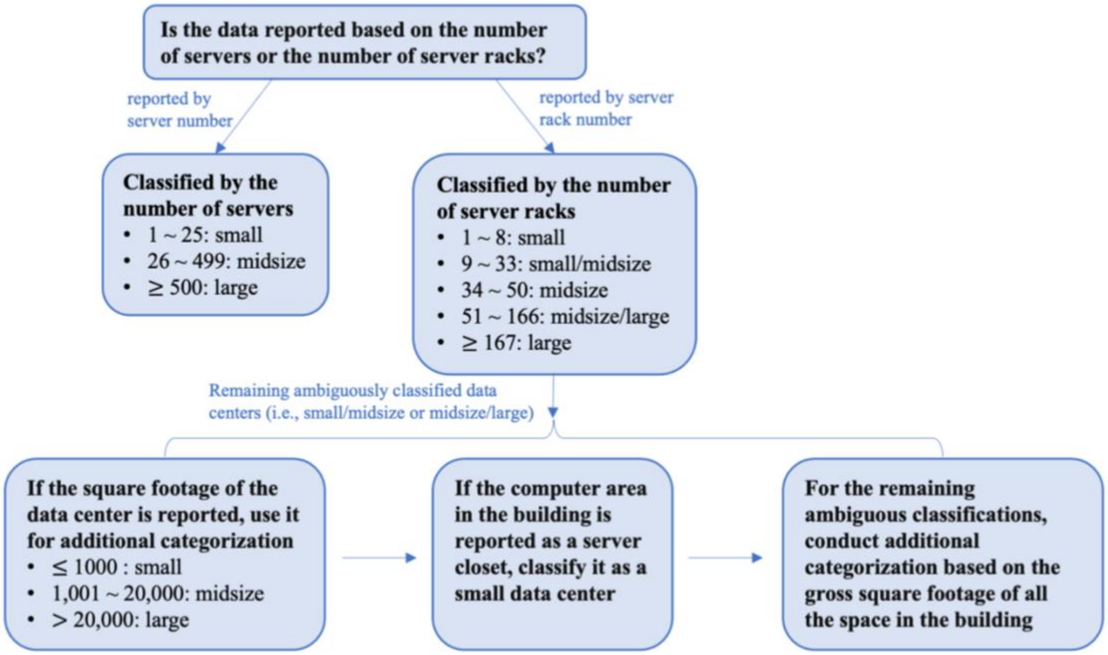

Fig A.1. Data Center Classification Scheme.

The data center classification scheme in Fig. A.1. was expanded using Mohan's framework (Ganeshalingam et al., 2017), with server count ranges directly adopted from Mohan's work. Rack number ranges for classification were determined by dividing the server count (bounding values of the server count ranges) by the average number of servers per rack. Overlaps exist in the derived rack number ranges for data center classification, resulting in data centers being classified as small/midsize or midsize/large data centers. These ambiguous cases were further categorized based on computer area square footage, reported data center information, or gross building square footage. For more information, readers can refer to the code available in the OSM.

Appendix B. OES Job Codes to CBECS PBAs Mapping

Table B.1. Mapping of OES Job Codes to CBECS PBAs

OES Job Code	CBECS PBA
Management Occupations	Office
Business and Financial Operations Occupations	Office
Computer and Mathematical Occupations	Office, Laboratory
Architecture and Engineering Occupations	Office, Laboratory
Life, Physical, and Social Science Occupations	Office, Laboratory
Community and Social Service Occupations	Office, Religious worship
Legal Occupations	Office
Education, Training, and Library Occupations	Education, Public assembly

Arts, Design, Entertainment, Sports, and Media Occupations	Public assembly
Healthcare Practitioners and Technical Occupations	Outpatient health care, Inpatient health care, Nursing
Healthcare Support Occupations	Outpatient health care, Inpatient health care, Nursing
Protective Service Occupations	Public order and safety
Food Preparation and Serving Related Occupations	Food service
Building and Grounds Cleaning and Maintenance Occupations	Office, Laboratory, Lodging
Personal Care and Service Occupations	Service
Sales and Related Occupations	Food sales, Retail other than mall, Enclosed mall, Strip shopping mall, Food service
Office and Administrative Support Occupations	Office, Lodging
Farming, Fishing, and Forestry Occupations	Agricultural
Construction and Extraction Occupations	Industrial
Installation, Maintenance, and Repair Occupations	Other
Production Occupations	Non-refrigerated warehouse, Refrigerated warehouse, Other
Transportation and Material Moving Occupations	Non-refrigerated warehouse, Other

Appendix C. Classification and Grouping of Cooling System

The classification and grouping of data center cooling systems in this study is based on the presence of specific cooling components as indicated in the CBECS microdata:

- Systems identified with Room Air Conditioner (RCAC), Packaged Air Conditioning Unit (PKGCL), or Central Air Conditioning with No Water Loop (ACWNWL) were categorized as using Direct Expansion for mechanical cooling;
- Systems identified with Air-Cooled Chillers (CHLAIRCL) were categorized as using Air-Cooled Chillers for mechanical cooling;
- Systems identified with Water-Cooled Chillers (CHLWTRCL) were categorized as using Water-Cooled Chillers for mechanical cooling;
- Systems identified with Absorption Chillers (CHLABSRP) were categorized as using Absorption Chillers for mechanical cooling;
- Systems employing other types of mechanical cooling equipment were grouped under the category "Other Chillers";
- If a system indicated the use of Evaporative Cooling and/or Air Economizer, these technologies were assumed to be the primary cooling methods, supplemented by the mechanical cooling systems described above, especially during periods of peak demand such as hot summer months;
- The remaining rare or uncommon cooling equipment types were consolidated into an “Others” category. Given that the above classification method could theoretically result in up to 2^8 permutations of cooling system configurations, categories with market shares below 3% for small

data centers and below 4% for midsize data centers were further grouped into the “Others” category to maintain clarity and practicality.

Notably, this categorization initially yielded more than ten distinct cooling system types for midsize data centers. To enhance the interpretability of Fig. 4, systems employing Air Economizers in combination with supplemental cooling methods—such as Direct Expansion, Air-Cooled Chillers, Water-Cooled Chillers, or Other Chillers—were consolidated into a single category titled “Air Economizer (Mechanical Cooling)”. The final grouped results are presented in Fig. 4, while detailed original subcategory data, along with comprehensive methodological specifics, are available in the OSM.

References

Baechler, M.C., Gilbride, T.L., Cole, P.C., Hefty, M.G., Ruiz, K., 2015. High-Performance Home Technologies: Guide to Determining Climate Regions by County. Pac. Northwest Natl. Lab. Oak Ridge Natl. Lab.

Barroso, L.A., Clidaras, J., Hölzle, U., 2013. The datacenter as a computer: An introduction to the design of warehouse-scale machines, second edition, Synthesis Lectures on Computer Architecture. https://doi.org/10.2200/S00516ED2V01Y201306CAC024

Bradbury, D., 2016. Super cool: Arctic data centres aren’t just for Facebook [WWW Document]. The Register. URL https://www.theregister.com/2016/05/12/power_in_a_cold_climate/ (accessed 12.29.23).

Choo, K., Galante, R.M., Ohadi, M.M., 2014. Energy consumption analysis of a medium-size primary data center in an academic campus. Energy Build. 76, 414–421. https://doi.org/10.1016/J.ENBUILD.2014.02.042

CISCO, 2018. Forecast and Methodology, 2016–2021 [WWW Document]. Cisco Glob. Cloud Index. URL (accessed 7.7.21).

Dreibholz, T., Becke, M., Adhari, H., 2007. Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431, Tdr.Wiwi.Uni-Due.De. LBNL, Berkeley, CA.

Energy Information Administration, 2018. Commercial Buildings Energy Consumption Survey (CBECS) [WWW Document]. URL https://www.eia.gov/consumption/commercial/ (accessed 12.24.23).

Energy Information Administration, 2012. Commercial Buildings Energy Consumption Survey (CBECS) [WWW Document]. URL https://www.eia.gov/consumption/commercial/ (accessed 12.24.23).

Energy Information Administration (EIA), 2018. Commercial Buildings Energy Consumption Survey (CBECS) Methodology [WWW Document]. URL https://www.eia.gov/consumption/commercial/data/2018/index.php?view=methodology (accessed 2.22.25).

Evans, T., 2012. The Different Technologies for Cooling Data Centers. Schneider Electr.

Evans, T., 2010. How and Why Mission-Critical Cooling Systems Differ From Common Air Conditioners White Paper #56. Schneider Electr.

Ganeshalingam, M., Shehabi, A., Desroches, L.-B., 2017. Shining a Light on Small Data Centers in the USA 1–31.

IDC, 2014a. IDC’s Worldwide Quaterly Server Shipment Tracker, 2010-2018.

IDC, 2014b. U.S. Datacenter Census and Construction 2014–2018 Forecast: Realigning Workloads, Managing Obsolescence, and Leveraging Hyperscale.

Jin, C., Bai, X., 2019. The study of servers’ arrangement and air distribution strategy under

partial load in data centers. Sustain. Cities Soc. 49, 101617. https://doi.org/10.1016/J.SCS.2019.101617
Kamiya, G., Kvarnström, O., 2019. Data centres and energy – from global headlines to local headaches? [WWW Document]. IEA. URL https://www.iea.org/commentaries/data-centres-and-energy-from-global-headlines-to-local-headaches (accessed 12.30.23).
Koomey, J., 2011. Growth in Data Center Electricity use 2005 to 2010. Anal. Press 1–24.
Koomey, J.G., 2008. Worldwide electricity used in data centers. Environ. Res. Lett. 3, 034008. https://doi.org/10.1088/1748-9326/3/3/034008
Lee, U., Han, J., Elgowainy, A., Wang, M., 2018. Regional water consumption for hydro and thermal electricity generation in the United States. Appl. Energy 210, 661–672. https://doi.org/10.1016/j.apenergy.2017.05.025
Lei, N., 2022. A Hybrid Physics-Based and Data-Driven Modeling Framework for Energy and Water Use Analysis of Data Centers with Spatio-Temporal Resolution. Northwestern University, Evanston, IL, USA.
Lei, N., Cheng, Z., Cao, Z., Masanet, E., 2021. An integrated decision-making framework for sustainable data center operation through intelligent load scheduling. J. Phys. Conf. Ser. 2042, 012091. https://doi.org/10.1088/1742-6596/2042/1/012091
Lei, N., Lu, J., Cheng, Z., Cao, Z., Shehabi, A., Masanet, E., 2023. Geospatial assessment of water footprints for hyperscale data centers in the United States. J. Phys. Conf. Ser. 2600, 172003. https://doi.org/10.1088/1742-6596/2600/17/172003
Lei, N., Masanet, E., 2022. 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. https://doi.org/10.1016/j.resconrec.2022.106323
Lei, N., Masanet, E., 2021. Global Data Center Energy Demand and Strategies to Conserve Energy, in: Data Center Handbook, 2nd Edition. WILEY, Palo Alto, CA. https://doi.org/10.1002/9781119597537.ch2
Lei, N., Masanet, E., 2020. Statistical analysis for predicting location-specific data center PUE and its improvement potential. Energy 201, 117556. https://doi.org/10.1016/j.energy.2020.117556
Masanet, E., Lei, N., 2020. How Much Energy Do Data Centers Really Use? Aspen Glob. Change Inst. 2021.
Masanet, E., Shehabi, A., Koomey, J., 2013. Characteristics of low-carbon data centres. Nat. Clim. Change 3, 627–630. https://doi.org/10.1038/nclimate1786
Masanet, E., Shehabi, A., Lei, N., Smith, S., Koomey, J., 2020. Recalibrating global data center energy-use estimates. Science 367, 984–986. https://doi.org/10.1126/science.aba3758
Mi, R., Bai, X., Xu, X., Ren, F., 2023. Energy performance evaluation in a data center with water-side free cooling. Energy Build. 295, 113278. https://doi.org/10.1016/J.ENBUILD.2023.113278
Mytton, D., 2021. Data centre water consumption. Npj Clean Water 4, 1–6. https://doi.org/10.1038/s41545-021-00101-w
Mytton, D., 2020. Hiding greenhouse gas emissions in the cloud. Nat. Clim. Change 10, 701. https://doi.org/10.1038/s41558-020-0837-6
Ni, J., Bai, X., 2017. A review of air conditioning energy performance in data centers. Renew. Sustain. Energy Rev. 67, 625–640. https://doi.org/10.1016/j.rser.2016.09.050
Rozite Vida, 2023. Data centres & networks - IEA [WWW Document]. URL https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-

networks (accessed 12.24.23).
Sharma, P., Pegus, P., Irwin, D., Shenoy, P., Goodhue, J., Culbert, J., 2017. Design and Operational Analysis of a Green Data Center. IEEE Internet Comput. https://doi.org/10.1109/MIC.2017.2911421
Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Nuoa, L., Siddik, M.A.B., Jonathan, K., Eric, M., Dale, S., 2024. 2024 United States Data Center Energy Usage Report. Berkeley Lab.
Shehabi, A., Smith, S.J., Sartor, D.A., Brown, R.E., Herrlin, M., Koomey, J.G., Masanet, E.R., Horner, N., Azevedo, I.L., Lintner, W., 2016. United States Data Center Energy Usage Report. Berkeley Lab 65.
Siddik, M.A.B., Chini, C.M., Marston, L., 2020. Water and Carbon Footprints of Electricity Are Sensitive to Geographical Attribution Methods. Environ. Sci. Technol. 54, 7533–7541. https://doi.org/10.1021/ACS.EST.0C00176/SUPPL_FILE/ES0C00176_SI_002.XLSX
Siddik, M.A.B., Shehabi, A., Marston, L.T., 2021. The environmental footprint of data centers in the United States. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/abfba1
U.S. Bureau of Labor Statistics, 2018. Occupational Employment and Wage Statistics [WWW Document]. URL https://www.bls.gov/oes/ (accessed 12.25.23).
U.S. Bureau of Labor Statistics, 2012. Occupational Employment and Wage Statistics [WWW Document]. URL https://www.bls.gov/oes/ (accessed 12.25.23).