Title
Mitigating Curtailment and Carbon Emissions through Load Migration between Data Centers

Journal
Joule, 4(10)

ISSN
2542-4785

Authors
Zheng, Jiajia
Chien, Andrew A
Suh, Sangwon

Publication Date
2020-10-01

DOI
10.1016/j.joule.2020.08.001

Peer reviewed

Manuscript

Joule

CellPress

Research Article

Mitigating curtailment and carbon emissions through data center load migration

Jiajia Zheng,[1] Andrew A. Chien,[2,3] and Sangwon Suh[1,4,*]
[1]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA
[2]Department of Computer Science, University of Chicago, Chicago, IL 60637, USA
[3]Argonne National Laboratory, Lemont, IL 60439, USA
[4]Lead Contact
*Correspondence: suh@bren.ucsb.edu

SUMMARY

As the share of variable renewable energy (VRE) grows in the electric grid, so does the risk of curtailment. While energy storage and hydrogen production have been proposed as solutions to the curtailment problem, they often pose technological and economic challenges. Here, we analyze the potential of data center load migration for mitigating curtailment and greenhouse gas (GHG) emissions. Using historical hourly electricity generation, curtailment, and typical data center server utilization data, we simulate the effect of migrating data center workloads from fossil fuel-heavy PJM to renewable-heavy CAISO. The results show that load migration within the existing data center capacity during the curtailment hours in CAISO has the potential to reduce 113-239 KtCO$_2$e yr^{-1} of GHG emissions and absorb up to 62% of the total curtailment with negative abatement cost in 2019. Our study demonstrates the overlooked role that data centers can play for VRE integration and GHG emissions mitigation.

Keywords: variable renewable energy, curtailment, data center, greenhouse gas emissions mitigation, load migration

INTRODUCTION

Driven by aggressive public policy and compelling economics, global capacity of Variable Renewable Energy (VRE), such as solar photovoltaics (PV) and wind electricity, is growing rapidly. European Union, for example, has a target to achieve at least 32% share of renewable energy by 2030,[1] and California aims at 60% renewable portfolio standard by 2030.[2]

As the penetration of VRE in the grid grows, so do the concerns of large-scale curtailment.[3–5] Curtailment is the reduction of output of a VRE resource below what it could have otherwise produced. It has been repeatedly reported in different world regions across Europe, America and Asia, significantly decreasing the market value of VRE.[5,6] Near-term reasons for VRE curtailment include minimum generation requirement for non-renewable energy sources and transmission constraints, but long term, fundamental causes drive increasing pressure for curtailment.[5,7] Large-scale energy storage and electricity transmission network expansion can mitigate VRE curtailment, but they are costly. With a system cost between \$380 to \$895 per kWh,[8] the battery storage capacity deployed globally (12 GWh in 2018)[9] is infinitesimal compared to the amount of global electricity consumption (about 23,000 TWh per year).[10] Long-distance transmission of VRE-generated electricity is possible but the construction of transmission infrastructure and the associated transmission losses are often cost-prohibitive.[11] Pumped hydro can be another storage solution, but it requires certain geographical features and may raise ecological concerns.[12]

Another approach to reduce curtailment is to use excess VRE electricity to produce more easily storable materials or products, such as hydrogen through water electrolysis, which can be shipped upon demand.[13,14] However, the logistics and handling of these materials and associated costs can pose additional challenge.[15] A potential solution to this logistics and handling problem is to use over-generated electricity to produce something that can be transported at minimal cost and energy: information.

Data centers can provide battery-like demand side management service by powering data processing with excess VRE. The technical and economic potential of zero-carbon cloud data centers which solely run on stranded renewable power has been explored.[16,17] Geographical load balancing has been widely studied to maximize renewable energy use by distributing the workloads among data centers in different locations.[18,19] Data centers are highly automated and monitored with little human interventions. Importantly, they have considerable flexible workloads which can be distributed geographically.[17,20] Provided the requisite data is available, those data centers with renewable energy access can process the requests routed from other regions and return the results to users while meeting Service Level Agreements (SLAs). Moreover, most data centers operate well below 100% capacity most of the time – over-provisioning for peaks leaves servers and network resources underutilized.[21] Furthermore, the peak time of data center loads often does not coincide with the peak time of VRE over-generation, providing room for data centers to use their excess capacity to process additional workloads with excess VRE.

Compared with building large-scale transmission infrastructure, building fiber optics networks and transmitting data are much cheaper, and takes significantly less time.[11] Therefore, the transmission of data is more economically favorable than the transmission of electricity, i.e. "moving bits, not watts".[22] Furthermore, the society's needs for data processing is growing rapidly. Global data centers used 205 TWh electricity in 2018 or 1% of global electricity consumption.[23] In 2014, U.S. data centers consumed 70 TWh electricity, which was 1.8% of the total annual U.S. electricity consumption.[24] It is estimated that the global datasphere will grow from 33 zettabytes (ZB) in 2018 to 175 ZB in 2025 at an annual growth rate of 27%, implying growing needs for data center infrastructures.[25] The decarbonization of data centers is imperative, and require combined efforts including maximizing IT-device efficiency, adoption of low-carbon electricity and improving infrastructure efficiency.[26] Load migration between data centers can collectively improve IT efficiency and utilization of renewable energy. Nevertheless, the potential for load migration between data centers to utilize excess VRE generation and reduce greenhouse gas (GHG) emissions has not been quantified.

In this study, we use two Independent System Operators (ISOs) in the U.S., California ISO (CAISO) and Pennsylvania-New Jersey-Maryland Interconnection (PJM), as a case study to explore the potential of workloads migration between data centers to mitigate curtailment and GHG emissions. PJM is the largest ISO in the U.S., which predominantly relies on thermal energy sources like coal and natural gas, with solar and wind accounted for only 3.2% in 2019.[27,28] The states covered by PJM host a large amount of data centers, with the most noteworthy area being North Virginia. In the second half of 2018, North Virginia saw over a third of the world's new data center capacity absorption, with an addition of 270 Megawatts (MW) data center power.[29] As the hub of technology and media companies, California also has many data centers, mostly located in the Bay Area and Southern California.[30] Of all the data center colocation establishments in the U.S., 13% are located in California and about 15% are in PJM region.[31,32]

Based on the historical hourly curtailment data of CAISO and a typical data center energy consumption profile, we evaluate the potential of the existing and additional data center capacities to absorb excess VRE and reduce GHG emissions by migrating data center workloads from PJM region. In this analysis, we use counterfactual scenarios as an illustration of the potential rather than as a record of historical accounts.

RESULTS

Historical curtailment of CAISO

We collected and analyzed the historical curtailment data of CAISO during 2015-2019. The total annual curtailment of CAISO grew from 188 GWh to 965 GWh from 2015 to 2019, at an average annual growth rate of 51%. Curtailment data at CAISO shows wide daily and seasonal variations, with an upward trend over time (Fig. 1a). Solar PV curtailment accounted for 90% of the total cumulative curtailment during this period and wind accounted for 10%. The majority of curtailment occurred in the first and second quarter, which combined accounted for 69% of the total curtailment in the period. Monthly curtailment peaked in April or May. This results from growing solar radiation strength and extended daytime length during spring, combined with mandatory runoffs from northwest hydro generation imports and cool weather. Both solar and wind curtailment occurred the least in the third quarter with July or August seeing the minimum, which can be explained by higher cooling demands in summer's warmer weather. The surge of solar curtailment during 2015-2019 mirrors the fact that the share of solar power in total CAISO generation had increased from 6.7% to 13.0% in this period. In comparison, the share of wind power in the generation mix increased from

5.3% to 7.2%, representing a milder growth than solar. When disaggregated at hourly resolution (Fig. 1b), curtailment took place rather randomly throughout the 24 hours of a day in 2015 and 2016, but as solar PV capacity grew and nighttime wind curtailment decreased during 2017-2019, total curtailment became increasingly more conspicuous in the daytime.

[INSERT FIGURE 1]

Instead of curtailing, the excess VRE generation in CAISO could be used to process data center workloads migrated from carbon-intensive grid regions. Many data centers operate at less than 50% average server utilization rate,[24,33,34] and the time-zone difference between PJM and CASIO helps avoid peaking load at the same time, allowing the data centers served by CAISO to take on additional data processing jobs migrated from PJM-served data centers during the off-peak hours. We use the historical curtailment data, which is referred to as “excess VRE” hereafter, to evaluate the potential of migrating workloads between data centers.

Hourly GHG intensity of CAISO and PJM

We collected and treated the electricity generation by energy resource data for CAISO and PJM during 2016-2019.[28,35] The life-cycle GHG intensities of the two ISOs during 2016-2019 (Fig. 2) were calculated on an hourly basis based on the historical generation data and U.S.-specific GHG emission factors, which include both combustion emissions and life-cycle emissions embodied in the inputs to power generation (Table S1). Imported electricity is not included in the calculation.

[INSERT FIGURE 2]

The annual average GHG intensity of PJM decreased from 499 to 452 kgCO_2e/MWh during 2016-2019. The monthly average intensity of PJM peaked in summer (July or August) and reached its lowest around April and October, with a range between 417-557 kgCO_2e/MWh. For CAISO, the annual average GHG intensity changed from 262 to 231 kgCO_2e/MWh during the same period. The monthly intensity of CAISO hit the lowest in April or May due to prominent solar and hydro power production.

During the hours when curtailment occurred in CAISO, only the intensity values of the curtailment (i.e. excess VRE) are shown (Fig. 2 - CAISO), which is assumed to be proportionally contributed by curtailed solar and wind power. In other words, the GHG emissions intensity of the excess generation is calculated as the average life-cycle GHG emissions intensity of solar and wind weighted by their shares in the total curtailment during that hour. While the average GHG intensity of CAISO excess generation during 2016-2019 was 41 kgCO_2e/MWh, the average GHG intensity of PJM during CAISO’s excess generation time was 476 kgCO_2e/MWh. The significant differences of the GHG intensities between the two grids during CAISO excess generation time present a great opportunity for GHG emissions mitigation by migrating the data center workloads geographically.

The capacity of data centers to absorb excess VRE

We first estimated the existing data center capacity in CAISO region. We collected the available data center location and power consumption data from a colocation data center industry website.[36] By examining the profiles of all the listed data centers in California, we calculated that the average annual total power consumption per colocation site is 9.92 MW, based on 26 data points that provided the information. We also identified that currently there are 288 data centers in CAISO region by the end of 2019.[37] We use a typical data center energy profile with an IT peak power (or critical power) of 10 MW as a standardized unit[38] in this study to estimate the excess VRE absorption capacity, GHG emissions reduction potential and abatement cost.

We then simulate the Dynamic Range (DR) and Power Usage Effectiveness (PUE) of the data centers served by CAISO. DR is the ratio of a server’s idling power to its maximum power,[24] based on which we calculate the energy consumption of servers given the rated power and utilization rate. PUE is defined as the ratio of the data center total energy consumption to IT equipment energy consumption, calculated, measured or assessed across the same period.[39] PUE values vary depending on data center type and geographical location. Here, we model the average PUE of colocation data centers in California. Detailed assumptions of the two parameters can be found in Experimental Procedures and Table S3. We also developed linear model between the hourly server utilization rate and the energy use of non-server components,[38] through which we can calculate the non-server energy consumption given a certain PUE value in a year.

We compare two scenarios for evaluating the excess VRE absorption potential of data center workloads migration: Baseline Scenario and Migration Scenario. In Baseline Scenario, workloads are processed by typical data centers served by PJM without any migration. In Migration Scenario, workloads are first migrated to and processed by the existing typical data centers served by CAISO. We assume that the migration occurs between data centers of similar scale with typical energy use characteristics in our model. Once the existing capacity is exhausted, we assume building additional data center capacity which run solely on the remaining excess VRE. We assume that the data centers all have advanced algorithms and automation mechanisms in place to enable the load migration.

[INSERT FIGURE 3]

Fig. 3 illustrates the excess VRE absorption potential of a typical data center in a week. The remaining capacity of an existing data center in an underutilized hour is calculated by subtracting the existent load of the data center in that hour from its maximum allowed load. Load migration is only enabled during the hours that the servers in data centers served by CAISO are underutilized. We test different scenarios by varying the assumption of the maximum allowed server utilization rate (UR) between 65% and 90% during underutilized time, representing an improved management and a maximized utilization scenario, respectively. Average utilization rate of large-scale cloud providers is estimated as 65%.[40] Once the remaining capacities of all existing data centers are exhausted, we calculate the respective additional data center capacity needed to absorb different portions of the total excess VRE. During excess generation hours, the servers in the additional data centers would be activated to process the workloads migrated from PJM region and operate at the maximum allowed UR assumed. The servers are assumed to be shut down at times when there is no available excess VRE.

GHG emissions reduction and abatement cost

We calculate the achieved total GHG emissions reduction by summing up the products of the hourly GHG intensity difference between the two scenarios and the amount of excess VRE absorbed for each year between 2016-2019. We then estimate the total abatement cost of the plan by comparing the cost difference between the two scenarios. For the excess VRE that fall within the remaining capacity of existing data centers, workloads migration causes only a change in electricity bills between the two scenarios. When additional data centers are built to absorb extra excess VRE, changes in electricity cost, amortized facility cost and additional cost are all captured.

We use the cost estimates developed specifically for zero-carbon cloud (ZCC) data centers that run on stranded renewable power[16] for the additional data center capacity in Migration Scenario. These intermittent data centers have lower facility cost because they use containers and can be located near renewable generation sites with less power distribution costs.[16] The electricity cost is also significantly lower for ZCC data centers than the traditional ones as the otherwise-curtailed VRE electricity is assumed to have zero cost. Additional cost for installing data and energy storage devices will incur due to the intermittent characteristic of the power supply for ZCC data centers. The cost of applications including software licenses, system and database administration are not considered as they vary greatly and do not constitute part of the infrastructure-related capital or operational cost. The total abatement cost sums up the difference of facility, electricity and additional costs between the two scenarios on an annual basis. The net abatement cost (in \$/metric ton CO_2e) is then calculated by dividing the total abatement cost by the total net GHG emissions reduction achieved.

Fig. 4 summarizes the results of total GHG emissions reduction and net GHG abatement cost using 2019 data. The existing data center capacity alone (i.e. when additional data center capacity is zero) can absorb 29%-62% of the total excess VRE in CAISO in 2019, assuming that the maximum server UR ranges between 65% and 90%. As we increase the maximum server UR and additional data center capacity, the excess VRE absorption level grows. At a given absorption level, a higher maximum server UR means a reduced need for additional data center capacity.

[INSERT FIGURE 4]

The resulting GHG emissions reduction is the net reduction after accounting for the embodied GHG emissions of the additional data centers, which are incurred due to the manufacturing of IT equipment and infrastructure materials. The embodied emissions are proportional to the total number of new data centers built and would offset a fair amount of the operational GHG emissions reduction achieved by data center workloads migration. A total net GHG emissions of 113-239 $KtCO_2e$ could have been reduced in 2019 given the maximum server UR range assumed, and additional data center capacity can bring the total reduction further up to 247 $KtCO_2e$ (Fig. 4a). The net GHG emissions reduction peaks at an absorption level between 68%-75% when the maximum server UR exceeds 85%. Absorbing 80% of the total excess VRE or above does not bring more GHG emissions reduction benefits because the embodied emissions from more additional data centers would outweigh the reduction from operational phase.

Fig. 4b shows the estimated net abatement cost of the Migration Scenario. A negative abatement cost means that data centers can generate extra profits by mitigating GHG emissions through load migration. Without building any additional data center capacity, the existing capacity has a net abatement cost of -$242/$tCO_2e$ under any maximum server UR, driven by a decrease in electricity cost. The economic break-even point of additional data center capacity that should be built decreases from 780 MW to 350 MW as the maximum server UR increases from 65% to 90% (Fig 4b, the white line of zero). The lowest abatement costs (up to -$688/$tCO_2e$) can be reached by keeping the maximum server UR below 70% and the additional data center capacity between 150 MW-350 MW, corresponding to an absorption level of 45%-60%. It is possible to absorb up to 77%-79% excess VRE while still manage to keep the net abatement cost negative, but an absorption goal of more than 80% does not make sense economically under any maximum server UR as the net abatement cost would always stay positive. The results of GHG emissions reduction and abatement costs for years 2016-2018 are shown in Figures S1-S3.

DISCUSSION

The inherent intermittent nature of VRE poses a major challenge to the stability and profitability of electric grids. Curtailment around the world is likely to grow as the share of VRE continues to rise, unless strong measures are taken to mitigate it[5]. Even with curtailment, over-generation still occurs, in which case, electricity is routinely sold at negative prices. In CAISO, for example, the share of 5-minute intervals with negative prices between 2014 and 2018 ranged 2%-4%.[41]

Our study shows that workloads migration between data centers can potentially absorb excess VRE, reducing both curtailment and GHG emissions at no or negative cost. The existing data center capacity served by CAISO has the potential to absorb up to 62% of the excess VRE and reduce GHG emissions of up to 239 $KtCO_2e$ with a net abatement cost of -$242/$tCO_2e$ in 2019, provided that server utilization rate is improved (Fig. 4). Additional data centers could further absorb the cumulative excess VRE up to 79% and reduce the GHG emissions up to 247 $KtCO_2e$ while still maintaining negative abatement cost.

Furthermore, the potential for workloads migration among data centers to mitigate curtailment and GHG emissions is likely to grow, as the needs for data processing services and data center infrastructure continue to expand. In order to capture such growing potential for workloads migration, a number of institutional and technological changes are due. In particular, development of the technology, policy, and protocols that enable real-time workloads migration among data centers during the time of excess renewable generation would be needed. In addition, it is essential to develop the mechanism to incentivize data centers on load migration between grid regions and to facilitate the fluid communication among multiple grid operators and data centers. Technologies should be developed to support highly dynamic data center operation based on instantaneous generation, load and capacity data. Reliable short-term VRE generation forecasting capability for accurate projection of VRE over-generation is indispensable for dynamic load migration during the excess generation hours.[42]

Besides the spatial flexibility of data centers, the temporal flexibility of certain types of workloads also holds great potential for demand response, which is not evaluated in this analysis. Flexibly scheduling delay-tolerant workloads can increase renewable energy usage and accommodate more renewable resources in the grid.[19,43] Some Internet service providers, for example Google, have developed carbon-aware scheduling technology to shift compute tasks across time to maximize renewable energy utilization, noting that their next step is to move the tasks between data centers in different locations.[44] Incentivized deadline deferral can be an effective way to change customers' behaviors and encourage workloads shifting.[45] Energy storage devices in data centers can also play a role at smoothing VRE power by charging during over-generation time and discharging at a later time.[46,47]

There are some challenges facing workloads migration, but potential solutions are available. First, workloads migration incurs additional network delay and thus potential service violations.[18,19] The network latency time from U.S. east to west is currently around 60 milliseconds,[48] which is short but not trivial. Fortunately, delay-tolerant workloads including scientific computation, big data analytics, medical image processing etc. are a major component of data center workloads, accounting for more than 50% of total workloads,[33] far more than the fraction of workloads we model the migration of. Interactive workloads which are inappropriate to shift because of the user-response latencies required, such as web search and videoconference, are assumed not to be migrated. Network latency has been decreasing thanks to the growing transmission speed and capacity of optic fibers; some companies already achieved speeds of hundreds of terabytes per second. The intermittent nature of excess renewable generation requires the workloads to be easily interruptible as needed, which makes it challenging for some workloads.[47] Nevertheless, this problem can be alleviated through installing more solid state drives and energy storage to checkpoint-restart the jobs.[16] The confidentiality of data center information remains another major concern. Data center owners are usually reluctant to share data about their facilities, including power consumption, to the public.[49] The confidentiality-related concerns can potentially be addressed through data reporting and aggregation protocols, advanced encryption technology and economic incentives.[23]

Our findings are applicable not only in the U.S. but also in other world regions with growing penetration of VRE. Data centers can and should play an important role in global VRE integration and GHG emission mitigation, especially in a future when the capacities of data processing and renewable energy are both rapidly growing.

EXPERIMENTAL PROCEDURES

Resource Availability

Lead Contact

Please contact the Lead Contact, Sangwon Suh (suh@bren.ucsb.edu) for information related to the data and code used in the following experimental procedures.

Materials Availability

No materials were used in this study.

Data and Code Availability

The datasets and codes generated during this study are available at Mendeley Data http://dx.doi.org/doi:10.17632/wngs282m48.1.

Historical curtailment and GHG intensity

We collected the historical solar and wind curtailment data of CAISO at a 5-min interval during 2015-2019.[50] We analyzed and visualized the curtailment data on an hourly basis. To calculate the hourly GHG intensity of the two grids, we first collected the data of electricity supply by energy resource type. We collected the hourly generation data by resource type with breakdown of renewable resources of CAISO during 2015-2019 from CAISO website.[35] For PJM, the generation data was obtained from the Generation by Fuel Type dataset from Data Miner 2 database.[28] The generation data for year 2015 was not available in PJM database so only 2016-2019 data was used. Due to daylight savings time change, the generation data of PJM at 2 a.m. in a certain day in March was missing, and we handled it by filling it with the average value of two adjacent hours; there were duplicate data points at 1 a.m. in a November day, and we only kept the latter one of the two duplicate hours.

To calculate the GHG emissions and intensities, we used the life-cycle GHG emissions data for each energy resource. Life-cycle GHG emissions are the total emissions from all stages of an energy resource's life cycle, covering upstream, operational and downstream processes. These processes include fuel/raw material extraction, transport, infrastructure construction/equipment manufacturing, combustion (for fossil fuels), equipment operation and maintenance and waste treatment. For natural gas, we used the life-cycle GHG emissions value of natural gas combined cycle in the U.S. reported by National Energy Technology Laboratory.[51] For coal, we used the generation-weighted average life-cycle GHG emissions data based on an investigation of over 300 coal power plants in the U.S.[52] For low-carbon energy resources, the median values of the life-cycle GHG emissions presented in the Fifth

Assessment Report by Intergovernmental Panel on Climate Change (IPCC) were applied.[53] The category "thermal" in CAISO generation data was treated as natural gas as the share of coal is negligible. The category "other renewables" in PJM generation data is considered as an equal mix of biomass and biogas. For any unspecified resource type such as "multiple fuels" and "other", an average unspecified emission value was used.[54] Table S1 summarizes the life-cycle GHG emissions values of electricity generation by energy resource that were used in this study. Imports of CAISO were not included when calculating the GHG intensity of generation as the energy mix was not clear. PJM was a net exporter of electricity as we summarized its interchange dataset.[55]

The hourly GHG intensity of electricity supply during 2016-2019 was calculated as the weighted average of the life-cycle GHG emissions of all the energy resources in that hour (equation 1).

$$GHG\_intensity_h = \sum_i (GHG_{LC,i} \times Generation_{i,h}) \div \sum_i Generation_{i,h} \quad (1)$$

In equation (1), h is a certain hour in a year, and i is a certain resource type. $GHG\_intensity_h$ (in $kgCO_2e/MWh$) is the GHG intensity of the grid in hour h, $GHG_{LC,i}$ (in $kgCO_2e/MWh$) is the life-cycle GHG emissions of resource i, $Generation_{i,h}$ (in MWh) is the electricity generated by resource i in hour h and $\sum_i Generation_{i,h}$ (in MWh) is the total electricity generation by all resources in hour h.

Reductions in curtailment and GHG emissions

The electricity consumption profile of a real-world data center on an hourly basis is difficult to obtain due to the secretive nature of the industry. Therefore we use the simulated energy consumption profile of a data center which has a critical (IT) power of 10 MW and a typical data center design.[38] We obtained the hourly electricity consumption data of the typical data center in a week and extended the weekly profile to a year. The data center has a total peak power of 21 MW and consumes approximately 114 GWh electricity annually.[38] The detailed technical specifications and the energy consumption data of this typical data center are shown in Table S2. Seasonal variations of the energy consumption are not considered.

Energy consumption of a data center is jointly determined by their IT and non-IT energy efficiency. We made assumptions of two key parameters to model future data center energy use, Dynamic Range (DR) and Power Usage Effectiveness (PUE), considering the energy efficiency improvement of both IT and non-IT components. The yearly values of the parameters assumed for 2016-2019 are shown in Table S3. DR determines the lowest power (idling power) consumption of servers, which serves as the intercept in the linear model between server power usage and server utilization rate. As server power efficiency improves, DR value gets lower. In other words, the power usage of servers while idling would decrease over time as a result of improved energy proportionality.[34] We assume that the average DR value drops from 0.25 to 0.11 from 2012 to 2019 based on the historical DR value changes reported in several literature sources.[34,56,57] For the energy efficiency improvement of non-IT components in data centers, we simulate the change of PUE value during the examined period. We assume that the average PUE of data centers served by CAISO decreases from 1.59 to 1.30 from 2012 to 2019 based on the PUE trend in the data center industry surveys conducted by Uptime Institute[58,59] and recent PUE values of California data centers from a colocation website.[36] We also developed a model based on the energy profile from the reference to simulate the linear relationship between hourly non-server energy consumption and server utilization rate ($R^2 = 0.988$). With this model, we calculated the hourly non-server energy consumption given certain annual average PUE value assumed for the year. Both PUE and DR values are assumed to change linearly during the examined period.

Under Migration Scenario, during the hours when there is excess VRE in CAISO, workloads from the data centers served by PJM are assumed to be migrated to the data centers served by CAISO. The time difference between CAISO and PJM regions of three hours is considered. We assume using the remaining capacity of existing data centers to respond to excess VRE generation of CAISO first, and then use additional data center capacity to absorb the rest of the excess generation. The remaining capacity of existing data centers is determined by the allowed maximum server utilization rate during underutilized hours, for which we set different levels between 65% and 90% as explained in the main text. While in theory, servers should be able to run at 100% utilization rate, in practice they are run at significantly lower utilizations to tolerate the burstiness of computations, and fluctuations of loads[34]. Our assumption of 90% as the upper bound of the maximum server UR is a conservative assumption with respect to the benefits of workloads migration. The additional data centers, assumed to be Zero-Carbon Cloud (ZCC) data centers, run at the maximum server utilization rate with the remaining excess VRE that exceeds the existing data centers' capacity. The GHG emissions reduction is then calculated as the difference of total GHG emissions between the Baseline Scenario and the Migration Scenario, i.e. by multiplying the amount of excess VRE generation absorbed by CAISO data centers and the GHG intensity difference between PJM generation and the excess VRE generation in CAISO.

To ensure a holistic perspective, the embodied GHG emissions of additional data centers are taken into account. They include all the non-operational emissions that come from the manufacturing of IT, electrical, mechanical equipment and building materials, etc. Life cycle assessment (LCA) studies of data centers are scarce in the literature, but we identified one study that has the energy consumption breakdown by data center component. The study shows that non-operational emissions account for 6.5% of the total life-cycle climate change impacts of a data center[60]. We calculated that the yearly non-operational energy consumption of a data center is around 432 MWh electricity per MW of critical (IT) power based on the data from the reference, assuming a 5-year IT refresh rate as the additional data centers run intermittently.[60] According to the Emissions & Generation Resource Integrated Database (eGRID) by U.S. EPA, the average GHG intensity of U.S. grid was 456 and 432 $kgCO_2$/MWh in 2016 and 2018, respectively. We extrapolated the two points linearly and estimated that the average GHG intensity in 2017 and 2019 was 444 and 421 $kgCO_2$/MWh, respectively. The embodied GHG emissions of a U.S. data center therefore amounted to 0.20-0.18 $KtCO_2e$/MW critical power per year during 2016-2019. We validated the number by analyzing the data from another earlier data center LCA study and it yielded similar estimate.[61]

Estimation of abatement cost

For the costs of additional data center capacity under the Migration Scenario, we use the cost estimates developed for the ZCC data centers that run solely on stranded renewable power.[16] They typically co-locate with existing renewable generation facilities and therefore the costs of power transmission and distribution can be reduced. The total abatement cost sums up the changes between the Baseline Scenario and Migration Scenario in facility cost, electricity cost and additional cost for a certain year.

(1) Facility cost. The amortized physical facility cost of ZCC data centers ($0.50 per watt of critical power) is markedly lower than that of traditional data centers ($5.25/W) by using containers and co-locating at renewable generation sites[16]. But ZCC data centers run intermittently depending on the availability of excess VRE generation, and traditional ones run continuously as a comparison. Under the Migration Scenario, the amortized facility cost of ZCC data centers is obtained by simply multiplying $0.50/W with the total IT power, e.g. a 10 MW ZCC data center has an amortized facility cost of $5 million. To capture the facility cost of typical data centers served by PJM under the Baseline Scenario, we divided the amortized facility cost (e.g. $52.5 million for a typical traditional 10 MW data center) by its annual energy consumption, and then multiplied this unit facility cost (in $/MWh) with the amount of excess VRE generation absorbed by additional data centers through workloads migration in that year. In other words, we allocated the facility cost of data centers under the Baseline Scenario based on the amount of load that could be migrated to additional data centers under the Migration Scenario. Supplemental Equation S15-S17 represent the mathematical expressions of calculating facility costs.

(2) Electricity cost. We assume zero cost for the over-generated electricity under the Migration Scenario since excess VRE generation is regarded as "stranded energy" and it would have been curtailed if not utilized. We used the historical average retail electricity prices for Virginia during 2016-2019 as representative values to calculate the electricity costs of typical data centers powered by PJM under the Baseline Scenario.

(3) Additional cost. Due to the fact that ZCC data centers run on intermittent excess VRE generation, additional solid state drives (SSDs) and energy storage devices to checkpoint-restart jobs interrupted by a power outage are needed.[16] Combined with hardware for free cooling, the total additional cost incurred is $0.175/W per year.

The amortized compute cost and network cost are assumed to be the same under the two scenarios. Workloads migration has the potential to increase software licensing costs, particularly if additional virtual instances are required. We do not model this cost, due to the lack of public information both on software use and licensing. The detailed cost breakdown is summarized in Table S4. Summing up all the cost changes across facility cost, electricity cost and additional cost due to workloads migration, we estimate the total abatement cost for each year. Then by dividing the total abatement cost by the total net GHG emissions reduction, we derive the net abatement cost standardized by one unit of GHG emissions reduction in a certain year. All the computation steps are presented in Equations S1-S21, with nomenclature listed in Table S5.

Limitations

There are a few uncertainties and limitations with this study. First, we used a typical data center energy consumption profile to estimate the remaining absorption capacity for excess VRE. In reality, it may not be representative enough as the data centers in CAISO region probably have various energy consumption patterns. We also extrapolated the weekly energy consumption data to the entire year, while in fact the profile may change because of temperature range under different climate conditions. Second, there is no complete and transparent database available on the current data center capacity in the U.S., and the information of the power consumption of the data centers is particularly scarce, so we had to use limited available data points to estimate the existing capacity of data centers in CAISO region. Third, the DR and PUE values of data centers in real world vary depending on the data center type, scale and location. We simulated the average values as a simplification when geographic-specific and fine-grained data are lacking. We also simplified the analysis by assuming that the migration occurs between data centers of similar scale with typical energy use characteristics, while the electricity consumption to process a same compute task may be different for data centers with contrasting characteristics. Lastly, there is uncertainty with the data center costs. The cost components of data centers fall in a broader spectrum in the real-world, and they may evolve in the future due to a variety of reasons including disruptive technology development. The abatement cost of workloads migration may involve more potential cost categories such as new devices and algorithms that are necessary to enable the load migration and communication between grid operators and data centers.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Table S1-S5, Figures S1-S3 and Supplemental Experimental Procedures.

ACKNOWLEDGMENTS

J.Z. is, in part, supported by Bren School Graduate Student Fellowship. S.S. and J.Z. are, in part, supported by the National Science Foundation under Grant No.1360445. A.A.C is supported by National Science Foundation under Grant No.1832230. We thank Eric Masanet, Mahnoosh Alizadeh, and Ranjit Deshmukh, all at the University of California, Santa Barbara, for their valuable comments.

AUTHOR CONTRIBUTIONS

Conceptualization, S.S.; Methodology, J.Z., S.S. and A.A.C.; Formal Analysis, J.Z.; Investigation, J.Z., S.S. and A.A.C.; Resources, S.S.; Visualization, J.Z.; Writing – Original Draft, J.Z.; Writing – Review & Editing, S.S. and A.A.C.; Supervision, S.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. European Union (2016). 2030 climate & energy framework. Clim. Action - Eur. Comm. https://ec.europa.eu/clima/policies/strategies/2030_en.
2. State of California. California Renewables Portfolio Standard (RPS) Program. https://www.cpuc.ca.gov/rps/.
3. Jenkins, J.D., Luke, M., and Thernstrom, S. (2018). Getting to Zero Carbon Emissions in the Electric Power Sector. Joule *2*, 2498–2510.
4. Frew, B., Cole, W., Denholm, P., Frazier, A.W., Vincent, N., and Margolis, R. (2019). Sunny with a chance of curtailment: Operating the U.S. grid with very high levels of solar photovoltaics. iScience, S2589004219303967.
5. Wiser, R.H., Mills, A., Seel, J., Levin, T., and Botterud, A. (2017). Impacts of Variable Renewable Energy on Bulk Power System Assets, Pricing, and Costs.
6. Bird, L., Lew, D., Milligan, M., Carlini, E.M., Estanqueiro, A., Flynn, D., Gomez-Lazaro, E., Holttinen, H., Menemenlis, N., Orths, A., et al. (2016). Wind and solar energy curtailment: A review of international experience. Renew. Sustain. Energy Rev. *65*, 577–586.

7. Denholm, P., O'Connell, M., Brinkman, G., and Jorgenson, J. (2015). Overgeneration from Solar Energy in California. A Field Guide to the Duck Chart (National Renewable Energy Laboratory).
8. Fu, R., Remo, T., and Magolis, R. (2018). 2018 U.S. Utility-Scale Photovoltaics-Plus-Energy Storage System Costs Benchmark.
9. Wood Mackenzie Power & Renewable (2019). Global energy storage outlook 2019.
10. Enerdata (2019). Global Energy Statistic Yearbook 2019. https://yearbook.enerdata.net/electricity/electricity-domestic-consumption-data.html.
11. Bird, S., Achuthan, A., Ait Maatallah, O., Hu, W., Janoyan, K., Kwasinski, A., Matthews, J., Mayhew, D., Owen, J., and Marzocca, P. (2014). Distributed (green) data centers: A new concept for energy, computing, and telecommunications. Energy Sustain. Dev. *19*, 83–91.
12. Rehman, S., Al-Hadhrami, L.M., and Alam, Md.M. (2015). Pumped hydro energy storage system: A technological review. Renew. Sustain. Energy Rev. *44*, 586–598.
13. Gahleitner, G. (2013). Hydrogen from renewable electricity: An international review of power-to-gas pilot plants for stationary applications. Int. J. Hydrog. Energy *38*, 2039–2061.
14. Barton, J.P., and Infield, D.G. (2004). Energy storage and its use with intermittent renewable energy. IEEE Trans. Energy Convers. *19*, 441–448.
15. Klerke, A., Christensen, C.H., Nørskov, J.K., and Vegge, T. (2008). Ammonia for hydrogen storage: challenges and opportunities. J. Mater. Chem. *18*, 2304–2310.
16. Yang, F., and Chien, A.A. (2018). Large-Scale and Extreme-Scale Computing with Stranded Green Power: Opportunities and Costs. IEEE Trans. Parallel Distrib. Syst. *29*, 1103–1116.
17. Chien, A.A., Wolski, R., and Yang, F. (2015). The Zero-Carbon Cloud: High-Value, Dispatchable Demand for Renewable Power Generators. In The Electricity Journal, pp. 110–118.
18. Liu, Z., Lin, M., Wierman, A., Low, S.H., and Andrew, L.L.H. (2011). Geographical Load Balancing with Renewables. SIGMETRICS Perform Eval Rev *39*, 62–66.
19. Kong, F., and Liu, X. (2014). A Survey on Green-Energy-Aware Power Management for Datacenters. ACM Comput Surv *47*, 30:1–30:38.
20. Goiri, Í., Haque, M.E., Le, K., Beauchea, R., Nguyen, T.D., Guitart, J., Torres, J., and Bianchini, R. (2015). Matching renewable energy supply and demand in green datacenters. Ad Hoc Netw. *25*, 520–534.
21. Jin, X., Zhang, F., Vasilakos, A.V., and Liu, Z. (2016). Green Data Centers: A Survey, Perspectives, and Future Directions. ArXiv160800687 Cs.
22. Liu, Z., Wierman, A., Chen, Y., Razon, B., and Chen, N. (2013). Data center demand response: Avoiding the coincident peak via workload shifting and local generation. Perform. Eval. *70*, 770–791.
23. Masanet, E., Shehabi, A., Lei, N., Smith, S., and Koomey, J. (2020). Recalibrating global data center energy-use estimates. Science *367*, 984–986.
24. Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., Masanet, E., Horner, N., Azevedo, I., and Lintner, W. (2016). United States Data Center Energy Usage Report.
25. Reinsel, D., Gantz, J., and Rydning, J. (2018). The Digitization of the World from Edge to Core.
26. Masanet, E., Shehabi, A., and Koomey, J. (2013). Characteristics of low-carbon data centres. Nat. Clim. Change *3*, 627–630.
27. PJM (2019). PJM At a Glance. https://www.pjm.com/~/media/about-pjm/newsroom/fact-sheets/pjm-at-a-glance.ashx.
28. PJM PJM Data Miner 2 - Generation by Fuel Type. https://dataminer2.pjm.com/feed/gen_by_fuel.
29. Sverdlik, Y. (2019). The Cloud Data Center Construction Boom, in Two Charts. Data Cent. Knowl. https://www.datacenterknowledge.com/cloud/cloud-data-center-construction-boom-two-charts.
30. Ghatikar, G., Ganti, V., Matson, N., and Piette, M.A. (2012). Demand Response Opportunities and Enabling Technologies for Data Centers: Findings From Field Studies.
31. DatacenterMap.com Colocation USA - Data Centers. https://www.datacentermap.com/usa/.
32. Curran, J. (2018). IBISWorld Industry Report OD5899 - Colocation Facilities in the US (IBISWorld Services).
33. Cortez, E., Bonde, A., Muzio, A., Russinovich, M., Fontoura, M., and Bianchini, R. (2017). Resource Central: Understanding and Predicting Workloads for Improved Resource Management in Large Cloud Platforms. In Proceedings of the 26th Symposium on Operating Systems Principles - SOSP '17 (ACM Press), pp. 153–167.
34. Barroso, L.A., Hölzle, U., and Ranganathan, P. (2019). The Datacenter as a Computer: Designing Warehouse-Scale Machines (Morgan & Claypool).
35. CAISO California ISO - Renewables and emissions reports. http://www.caiso.com/market/Pages/ReportsBulletins/RenewablesReporting.aspx.
36. CloudandColocation.com California Data Center Location Map. Cloud Coloca. https://cloudandcolocation.com/state/california/.
37. Cloudscene Data Centers Market in United States of America. Cloudscene. https://cloudscene.com/market/data-centers-in-united-states/all.
38. Rahmani, R., Moser, I., and Seyedmahmoudian, M. (2018). A Complete Model for Modular Simulation of Data Centre Power Load.
39. International Organization for Standardization (2016). ISO/IEC 30134-2:2016(en), Information technology — Data centres — Key performance indicators — Part 2: Power usage effectiveness (PUE). https://www.iso.org/obp/ui/#iso:std:iso-iec:30134:-2:ed-1:v1:en.
40. AWS News Blog (2015). Cloud Computing, Server Utilization, & the Environment. Amaz. Web Serv. https://aws.amazon.com/blogs/aws/cloud-computing-server-utilization-the-environment/.
41. CAISO (2019). 2018 Annual Report on Market Issues & Performance.
42. Sun, Y., Szűcs, G., and R. Brandt, A. (2018). Solar PV output prediction from video streams using convolutional neural networks. Energy Environ. Sci. *11*, 1811–1818.
43. Alstone, P., Potter, J., Piette, M., Schwartz, P., Berger, M., Dunn, L., Smith, S., Sohn, M., Aghajanzadeh, A., Stensson, S., et al. (2017). 2025 California Demand Response Potential Study - Charting California's Demand Response Future: Final Report on Phase 2 Results.
44. Radovanovic, A. (2020). Our data centers now work harder when the sun shines and wind blows. Google. https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/.
45. Paul, D., Zhong, W., and Bose, S.K. (2017). Demand Response in Data Centers Through Energy-Efficient Scheduling and Simple Incentivization. IEEE Syst. J. *11*, 613–624.

46. Goiri, Í., Katsak, W., Le, K., Nguyen, T.D., and Bianchini, R. (2013). Parasol and GreenSwitch: Managing Datacenters Powered by Renewable Energy. In Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems ASPLOS '13. (ACM), pp. 51–64.
47. Kong, F., and Liu, X. (2016). GreenPlanning: Optimal Energy Source Selection and Capacity Planning for Green Datacenters. In 2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS), pp. 1–10.
48. AT&T (2020). Global IP Network Latency. http://ipnetwork.bgtmo.ip.att.net/pws/network_delay.html.
49. Avgerinou, M., Bertoldi, P., and Castellazzi, L. (2017). Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency. Energies, 1470.
50. CAISO (2019). California ISO - Managing Oversupply. http://www.caiso.com/informed/Pages/ManagingOversupply.aspx.
51. Littlefield, J., Roman-White, S., Augustine, D., Pegallapati, A., and Zaimes, G. (2019). Life Cycle Analysis of Natural Gas Extraction and Power Generation (National Energy Technology Laboratory).
52. Gilbert, A.Q., and Sovacool, B.K. (2017). Benchmarking natural gas and coal-fired electricity generation in the United States. Energy *134*, 622–628.
53. Intergovernmental Panel on Climate Change (2014). Annex III: Technology-specific cost and performance parameters. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press).
54. Hundiwale, A. (2016). CAISO Greenhouse Gas Emission Tracking Methodology. https://www.caiso.com/Documents/GreenhouseGasEmissionsTracking-Methodology.pdf.
55. PJM PJM Data Miner 2 - Actual/Schedule Summary Report. https://dataminer2.pjm.com/feed/act_sch_interchange/definition.
56. Jiang, C., Wang, Y., Ou, D., Luo, B., and Shi, W. (2017). Energy Proportional Servers: Where Are We in 2016? In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 1649–1660.
57. Malla, S., and Christensen, K. (2020). The effect of server energy proportionality on data center power oversubscription. Future Gener. Comput. Syst. *104*, 119–130.
58. Uptime Institute (2013). 2013 Data Center Industry Survey. https://uptimeinstitute.com/resources/asset/2013-data-center-industry-survey.
59. Lawrence, A. (2019). Is PUE actually going UP? Uptime Inst. Blog. https://journal.uptimeinstitute.com/is-pue-actually-going-up/.
60. Whitehead, B., Andrews, D., and Shah, A. (2015). The life cycle assessment of a UK data centre. Int. J. Life Cycle Assess. *20*, 332–349.
61. Shah, A.J., Chen, Y., and Bash, C.E. (2012). Sources of variability in data center lifecycle assessment. In 2012 IEEE International Symposium on Sustainable Systems and Technology (ISSST), pp. 1–6.

Figure 1. Historical curtailment of CAISO, 2015-2019.
(A) Solar and wind curtailment by day.
(B) Solar, wind and total curtailment by day and hour.

Figure 2. Hourly GHG intensity of PJM and CAISO, 2016-2019.
During the time when there was curtailment in CAISO, only the intensity of curtailment (assumed proportionally contributed by solar and wind curtailment) is shown. Intensity of imports is not included.

Figure 3. Illustration of the energy consumption profile change of a typical data center in the Baseline Scenario (before workloads migration) and the Migration Scenario (after workloads migration) in a week.
A maximum allowed server utilization rate of 65% during underutilized hours is assumed in this graph as an example.

Figure 4. Estimated net GHG emissions reduction and net abatement cost as a result of assumed maximum server utilization rate and additional data center capacity (2019).
(A) GHG emissions reduction in $KtCO_2e$.
(B) Net abatement cost in \$/tonCO_2e reduction − negative net abatement cost indicates profitable GHG mitigation.
The annotated black lines represent the percentages of total excess VRE absorbed. See Figure S1-S3 for 2016-2018 results.

Figure 1

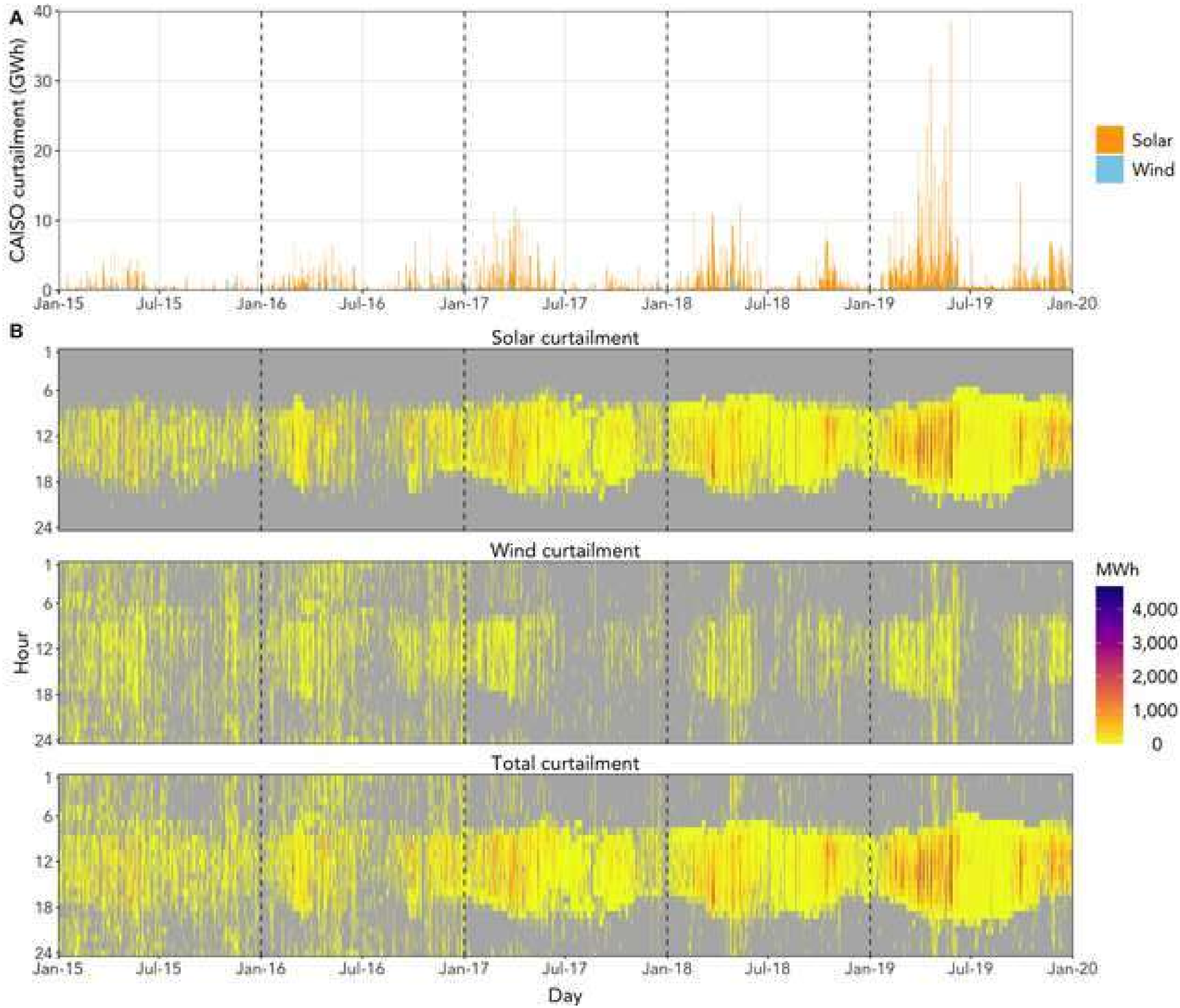

A
CAISO curtailment (GWh)
40
30
20
10
0
Jan-15
Jul-15
Jan-16
Jul-16
Jan-17
Jul-17
Jan-18
Jul-18
Jan-19
Jul-19
Jan-20
Solar
Wind
B
Solar curtailment
Wind curtailment
Total curtailment
Hour
1
6
12
18
24
MWh
4,000
3,000
2,000
1,000
0
Day

Figure 2

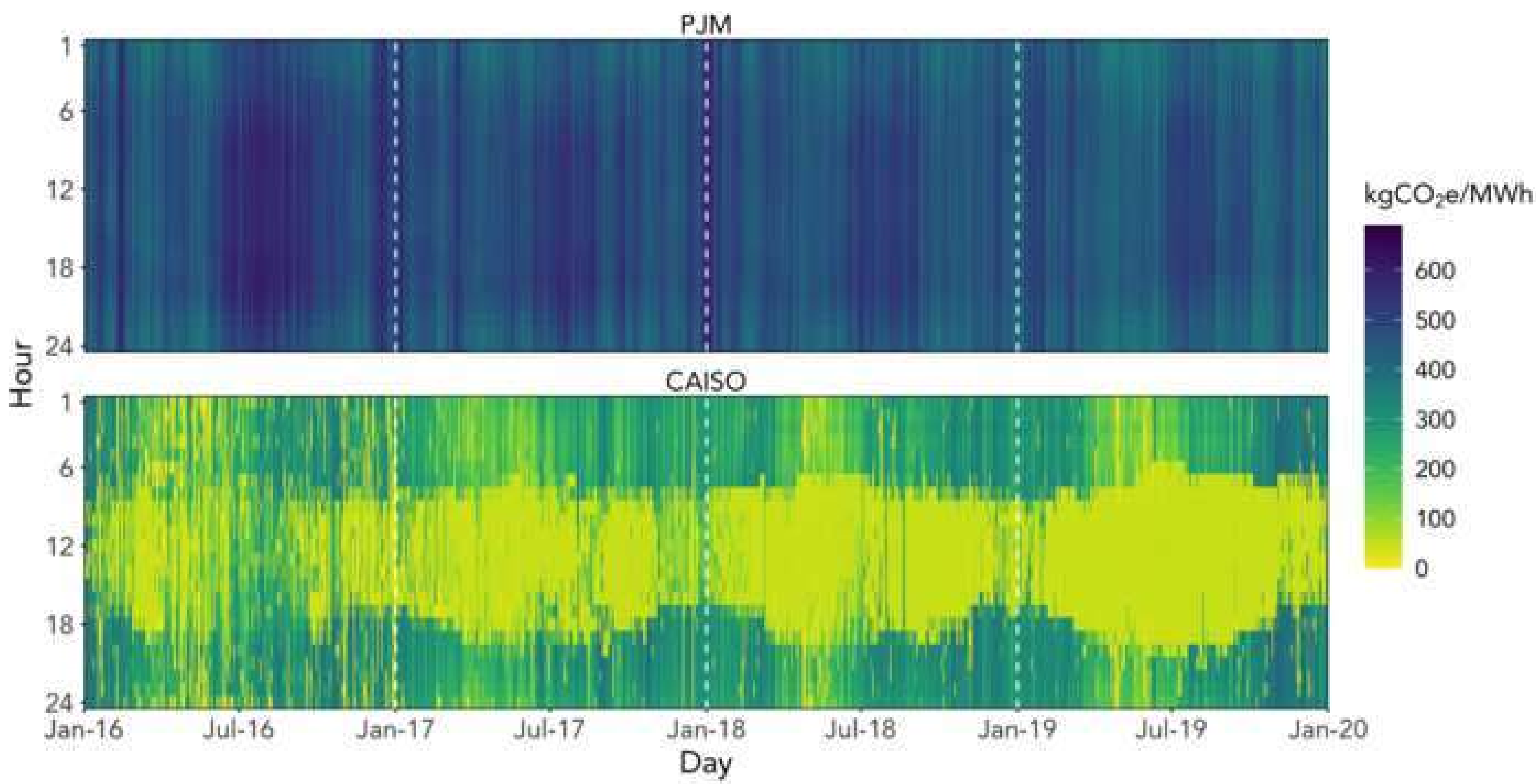
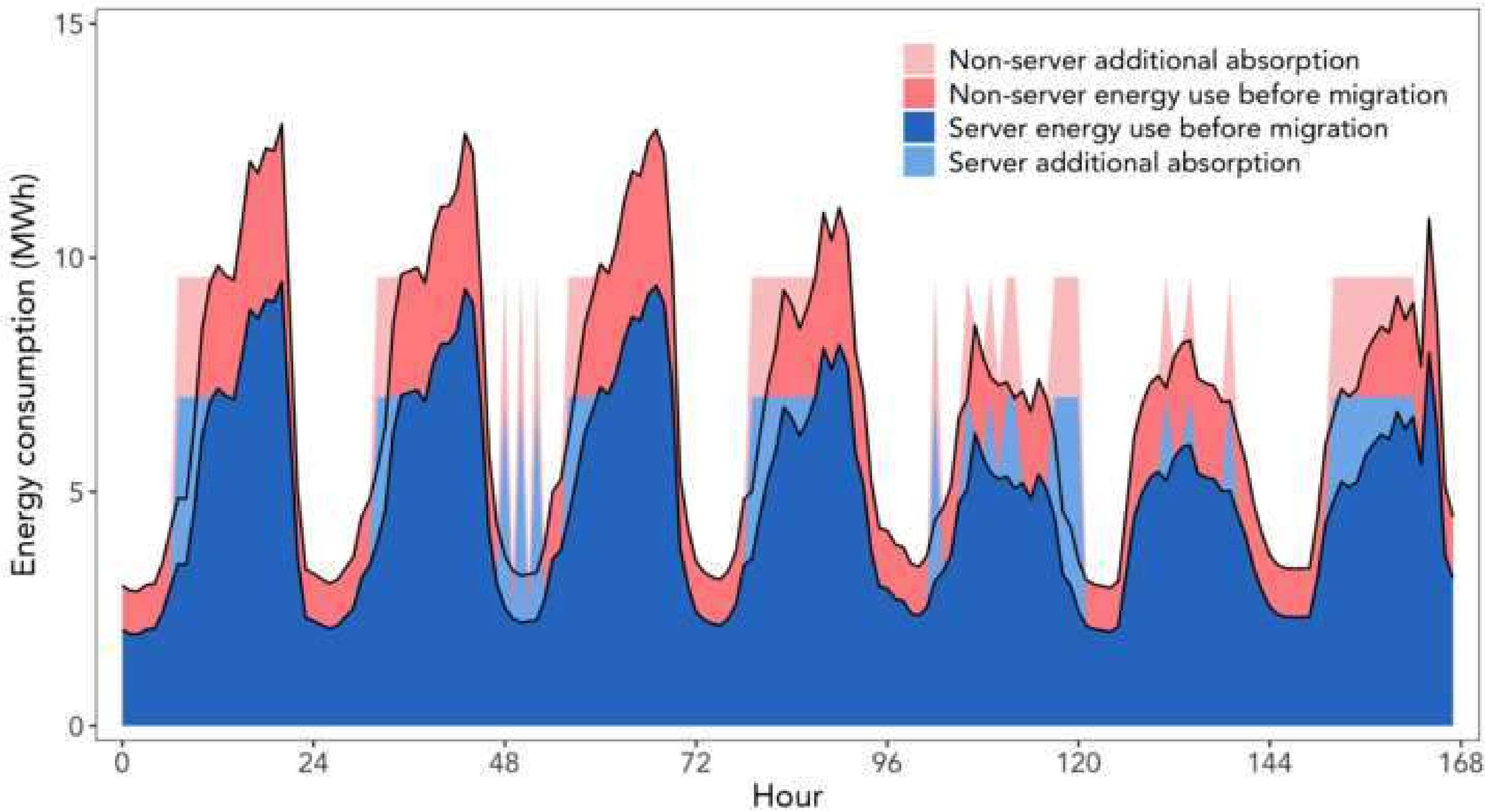
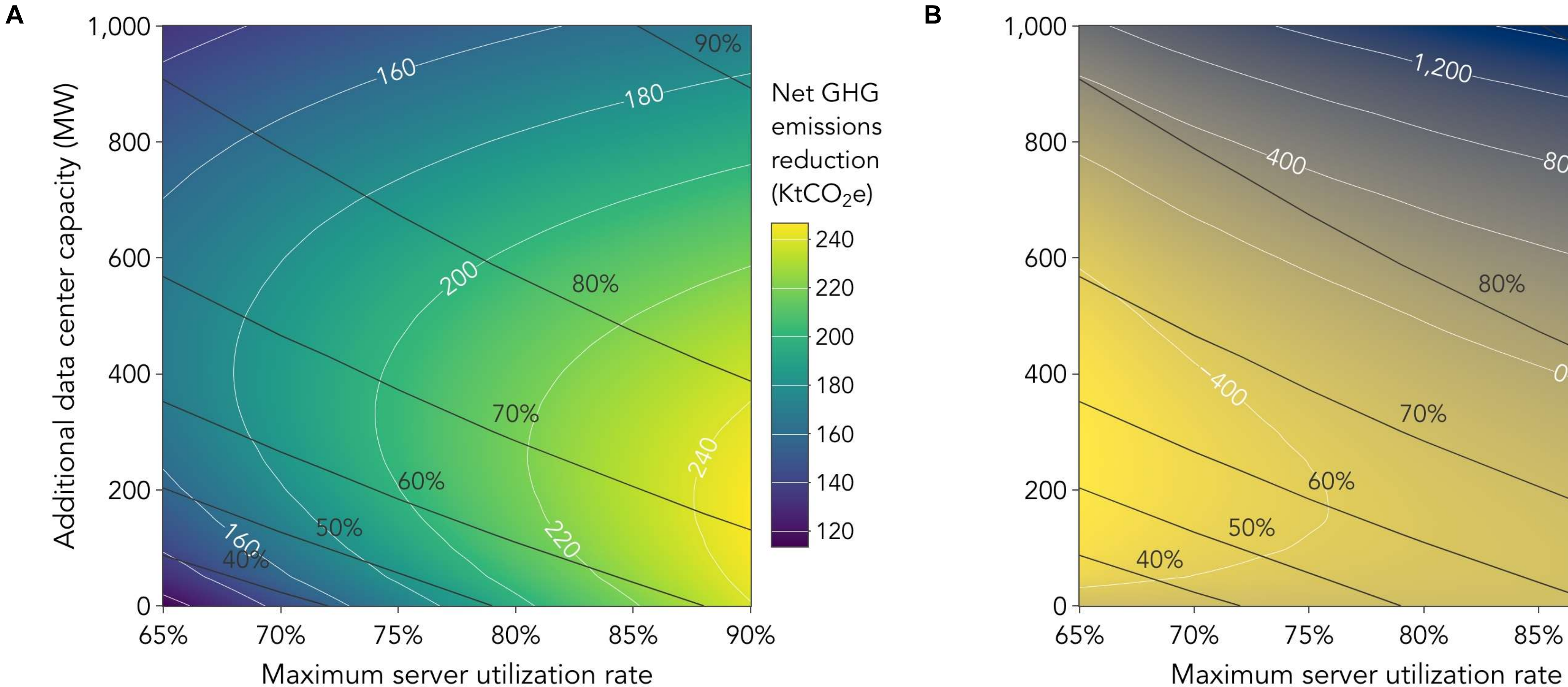

PJM
CAISO
Hour
1
6
12
18
24
Jan-16
Jul-16
Jan-17
Jul-17
Jan-18
Jul-18
Jan-19
Jul-19
Jan-20
Day
kgCO₂e/MWh
600
500
400
300
200
100
0

Figure 3



Energy consumption (MWh)
15
10
5
0
0
24
48
72
96
120
144
168
Hour
Non-server additional absorption
Non-server energy use before migration
Server energy use before migration
Server additional absorption

Figure 4


A
B
Additional data center capacity (MW)
1,000
800
600
400
200
0
65%
70%
75%
80%
85%
90%
Maximum server utilization rate
Net GHG emissions reduction ($KtCO_2e$)
240
220
200
180
160
140
120
90%
80%
70%
60%
50%
40%
160
180
200
220
240
1,200
400
−400
Maximum server utilization rate

Supplemental Data

In this section, we present the data sources and parameter assumptions used for the analysis.

- The life-cycle GHG emissions data of electricity generation by energy source that we applied to derive the hourly GHG intensities are listed in Table S1 with references.
- We used a typical data center energy consumption profile to calculate the excess VRE absorption and resulting GHG emissions reduction. The characteristics of this data center including server power usage, utilization rate (UR) and peak load etc. are presented in Table S2, based on which we calculated the server and non-server energy consumption for the modeled years.
- The assumptions of two key parameters that affect server and non-server energy consumption, Dynamic Range and Power Usage Effectiveness, are shown in Table S3. Combined with the data from Table S2, we developed the server and non-server energy consumption model.
- To calculate the GHG emissions abatement cost of data center workloads migration, we used the cost estimates with breakdown of different cost components for both Zero-Carbon Cloud (ZCC) data centers and traditional data centers in Table S4.

Table S1. Life-cycle GHG emissions of electricity generation technology by energy resource

Energy resource	Life-cycle GHG emissions ($kgCO_2e/MWh$)	Reference
Solar	48	1
Wind	11	
Geothermal	38	
Biomass	230	
Hydro	24	
Nuclear	12	
Biogas	253	2
Natural Gas	537	3
Coal	1,046	4
Oil	733	5
Unspecified	428	6

Table S2. Characteristics of the 10 MW* typical data center used for analysis[7]

Metric	Value
Number of servers	40,000
Idling power per server	120 W
Maximum power per server	250 W
Range of hourly server utilization rate	5.4% - 94.0%
Annual average server utilization rate	40%
Peak total IT load	10 MW
Peak total load	20.7 MW
Estimated annual total energy consumption	114,234 MWh

*IT (or critical) power.

Table S3. Assumptions of annual average Dynamic Range (DR) and Power Usage Effectiveness (PUE) values of data centers served by CAISO[8–11]

Year	PUE	DR
2012	1.59	0.25
2013	1.55	0.23
2014	1.51	0.21
2015	1.47	0.19
2016	1.42	0.17
2017	1.38	0.15
2018	1.34	0.13
2019	1.30	0.11

*Only values assumed for the modeled years 2016-2019 are used in the analysis.

Table S4. Estimated amortized cost for Zero-Carbon Cloud (ZCC) data centers and traditional data centers[12]

Category	ZCC data centers powered by CAISO excess VRE generation (Migration Scenario)	Traditional data centers powered by PJM (Baseline Scenario)
Compute cost ($/W*)	5.18	5.18
Physical facility cost ($/W)	0.50	5.25
Network cost ($/W)	0.20	0.20
Electricity cost (cent/kWh)	0	9.09, 9.18, 9.48, 9.68** (2016-2019, respectively)
Total additional cost ($/W)	0.175	-
- SSD cost	0.075	-
- Battery cost	0.025	-
- Hardware for free cooling	0.075	-

*The unit $/W is dollar per watt of IT power (2015 dollars).
**Historical average retail electricity rate of Virginia[13].

Supplemental Figures

In the following Figures S1-S3, we show the GHG emissions reduction and net abatement cost for the year 2016-2018, respectively.

The total curtailment in CAISO was 307 GWh, 379 GWh and 461 GWh in 2016, 2017 and 2018, respectively. Therefore, the absorption level of the excess VRE (which would otherwise be curtailed) achieved by the same combination of maximum server UR and additional data center capacity decreases over time, comparing across Figures S1-S3. Particularly, the existing data centers alone could reduce 38%-73% of the total cumulative curtailment in CAISO during 2016-2019 (53%-89% in 2016, 46%-81% in 2017, 41%-78% in 2018, and 29%-62% in 2019), with the maximum server UR ranging between 65% and 90% during underutilized hours.

During 2016-2018, the net GHG emissions reduction can be up to 120-150 KtCO_2e per year if the maximum server UR falls in higher range, similar to the observation from 2019 results. The most reduction falls in the absorption level of between 75%-90%, and a further absorption beyond 85% would potentially have negative effects as the embodied emissions of additional data centers would offset the mitigation efforts, which is similar with 2019 results. The total cumulative net GHG emissions reduction during 2016-2019 ranges between 342-647 KtCO_2e for the existing data center capacity, depending on the maximum server UR assumed.

The net abatement cost of the existing data centers alone is -\$202/tCO_2e, -\$210/tCO_2e and -\$226/tCO_2e for 2016, 2017 and 2018, respectively. The slight drop in net abatement cost results from higher availability of excess VRE electricity along the years. The maximum additional data center capacity that permits negative abatement cost is 50-100 MW, 65-120 MW and 90-220 MW for year 2016, 2017 and 2018, respectively, depending on the maximum server UR. It is possible to absorb up to 85% of the excess VRE while still maintain a negative abatement cost for 2017 and 2018, if the maximum server UR can be improved to a range above 85%. Absorption goals beyond the limit would entail net positive abatement costs. For 2016, it is possible to absorb over 80% even 90% of the excess VRE with negative abatement cost when the maximum server UR can reach a level above 75%.

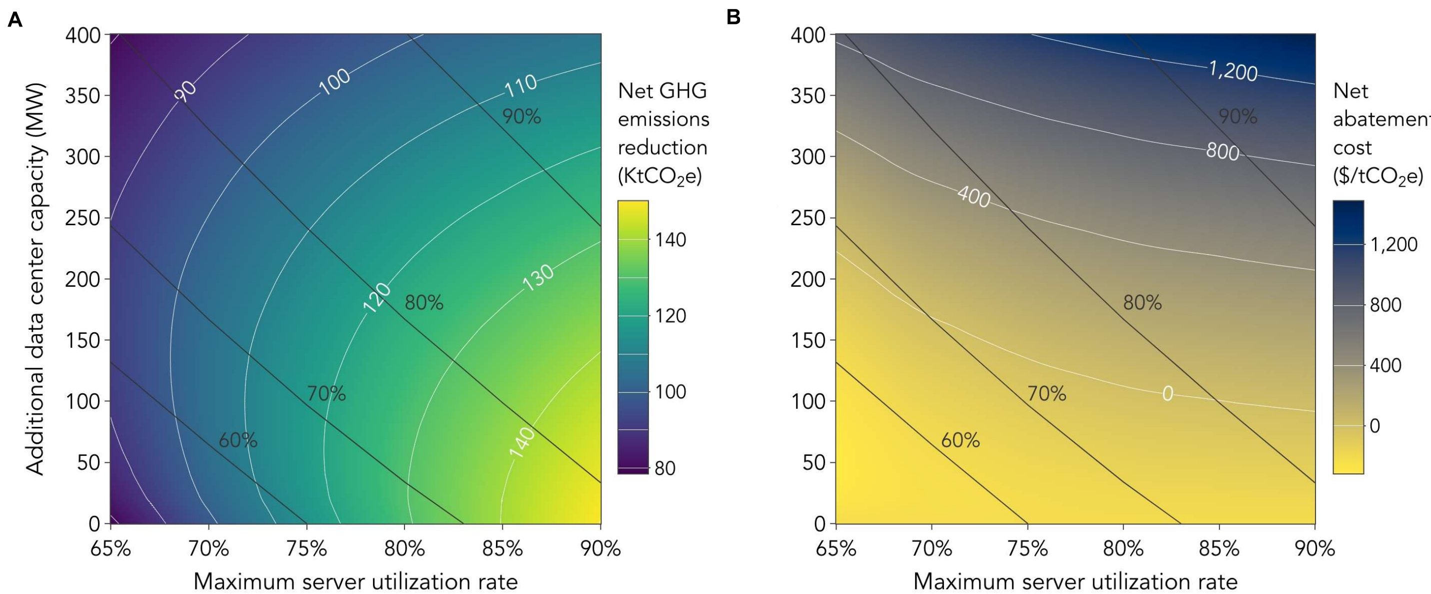


Figure S1. GHG emissions reduction and net abatement cost (2018). Related to Figure 4.
(A) GHG emissions reduction (in $KtCO_2e$).
(B) Net abatement cost (in \$/metric ton CO_2e reduction).
The annotated black lines represent the percentages of yearly total excess VRE absorbed.

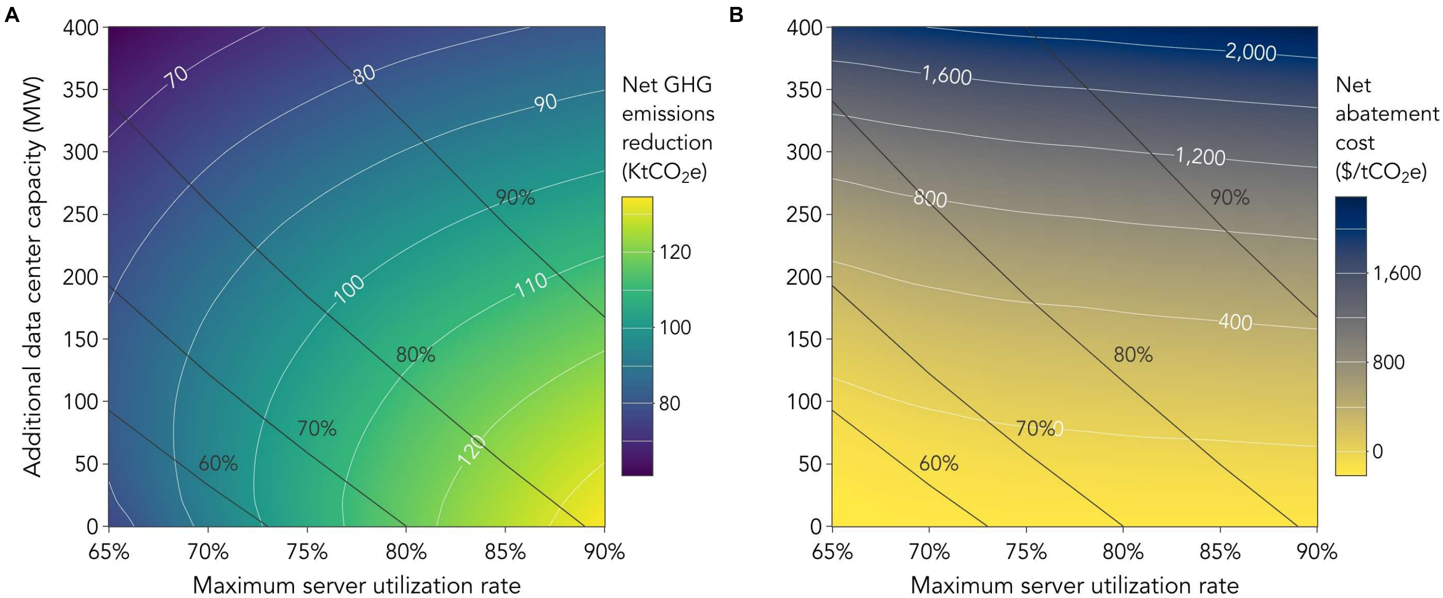


Figure S2. GHG emissions reduction and net abatement cost (2017). Related to Figure 4.
(A) GHG emissions reduction (in $KtCO_2e$).
(B) Net abatement cost (in \$/metric ton CO_2e reduction).
The annotated black lines represent the percentages of yearly total excess VRE absorbed.

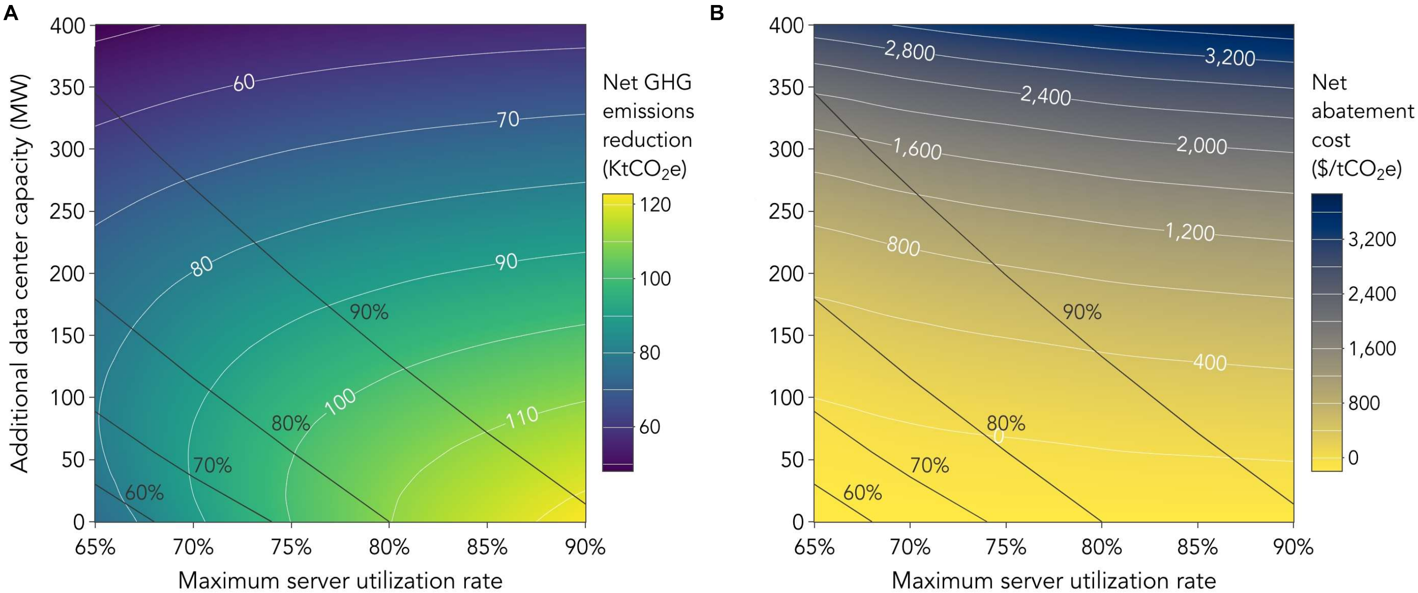


Figure S3. GHG emissions reduction and net abatement cost (2016). Related to Figure 4.
(A) GHG emissions reduction (in $KtCO_2e$).
(B) Net abatement cost (in $/metric ton CO_2e reduction).
The annotated black lines represent the percentages of yearly total excess VRE absorbed.

Supplemental Experimental Procedures

As we discussed in detail in Experimental Procedures, we compute the excess VRE in CAISO that can be absorbed by the remaining capacity of existing data centers and additional data center capacity, by varying the maximum server utilization rate in underutilized hours and allowable additional data enter capacity, and then calculate the resulting GHG emissions reduction and net abatement cost.

Below Supplemental Equations illustrate the computation steps of: (A) absorption of excess VRE, (B) GHG emissions reduction, (C) energy consumption by data centers and (D) net abatement cost, with nomenclature listed in Table S5.

Equations S1 – S21

(A) Absorption of excess VRE by existing and additional data center capacities powered by CAISO

$$Rmn\_Cpty_{exstDC_{h,y}} = \left(DCload_{max,y} - DCload_{h,y}\right) * N\_ExstDC \tag{S1}$$

$$Absp_{exstDC_{h,y}} = \begin{cases} Excess\_VRE_{h,y}, & if\ Excess\_VRE_{h,y} \leq Rmn\_Cpty_{exstDCs_h} \\ Rmn\_Cpty_{exstDCs_{h,y}}, & otherwise. \end{cases} \tag{S2}$$

$$Need\_N_{addDC_{h,y}} = \begin{cases} 0, & if\ Excess\_VRE_{h,y} \leq Rmn\_Cpty_{exstDC_h} \\ \frac{Excess\_VRE_{h,y} - Absp_{exstDC_{h,y}}}{DCload_{max,y}}, & otherwise. \end{cases} \tag{S3}$$

$$Absp_{addDC_{h,y}} = \begin{cases} 0, & if\ Need\_N_{addDC_{h,y}} = 0 \\ Need\_N_{addDC_{h,y}} \times DCload_{max,y}, & if\ 0 < Need\_N_{addDC_{h,y}} \leq Thld\_N_{addDC_y} \\ Thld\_N_{addDC_y} \times DCload_{max,y}, & otherwise. \end{cases} \tag{S4}$$

$$Absp_{exstDC_y} = \textstyle\sum_{h\in[1,8760]} Absp_{exstDCs_{h,y}} \tag{S5}$$

$$Absp_{addDC_y} = \textstyle\sum_{h\in[1,8760]} Absp_{addDCs_{h,y}} \tag{S6}$$

$$Total\_Absp_y = Absp_{exstDC_y} + Absp_{addDC_y} \tag{S7}$$

(B) GHG emissions reduction

$$\Delta GHG_{h,y} = (Intst_{PJM_{h,y}} - Intst_{CAISOexcess_{h,y}}) \times (Absp_{exstDCs_{h,y}} + Absp_{addDCs_{h,y}}) \tag{S8}$$

$$\Delta GHG_y = \textstyle\sum_{h\in[1,8760]} \Delta GHG_{h,y} \tag{S9}$$

(C) Energy consumption by data centers

$$PUE_y = \frac{Energy\_Total_y}{Energy\_IT_y} \tag{S10}$$

$$DR_y = \frac{Power\_Idle_y}{Power\_Max} \tag{S11}$$

$$Power\_Server_{h,y} = Power\_Idle_y + \left(Power\_Max - Power\_Idle_y\right) \times UR_{h,y} \tag{S12}$$

$$Energy\_IT_{h,y} = Power\_Server_{h,y} \times N\_Servers \tag{S13}$$

$$Energy\_nonIT_{h,y} = (m_0 \times UR_{h,y} + b_0) \times m_y \quad s.t.\ \frac{Energy\_Total_y}{Energy\_IT_y} = PUE_y \tag{S14}$$

(D) GHG emissions abatement cost

$$\Delta Fac\_Cost_y = Fac\_Cost_{y,MS} - Fac\_Cost_{y,BS} \tag{S15}$$

$$Fac\_Cost_{y,MS} = Unit\_Fac\_Cost_{MS} \times Total\_Cpty_{addDC_y} \tag{S16}$$

$$Fac\_Cost_{y,BS} = \frac{Unit\_Fac\_Cost_{BS} \times IT\_Power_i}{Energy\_Total_{i,y}} \times Absp_{addDC_y} \tag{S17}$$

$$\Delta Elec\_Cost_y = 0 - Elec\_Rate_{y,BS} \times Total\_Absp_y \tag{S18}$$

$$\Delta Other\_Cost_y = Unit\_Add\_Cost_{MS} \times Total\_Cpty_{addDC_y} - 0 \tag{S19}$$

$$Total\_Abate\_Cost_y = \Delta Fac\_Cost_y + \Delta Elec\_Cost_y + \Delta Other\_Cost_y \tag{S20}$$

$$Net\_Abate\_Cost_y = \frac{Total\_Abate\_Cost_y}{\Delta GHG_y} \tag{S21}$$

Table S5. Nomenclature for supplemental equations

Symbol	Unit	Description
$Rmn\_Cpty_{exstDC_{h,y}}$	MWh	Total remaining capacity of existing data centers powered by CAISO to absorb excess VRE during hour h in year y
$DCload_{max,y}$	MW	Hourly total peak load of a typical 10 MW data center in year y, determined by the assumed maximum server utilization rate which ranges between 65%-90%
$DCload_{h,y}$	MW	Actual existent total load of a typical 10 MW data center at hour h in year y
N_ExstDC	EA	Number of existing 10 MW-equivalent data centers powered by CAISO
$Absp_{exstDC_{h,y}}$	MWh	Absorption of excess VRE by existing data centers powered by CAISO during hour h in year y
$Excess\_VRE_{h,y}$	MWh	CAISO's total excess VRE (i.e. curtailment) at hour h in year y
$Need\_N_{addDC_{h,y}}$	EA	Total number of additional data centers (10 MW IT power) required to absorb all the rest of excess VRE in CAISO that exceeds existing data center capacity during hour h in year y
$Absp_{addDC_{h,y}}$	MWh	Absorption of excess VRE by additional data center capacity powered by CAISO during hour h in year y
$Thld\_N_{addDC_y}$	EA	Threshold number of additional data centers (10 MW critical power) that are allowed to be built in CAISO region in year y
$Absp_{exstDC_y}$	MWh	Annual total absorption of excess VRE by existing data center capacity powered by CAISO in year y
$Absp_{addDC_y}$	MWh	Annual total absorption of excess VRE by additional data center capacity powered by CAISO in year y
$Total\_Absp_y$	MWh	Annual total absorption of excess VRE by both existing and additional data center capacity powered by CAISO in year y
$\Delta GHG_{h,y}$	$kgCO_2e$	GHG emissions reduction at hour h achieved by processing migrated workloads with excess VRE in year y
$Intst_{PJM_{h,y}}$	$kgCO_2e/MWh$	GHG intensity of PJM generation at hour h in year y
$Intst_{CAISOexcess_{h,y}}$	$kgCO_2e/MWh$	GHG intensity of CAISO curtailment at hour h in year y
ΔGHG_y	$KtCO_2e$	Annual total GHG emissions reduction in year y
PUE_y	Unit-less	Annual average Power Usage Effectiveness value assumed for a data center in year y

$Energy\_Total_{i,y}$	MWh	Total energy consumption (IT + non-IT) of a data center in year y
$Energy\_IT_{i,y}$	MWh	Total energy consumption of IT equipment in a data center in year y
DR_y	Unit-less	Dynamic range assumed for servers in year y
$Power\_Server_{h,y}$	W	Power usage of a server at hour h in year y
$Power\_Idle_y$	W	Power usage of a server when they are idling in year y
$Power\_Max_y$	W	Rated power usage of a server in year y
$UR_{h,y}$	%	Utilization rate of the servers at hour h in year y
$N\_Servers$	EA	Number of servers in the data center
$Energy\_nonIT_{h,y}$	MWh	Total energy consumption of non-IT components in a data center at hour h in year y
m_0, b_0, m_y	Unit-less	m_0, b_0 – Slope and intercept of linear model of non-IT vs. server utilization rate, respectively. m_y – Co-efficient of the linear model such that the assumed PUE value is met
$\Delta\, Fac\_Cost_y$	million $	Annual facility cost change due to workloads migration in year y
$Fac\_Cost_{y,MS}$	million $	Amortized facility cost of additional (ZCC) data centers powered by CAISO in year y under Migration Scenario (MS)
$Fac\_Cost_{y,BS}$	million $	Amortized facility cost of traditional data centers powered by PJM in year y under Baseline Scenario (BS)
$Unit\_Fac\_Cost_{MS}$	$/W	Unit cost of amortized facility cost of ZCC data centers under MS
$Total\_Cpty_{addDC_y}$	MW	Total capacity of additional data centers in year y under MS
$Unit\_Fac\_Cost_{BS}$	$/W	Unit cost of amortized facility cost of traditional data centers under BS
IT_Power_i	MW	IT peak load of a typical data center i (i.e. 10 MW)
$\Delta\, Elec\_Cost_y$	million $	Annual electricity cost change due to workloads migration in year y
$Elect\_Rate_{y,BS}$	$/MWh	The average retail electricity rate in year y under BS
$\Delta\, Other\_Cost_y$	million $	Annual other costs change due to workloads migration in year y
$Unit\_Add\_Cost_{MS}$	$/W	Unit annual additional cost of ZCC data centers under MS
$Total\_Abate\_Cost_y$	million $	Total abatement cost in year y under MS
$Net\_Abate\_Cost_y$	$/ton$CO_2$e	Net abatement cost in year y under MS

Supplemental References

1. Intergovernmental Panel on Climate Change (2014). Annex III: Technology-specific cost and performance parameters. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press).

2. Meyer-Aurich, A., Schattauer, A., Hellebrand, H.J., Klauss, H., Plöchl, M., and Berg, W. (2012). Impact of uncertainties on greenhouse gas mitigation potential of biogas production from agricultural resources. Renew. Energy *37*, 277–284.

3. Littlefield, J., Roman-White, S., Augustine, D., Pegallapati, A., and Zaimes, G. (2019). Life Cycle Analysis of Natural Gas Extraction and Power Generation (National Energy Technology Laboratory). DOE/NETL-2019/2039. https://www.netl.doe.gov/energy-analysis/details?id=3198

4. Gilbert, A.Q., and Sovacool, B.K. (2017). Benchmarking natural gas and coal-fired electricity generation in the United States. Energy *134*, 622–628.

5. World Nuclear Association (2011). Comparison of Lifecycle Greenhouse Gas Emissions of Various Electricity Generation Sources. https://www.world-nuclear.org/uploadedFiles/org/WNA/Publications/Working_Group_Reports/comparison_of_lifecycle.pdf.

6. Hundiwale, A. (2016). CAISO Greenhouse Gas Emission Tracking Methodology. https://www.caiso.com/Documents/GreenhouseGasEmissionsTracking-Methodology.pdf.

7. Rahmani, R., Moser, I., and Seyedmahmoudian, M. (2018). A Complete Model for Modular Simulation of Data Centre Power Load. http://arxiv.org/abs/1804.00703

8. Lawrence, A. (2019). Is PUE actually going UP? Uptime Inst. Blog. https://journal.uptimeinstitute.com/is-pue-actually-going-up/.

9. Barroso, L.A., Hölzle, U., and Ranganathan, P. (2019). The Datacenter as a Computer: Designing Warehouse-Scale Machines (Morgan & Claypool).

10. Jiang, C., Wang, Y., Ou, D., Luo, B., and Shi, W. (2017). Energy Proportional Servers: Where Are We in 2016? In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 1649–1660.

11. Malla, S., and Christensen, K. (2020). The effect of server energy proportionality on data center power oversubscription. Future Gener. Comput. Syst. *104*, 119–130.

12. Yang, F., and Chien, A.A. (2018). Large-Scale and Extreme-Scale Computing with Stranded Green Power: Opportunities and Costs. IEEE Trans. Parallel Distrib. Syst. *29*, 1103–1116.

13. EIA - State Electricity Profiles https://www.eia.gov/electricity/state/virginia/.