Towards Carbon-Aware Spatial Computing: Challenges and Opportunities

Bharat Jayaprakash[1], Matthew Eagon[1], Mingzhou Yang[2], William F. Northrop[1], and Shashi Shekhar[2,*]

***Abstract*— Carbon-aware spatial computing (CASC) is focused on reducing the carbon footprint of spatial computing itself and leveraging spatial computing techniques to minimize carbon emissions in other domains. The significance of CASC lies in its potential to mitigate anthropogenic climate change by offering numerous societal applications, such as carbon-aware supply chain development and carbon-aware site selection. CASC is challenging because of the spatiotemporal variability and the high dimensionality of carbon emissions data, involving spatial coordinates and timestamps. Related work, known as carbon-aware computing, mostly focuses on job scheduling of cloud computing, and there is a lack of surveys and review papers detailing the potential of CASC on variant domains and applications. In this paper, we provide the vision of CASC by proposing a taxonomy of sub-domains within CASC and introducing ideas beyond job scheduling, such as carbon-smart site selection. We also briefly review the literature in selected sub-domains and highlight research challenges and opportunities. Given the societal importance of the topic, we encourage the scientific community to use this brief survey to expand the field of study into other related sub-domains and advance CASC more broadly.**

***Index Terms*— Carbon-aware, spatial computing, workload shifting, carbon complexity, climate risk**

I. Introduction

Carbon-aware spatial computing (CASC) has two primary objectives. First, CASC aims to reduce the carbon footprint associated with spatial computing itself. This involves implementing innovative strategies that optimize the allocation of computing resources, such as job scheduling with the aim of consuming clean energy for computation. Second, it offers the potential to minimize carbon emissions and promote sustainability in diverse sectors by leveraging spatial computing techniques. For example, site selection through spatial computing can reduce emissions for many vertical markets ranging from food (e.g., local produce sourcing) and supply chains (e.g., co-locating manufacturing with major consumption sites) to energy, transportation, data centers, and more.

CASC is important because it offers many untapped opportunities to help achieve the all-important goal of achieving net-zero carbon emissions, a global effort to reduce greenhouse gas (GHG) emissions to zero by 2050 set by the Paris Agreement [1]. Specifically, carbon emissions from the transportation and electric power industries have emerged as the major contributors to climate change with annual CO_2 emissions reaching 28% and 25% of overall GHG emissions

[1]Department of Mechanical Engineering, University of Minnesota, MN
[2]Department of Computer Science, University of Minnesota, MN
*Corresponding author.
Email: {jayap015, eagon012, yang7492, wnorthro, shekhar}@umn.edu

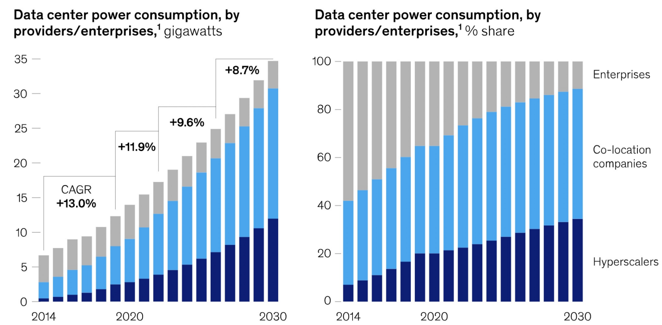

Fig. 1. US data center demand forecast till 2030 [3].

in the U.S., respectively [2]. On the other hand, the demand for computational capabilities also projects multi-fold growth (see Figure 1). All these domains can benefit from the advancement of CASC.

The challenges around CASC arise from the spatiotemporal nature of carbon emission data. For example, emissions related to computation largely depend on the carbon intensity of producing the energy that the computing operation consumes, which consequently depends on the energy mix of the local power grid. Therefore, the carbon intensity not only varies temporally in a given region but also spatially across different regions, as shown by Figures 3 and 4. In addition, the carbon emission data usually have high dimensionality due to the inclusion of spatial coordinates and timestamps.

Despite the promise of CASC, current work on carbon-aware computing mostly focuses only on job scheduling of data centers, and there is a notable lack of surveys and review papers detailing CASC. To address these limitations, this paper makes the following contributions. First, we introduce the domain of CASC and present a hierarchical taxonomy of research areas as a navigational guide (Section II). Second, we review the research and open problems in four sub-domains where optimized spatiotemporal techniques could be applied (Sections II-B.3, II-B.1, II-B.2, and II-B.4). As a prerequisite, we discuss how the carbon footprint of computation is calculated and the particular importance that carbon intensity plays in the calculation (Section II-A). While our review is necessarily brief, we believe it offers a comprehensive overview of major research opportunities in the domain. This is a vision paper, so the following works are out of scope: specific methodologies, experiments, results, etc.

II. Carbon-Aware Spatial Computing

To help conceptualize the research opportunities in CASC, we present a hierarchical taxonomy (Figure 2) with three

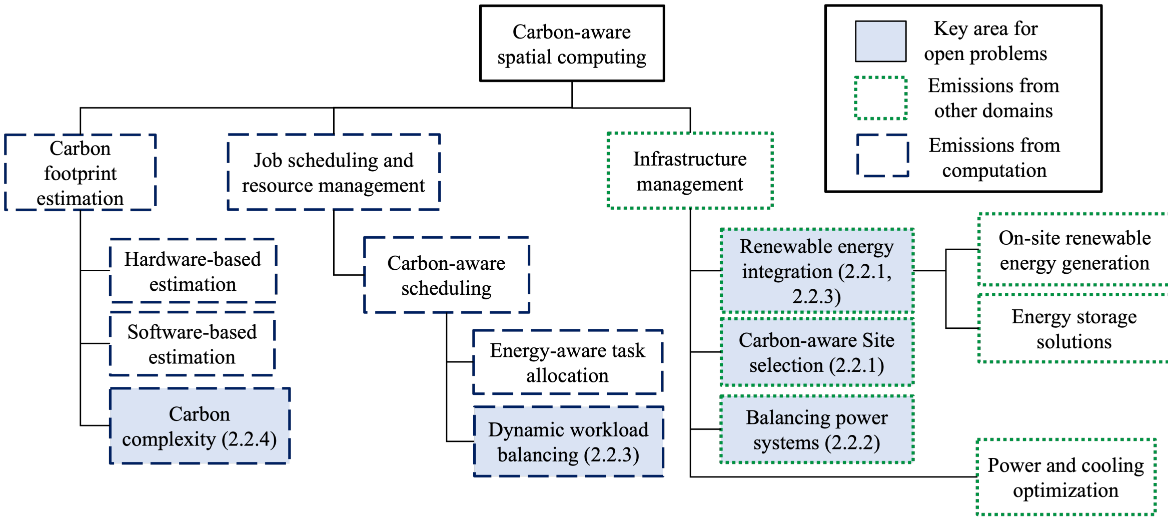

Fig. 2. Taxonomy of carbon-aware computing methods and key areas for open problems (blue).

main branches: carbon footprint estimation, job scheduling and resource management, and infrastructure management.

A. Calculating Carbon Footprint

An essential prerequisite of CASC is quantifying carbon footprint. Quantifying the carbon footprint in the domain applications of CASC (e.g., supply chains, energy grids, etc.) can be multi-faceted due to diverse emission sources (e.g., tailpipe emissions, grid-side emissions, etc.). This section focuses on an example of quantifying the carbon footprint of one domain: computing operations.

The carbon footprint of computing operations depends on the total energy expended during the operation, the associated carbon emissions from energy generation (supply-side emissions), and additional emissions related to auxiliary demands, such as GHG emissions from cooling systems. The total energy expended is influenced by the energy requirements of the computing resources used such as the computing cores, memory units, etc., and the characteristics of the computation to be executed such as running time. The supply-side emissions vary with the time and location of energy generation.

The energy consumption of computing operations can be modeled as a combination of the energy drawn by computing cores (e.g., CPU) and by that of memory. Additionally, if these operations are carried out in data centers, the efficiency of the data center should also be taken into consideration. The efficiency represents how much extra power is necessary to run the facility (e.g., cooling and lighting) [4]. Assuming consistent power and efficiency, the energy consumption of computing operations can be calculated as [4]:

$$E(loc,t) = runtime*(n_c*u_c*P_c+n_m*P_m)*PUE(loc,t) \quad (1)$$

where $runtime$ denotes the running time, n_c denotes the number of cores used, u_c denotes the core usage factor between 0 and 1, P_c denotes the power draw of a computing core, n_m denotes the size of memory available, P_m denotes the power draw of memory, and PUE denotes the efficiency coefficient of the data center and is dependent on the spatial (loc) and temporal (t) coordinates. The carbon intensity of computing can be represented by a function of the location and time of operation, denoted as $CI(loc,t)$. Then, given a quantity of energy E consumed by computing operations, the carbon footprint C is obtained as:

$$C(loc,t) = E(loc,t) * CI(loc,t) \quad (2)$$

By combining Equations 2 and 1, the carbon footprint of computation operations can be calculated by:

$$C = (n_c*u_c*P_c+n_m*P_m)*\int_{t_0}^{t_0+runtime} PUE*CIdt \quad (3)$$

where t_0 denotes the starting timestamp of the computation and loc denotes the location where the computation is performed. In conclusion, Equation 3 shows that the carbon footprint of computing operations is influenced by five factors: run time, power draw from computing cores, power draw from memory, energy efficiency of data centers, and carbon intensity.

Carbon intensity data is crucial for the estimation of carbon footprint using Equation 3 as well as in numerous methodologies implemented to mitigate carbon emissions, which will be discussed in detail in Section II. Until recently, real-time energy production and carbon emissions data from the balancing authorities were not available to consumers. However, authorities have started providing this information to consumers through web APIs. WattTime [6] and electricityMap [5] are two such examples that not only provide historical and real-time data but also forecast data for grids around the world. An electrical grid's marginal emissions rate data [6], [5] represents the emissions rate of electricity generators responding to changes in load on the

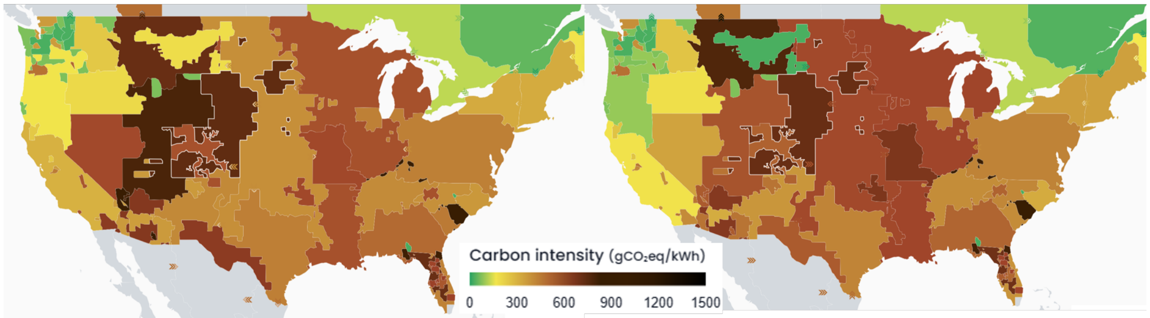

Fig. 3. Electrical grid marginal emissions rate (CI in Equation 2) data of US at different time stamps [5].

local grid by location and time. As shown in Figure 3 [1], the carbon intensity can change drastically for different regions and different time stamps.

B. Challenges and Opportunities

We focus our review on four sub-domains highlighted in blue in Figure 2. We begin our review with the topic of carbon-aware site selection within the sub-domain of infrastructure management

1) Carbon-Aware Site Selection

Carbon-aware site selection is the process of strategically choosing locations for facilities with the goal of minimizing carbon emissions. An important application of carbon-aware site selection is the supply chain industry. Every site within a supply chain network such as suppliers, manufacturing plants, distribution centers, customer locations etc. constitutes a distinct element in the local transportation ecosystem, characterized by infrastructure, transport modes, reliability and cost considerations. By strategically relocating sites or adding new sites to the network, companies can minimize transportation distances, leading to reduced emissions associated with product distribution. Current research works implement site selection criteria such as road connectivity ([8], [9]), water supply ([8], [9]), and proximity to main markets ([10], [8]). Optimizing for component-wise carbon footprint is an unexplored area.

Another application is in the data center market. With data center vacancy rates dropping to record lows [11] and increasing projected demand for high-performance computing (HPC) [3], data center and cloud computing enterprises face significant pressure to establish new facilities or expand existing ones. Current research in the data center placement problem (DCP) focus on minimizing different cost factors and addressing specific objectives. The authors of [12] consider the total data center ownership cost (split into capital and operating expenses), while [13] emphasizes minimizing network costs during disaster failure scenarios. [14] aims to minimize the consumption of dirty energy and data center ownership costs.

[1]Note that Figure 3 shows the spatiotemporal variability of the carbon intensity instead of the absolute values of the carbon emissions. This distinction sets it apart from emission maps such as those in [7], which display aggregated emissions for the research area.

Open Problems: The selection of suitable sites for various vertical markets pose complex challenges. In supply chains, finding locations that optimize the distribution's carbon footprint while satisfying operational demands remains challenging. Sites can be strategically positioned to reduce the transportation of heavy materials that result in higher emissions, which can particularly benefit industries with high transportation-related carbon footprints, such as heavy machinery, electronics, and construction materials. Similarly, cloud computing companies prioritize low operation costs including energy prices. Unfortunately, energy prices do not strongly correlate with the carbon intensity of energy production, necessitating the consideration of carbon intensity in the placement of computing facilities. Developing robust spatial methodologies like clustering and geospatial optimization can integrate these additional factors with carbon intensity into the site selection process.

2) Balancing Power Systems

The idea of balancing the load of a power system (or grids) through strategic management of consumer-side activities can help minimize the reliance on fossil fuel-based power generation, thereby potentially reducing carbon footprint. One such example of strategic management is the Vehicle-to-Grid (V2G) system. V2G is a bi-directional charging system that enables electric vehicles (EVs) to both absorb excess power as well as push energy back to the grid, thus promoting grid stability. Umoren and Shakir [15] introduced the concept of Electric-Vehicle-as-a-service (EVaaS) which focuses on the allocation of EVs in a microgrid with the aim of outage mitigation and grid balancing. However, consumers would be incentivized mostly only by net-positive monetary benefits.

Another such example is the implementation of a technique called spatial workload migration. One study [16] evaluated the economic feasibility of balancing power load by spatially migrating workload using geographically distributed data centers, where one data center can modify its workload by taking on a fractional share of the overall workload. This adjustment affects the power demand and contributes to grid stability. However, this workload shift impacts the load at the other location, potentially leading to high load volatility. The authors explore this phenomenon by focusing on the spatial migration of load to locations that can satisfy unbalanced

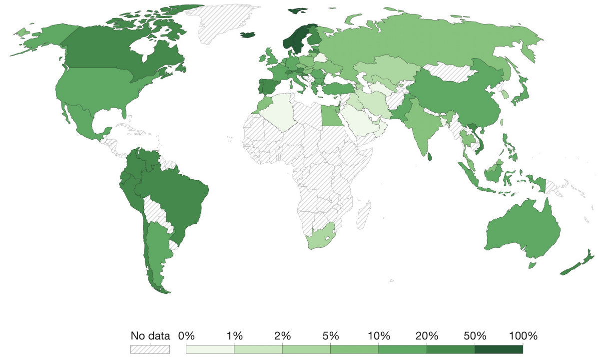

Fig. 4. Share of primary energy from renewable sources including hydropower, solar, wind, geothermal, bioenergy, wave, and tidal, 2021. [23].

demand in a cheaper or less carbon-intensive way. The authors of [17] extend this approach to a larger scale by virtually interconnecting multiple distant markets using data centers, addressing the drawback of limited balancing potential [16].

Open Problems: In seeking to minimize carbon emissions, scheduling algorithms can create bottlenecks as demand surges in areas with the lowest carbon intensity, potentially destabilizing power demand. This often leads to the use of backup power sources to stabilize the grid which can increase the local carbon intensity. Future methods may be able to solve this problem by accounting for the interaction between the decisions made by scheduling algorithms and the grid supplying large computing resources.

3) Spatio-temporal Workload Shifting

Supply-side emissions are the primary contributors to the carbon footprint of computing operations. It has been widely acknowledged that promoting sustainable computing necessitates not only energy-awareness but also carbon-awareness [18]. The grid's energy demand, and consequently the carbon intensity of generation, vary based on consumer behavioral patterns. Daytime typically experiences higher energy demand and high carbon intensity. Additionally, weather conditions impact heating and cooling requirements, and the availability of renewable energy sources. As a result, the practice of temporally shifting computing workloads to low carbon-intensive periods has become prevalent [19], [20], [21]. Google's carbon-intelligent platform reduces emissions by leveraging the temporal flexibility of Google's workloads that tolerate delays of up to 24 hours [19], [20]. The temporal flexibility of such time-shiftable workloads depends on characteristics such as estimated running time, interruptibility, and deferrability [22]. In solving the problem of temporal workload shifting, appropriate weights are assigned to these factors, along with other modeled performance objectives.

In light of the spatial variation in carbon intensity (Figure 4), another approach is spatial workload shifting [24], [25], [26], [27]. It involves migrating computing jobs to cloud centers that utilize comparatively greener energy. However, the trade-offs of migration include increased time delays due to network latencies and data transfer requirements, and additional energy costs for the migration, especially for long-running and memory-intensive jobs.

Open Problems: Existing methods are able to effectively reduce the carbon footprint of computation by scheduling jobs to run in areas and at times where the grid carbon intensity is lower. However, this incentivizes the use of computing resources in countries with cleaner energy grids, which are concentrated in the global north, and may incur overlooked time, space, and energy costs. Methods for dynamic workload balancing should account for these variable costs of transferring jobs. Furthermore, future work could investigate whether it is less expensive in the long term to integrate renewable energy sources into data centers located in areas with high grid carbon intensity. Creating clean data center microgrids within otherwise carbon-intense areas could help reduce overall job times and reduce HPC resource bottlenecks while keeping carbon emissions low.

4) Carbon Complexity

Time complexity, or the total amount of time required by an algorithm to complete its execution, and space complexity, the total space taken by an algorithm with respect to the input size, are widely used metrics for algorithm analysis. Kansal et al. [28] presented a power consumption estimation model for virtual machines using resource usage data at runtime. The authors of [29] then posed the question "Should software applications be redesigned based on energy-optimality?" and consequently introduced an energy complexity model. Various other energy consumption models have also been proposed toward the same goal [30], [31]. However, these models to estimate the energy usage of computational operations differ in important ways from the notions of time and space complexity, which do not account for location and time and can be determined through the relatively simple analysis of algorithms. Space and time complexities are often measured using Big O notation representing the worst-case performance with respect to the input size, but no equivalent notation exists for energy complexity. Furthermore, energy usage is one step removed from carbon emissions, which must take into account the carbon intensity of the local grid and therefore depends on the place and time in which computing resources are being used.

Open Problems: Introducing a new algorithmic definition of energy complexity, and a standard process to determine carbon complexity could be a valuable development for the computing industry. An apt representation of carbon complexity could provide a standardized and quantitative way for facilities to assess the carbon footprint implications of employing different algorithms and computational processes and would facilitate the comparison of computing operations based on their carbon complexity. Current methods of defining energy complexity lack the simplicity and intuition that make space and time complexity popular metrics with which to compare algorithms. It is an open problem to explore ways to explicitly incorporate location and time into computational complexity models. Toward this end, empirical analysis of the relationship between energy utilization and space and time complexity of algorithms run with inputs of various sizes may prove constructive.

III. Conclusion

CASC is concerned with reducing the carbon footprint of spatial computing itself and leveraging spatial computing techniques to minimize carbon emissions in other domains. We identify and elaborate on four key problem areas within CASC that could benefit from optimized spatiotemporal methods. The applications discussed in this paper exemplify the significant impact CASC can have in achieving sustainability goals within various vertical markets. In addition to summarizing recent literature on these topics and highlighting open research problems within these areas, we introduce a hierarchical taxonomy that researchers may consult to navigate this field and explore areas that could lead to other research opportunities.

IV. Future Work

In future work, we will investigate new algorithms and present experimental findings to address the open problems discussed in this paper. Furthermore, we aim to quantify the theoretical benefits of using specific techniques from CASC as well as to initiate discussions around the economic and political feasibility of CASC.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (Grant No. 1901099). The authors would also like to thank Kim Koffolt and the Spatial Computing Research Group at the University of Minnesota for their helpful comments and suggestions.

References

[1] "Paris agreement to the united nations framework convention on climate change." T.I.A.S. No. 16-1104, December 12 2015.

[2] U. E. P. A. E. 430-R-23-002., "Inventory of u.s. greenhouse gas emissions and sinks: 1990-2021.," 2023.

[3] S. Bangalore, A. Bhan, A. Del Miglio, P. Sachdeva, V. Sarma, R. Sharma, and B. Srivathsan, "Investing in the rising data center economy," 2023.

[4] L. Lannelongue, J. Grealey, and M. Inouye, "Green algorithms: quantifying the carbon footprint of computation," *Advanced science*, vol. 8, no. 12, p. 2100707, 2021.

[5] "Electricity maps." https://app.electricitymaps.com/zone/US-NW-WACM?aggregated=false. Accessed: July. 2023.

[6] "Marginal emissions methodology." https://www.watttime.org/marginal-emissions-methodology/. Accessed: Jan. 2023.

[7] K. R. Gurney, J. Liang, R. Patarasuk, Y. Song, J. Huang, and G. Roest, "The vulcan version 3.0 high-resolution fossil fuel co2 emissions for the united states," *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 19, p. e2020JD032974, 2020.

[8] A. Rikalovic, I. Cosic, and D. Lazarevic, "Gis based multi-criteria analysis for industrial site selection," *Procedia Engineering*, vol. 69, pp. 1054–1063, 2014. 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013.

[9] R. K. Singh, N. Chaudhary, and N. Saxena, "Selection of warehouse location for a global supply chain: A case study," *IIMB Management Review*, vol. 30, no. 4, pp. 343–356, 2018.

[10] M. Ashrafzadeh, F. M. Rafiei, N. M. Isfahani, and Z. Zare, "Application of fuzzy topsis method for the selection of warehouse location: A case study," *Interdisciplinary journal of contemporary research in business*, vol. 3, no. 9, pp. 655–671, 2012.

[11] C. Research, "North america data center trends h2 2022." https://www.cbre.com/insights/reports/north-america-data-center-trends-h2-2022, 2023.

[12] I. Goiri, K. Le, J. Guitart, J. Torres, and R. Bianchini, "Intelligent placement of datacenters for internet services," in *2011 31st International Conference on Distributed Computing Systems*, pp. 131–142, 2011.

[13] J. Xiao, B. Wu, X. Jiang, P.-H. Ho, and S. Fu, "Data center network placement and service protection in all-optical mesh networks," in *2013 9th International Conference on the Design of Reliable Communication Networks (DRCN)*, pp. 88–94, 2013.

[14] Y. Wu, M. Tornatore, S. Ferdousi, and B. Mukherjee, "Green data center placement in optical cloud networks," *IEEE Trans. on Green Communications and Networking*, vol. 1, no. 3, pp. 347–357, 2017.

[15] I. Umoren and M. Shakir, "Evaas: A novel on-demand outage mitigation framework for electric vehicle enabled microgrids," pp. 1–6, 12 2018.

[16] G. Fridgen, R. Keller, M. Thimmel, and L. Wederhake, "Shifting load through space–the economics of spatial demand side management using distributed data centers," *Energy Policy*, vol. 109, pp. 400–413, 2017.

[17] M. Thimmel, G. Fridgen, R. Keller, and P. Roevekamp, "Compensating balancing demand by spatial load migration – the case of geographically distributed data centers," *Energy Policy*, vol. 132, pp. 1130–1142, 2019.

[18] J. Guitart, "Toward sustainable data centers: a comprehensive energy management strategy," *Computing*, vol. 99, pp. 597–615, June 2017.

[19] A. Radovanovic, "Our data centers now work harder when the sun shines and wind blows." https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/, 2020.

[20] A. Radovanović, R. Koningstein, I. Schneider, B. Chen, A. Duarte, B. Roy, D. Xiao, M. Haridasan, P. Hung, N. Care, S. Talukdar, E. Mullen, K. Smith, M. Cottman, and W. Cirne, "Carbon-Aware Computing for Datacenters," *IEEE Trans. on Power Systems*, vol. 38, pp. 1270–1280, Mar. 2023.

[21] J. L. Berral, I. Goiri, T. D. Nguyen, R. Gavalda, J. Torres, and R. Bianchini, "Building Green Cloud Services at Low Cost," in *2014 IEEE 34th International Conference on Distributed Computing Systems*, (Madrid, Spain), pp. 449–460, IEEE, June 2014.

[22] P. Wiesner, I. Behnke, D. Scheinert, K. Gontarska, and L. Thamsen, "Let's wait awhile: how temporal workload shifting can reduce carbon emissions in the cloud," in *Proceedings of the 22nd International Middleware Conference*, (Québec city Canada), pp. 260–272, ACM, Dec. 2021.

[23] H. Ritchie, M. Roser, and P. Rosado, "Energy," *Our World in Data*, 2022. https://ourworldindata.org/energy.

[24] F. F. Moghaddam, R. F. Moghaddam, and M. Cheriet, "Carbon-aware distributed cloud: multi-level grouping genetic algorithm," *Cluster Computing*, vol. 18, pp. 477–491, Mar. 2015.

[25] J. Zheng, A. A. Chien, and S. Suh, "Mitigating Curtailment and Carbon Emissions through Load Migration between Data Centers," *Joule*, vol. 4, pp. 2208–2222, Oct. 2020.

[26] Z. Zhou, F. Liu, Y. Xu, R. Zou, H. Xu, J. C. Lui, and H. Jin, "Carbon-aware load balancing for geo-distributed cloud services," in *2013 IEEE 21st International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems*, pp. 232–241, 2013.

[27] G. Neglia, M. Sereno, and G. Bianchi, "Geographical Load Balancing across Green Datacenters: A Mean Field Analysis," *ACM SIGMETRICS Performance Evaluation Review*, vol. 44, pp. 64–69, Sept. 2016.

[28] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, (New York, NY, USA), p. 39–50, Association for Computing Machinery, 2010.

[29] S. Roy, A. Rudra, and A. Verma, "An energy complexity model for algorithms," in *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, (Berkeley California USA), pp. 283–304, ACM, Jan. 2013.

[30] L. Luo, W.-J. Wu, and F. Zhang, "Energy modeling based on cloud data center," *Ruan Jian Xue Bao/Journal of Software*, vol. 25, pp. 1371–1387, 07 2014.

[31] W. Lin, H. Wang, Y. Zhang, D. Qi, J. Wang, and V. Chang, "A cloud

server energy consumption measurement system for heterogeneous cloud environments," *Information Sciences*, vol. 468, 08 2018.