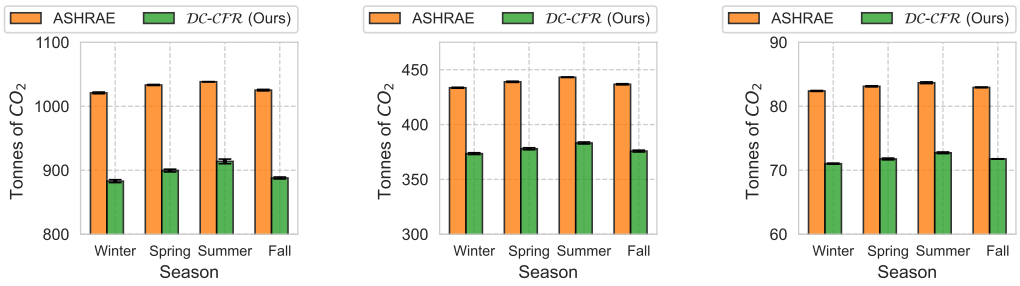
Real-time Carbon Footprint Minimization in Sustainable Data Centers with Reinforcement Learning

Soumyendu Sarkar[†*], **Avisek Naug**[†], **Ricardo Luna Gutierrez**[†], **Antonio Guillen**[†], **Vineet Gundecha**, **Ashwin Ramesh Babu, Cullen Bash**

Hewlett Packard Enterprise
soumyendu.sarkar, avisek.naug, rluna, antonio.guillen,
vineet.gundecha, ashwin.ramesh-babu, cullen.bash @hpe.com

Abstract

As machine learning workloads significantly increase energy consumption, sustainable data centers with low carbon emissions are becoming a top priority for governments and corporations worldwide. There is a pressing need to optimize energy usage in these centers, especially considering factors like cooling, balancing flexible load based on renewable energy availability, and battery storage utilization. The challenge arises due to the interdependencies of these strategies with fluctuating external factors such as weather and grid carbon intensity. Although there's currently no real-time solution that addresses all these aspects, our proposed Data Center Carbon Footprint Reduction ($\mathcal{DC}$-$\mathcal{CFR}$) framework, based on multi-agent Reinforcement Learning (MARL), targets carbon footprint reduction, energy optimization, and cost. Our findings reveal that $\mathcal{DC}$-$\mathcal{CFR}$'s MARL agents efficiently navigate these complexities, optimizing the key metrics in real-time. $\mathcal{DC}$-$\mathcal{CFR}$ reduced carbon emissions, energy consumption, and energy costs by over 13% with EnergyPlus simulation compared to the industry standard ASHRAE controller controlling HVAC for a year in various regions.



(a) Arizona, USA. (b) New York, USA. (c) Washington, USA.

Figure 1: CO_2 generation (Tonnes) for data center control approaches in a 1.2 MWh DC in different locations. ASHRAE is an industry-standard controller for HVAC in DCs.

1 Introduction

Sustainability shifts have led to the need for innovative optimization in data center (DC) operations. Traditional methods, focused mainly on energy and cost, fall short in holistic carbon footprint

*Corresponding author
†These authors contributed equally

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

reduction. Our research focuses on the holistic approach and introduces DC Carbon Footprint Reduction ($\mathcal{DC}$-$\mathcal{CFR}$), a framework that uses Deep Reinforcement Learning (DRL) to optimize energy usage, load shifting, and battery DC operations simultaneously in real-time, based on external and internal parameters like grid Carbon Intensity (CI), exterior weather, workload, etc. $\mathcal{DC}$-$\mathcal{CFR}$ effectively handles complexities of current methods, emphasizing:

- Carbon footprint focused data center (DC) operations, including workload redistribution, efficient cooling, and opportunistic energy storage.
- Real-time controls using shared performance indicators.
- Integration with leading industry simulators.

In a year-long evaluation across multiple DC setups and locations with EnergyPlus (Crawley et al., 2000), $\mathcal{DC}$-$\mathcal{CFR}$ reduced carbon emissions by 14.46%, energy consumption by 14.35%, and energy costs by 13.69%, marking its potential as a transformative tool for sustainable DC operations.

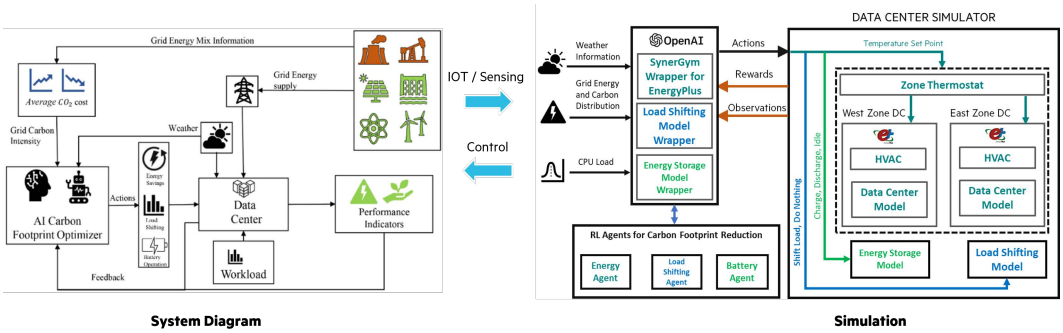


Figure 2: Overview of the physical and digital systems. For this work, we have used EnergyPlus data center simulation with external battery model and our load shifting model.

2 Related Work

Reinforcement Learning (RL) is a game changer for controllers that optimize energy and resource distribution in buildings and data centers. Notable examples include Sinergym for RL evaluation in building models, Energym (Scharnhorst et al., 2021) for climate and energy control assessments, CityLearn (Vázquez-Canteli et al., 2019) for urban energy coordination, and RL-Testbed (Moriyama et al., 2018) for dynamic DC cooling management. All of these rely on EnergyPlus (Crawley et al., 2000), a widely-used energy simulation software. Facebook's "*Carbon Explorer*" (Acun et al., 2023) and Google's "*Carbon-Aware Computing for Datacenters*" (Radovanović et al., 2023) papers both aim to reduce DC carbon emissions. Facebook reallocates DC load to hours with lower carbon footprints and uses energy storage, resulting in approximately a 4% reduction. Meanwhile, Google's workload distribution adjustments led to a 2% reduction. Both strategies hinge on static optimization dependent on accurate long-term forecasts. This reliance makes them susceptible to unpredictabilities like shifting weather patterns, thereby compromising their consistent effectiveness.

3 Problem Definition

Existing heating, ventilation, and air conditioning (HVAC) cooling solutions lack a comprehensive framework that simultaneously optimizes cooling, load shift, and energy in real-time. Our solution addresses this issue to help reduce carbon footprint in DC while optimizing energy consumption for both information technology (IT) and HVAC.

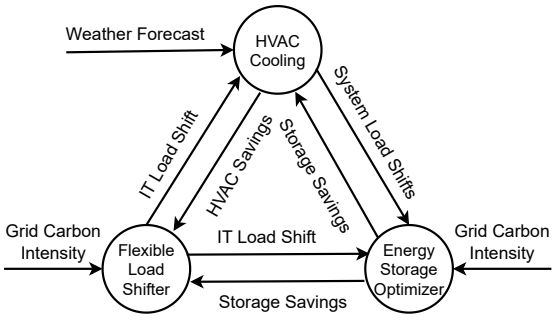


Figure 3: Internal and External System Dependencies

We introduce a strategy to reduce DC energy and carbon footprints using three Markov Decision Processes (MDPs). These MDPs help formulate the different aspects of the carbon footprint reduction problem: server workload shifting, energy-saving via cooling optimization, and energy storage using batteries.

	$\mathbf{MDP_{LS}}$ Load Shifter	$\mathbf{MDP_{E}}$ HVAC Energy	$\mathbf{MDP_{BAT}}$ Battery
State: S_t	Time, DC temperature, IT Load, Unassigned Flexible Load, DC Energy, Carbon Intensity, Battery Charge	Time, DC temperature, Weather, DC Energy, IT Load, HVAC Setpoint	Time, DC Energy, Battery Charge, Carbon Intensity
Action: A_t	Assign Flexible Load, Idle	HVAC Setpoint	Charge, Supply, Idle
Reward: $R_{t+1}(S_t, A_t)$	$0.8 \times r_{LS} + 0.1 \times r_E + 0.1 \times r_{BAT}$	$0.1 \times r_{LS} + 0.8 \times r_E + 0.1 \times r_{BAT}$	$0.1 \times r_{LS} + 0.1 \times r_E + 0.8 \times r_{BAT}$

Table 1: MDPs for Load Shifting ($\mathbf{MDP_{LS}}$), HVAC Energy Optimization ($\mathbf{MDP_{E}}$), and Battery Operation ($\mathbf{MDP_{BAT}}$).
$r_{LS} = -(CO_2\ Footprint + LS_{Penalty})$,
$r_E = -(Total\ Energy\ Consumption \times Cost\ per\ kWh)$, and
$r_{BAT} = -(CO_2\ Footprint)$, where $LS_{Penalty}$ is the scalar value of the unassigned flexible IT workload.

Figure 3 highlights the data flow and decision-making dependencies between these controls. The actions of the flexible load shifter decide the current DC workload, which subsequently decides the amount of HVAC cooling load required for the DC. The cooling optimization process determines the cooling setpoint based on this and the external weather conditions. The grid and the in-house battery banks used for Energy Storage partly provide the net energy required for the IT systems, along with the associated cooling energy. The amount of energy offset provided by the Energy Storage Optimizer depends on the current cooling load, external weather conditions, and the grid CI. This complex interdependence requires a multi-agent approach to optimize the energy and carbon footprint in real-time, considering these relations and the external variables.

The detailed formulation of the individual MDPs is shown in Table 1. We specify the States S_t, Actions A_t, and Reward formulation $R_{t+1}(S_t, A_t)$ for each problem. The variables in the state space are motivated by the discussions based on Figure 3. The Flexible Load Shifter action decides how much workload is assigned or removed from the default workload for a specific time instant t. The HVAC Energy Optimizer decides the cooling setpoint for the IT Room, and the battery agent decides whether to charge the battery storage, stay idle, or offset part of the cooling load using the stored charge.

To help the agents understand the effects of their actions on different performance metrics, we formulate the reward feedback for each agent as a weighted contribution of the incentives for load-shifting r_{LS}, reducing cooling energy r_E and reducing carbon footprint r_{BAT}. The weight terms are chosen to focus on one particular reward signal for each agent and shape cooperative multi-agent RL. Here $LS_{penalty}$ refers to the reduction in incentive when the shifted workloads are not assigned or completed in other parts of the day. This reward-shaping encourages scheduling all workloads as part of the trade-off between carbon footprint and workload assignment.

4 $\mathcal{DC}$-$\mathcal{CFR}$: A Multi-Agent Reinforcement Learning Solution

Addressing the challenges of sustainable DC operation entails navigating interdependent sub-problems, making MARL (Buşoniu et al., 2010) ideal due to its cooperative nature. It allows agents to achieve individual goals while collaborating through a shared reward system.

The $\mathcal{DC}$-$\mathcal{CFR}$ approach, illustrated in Figure 2, aims to reduce the DC's carbon footprint by simultaneously solving the MDPs for Load Shifting (MDP_{LS}), Energy Reduction (MDP_E), and Battery Operation (MDP_{BAT}). This MARL approach accounts for interdependencies among agents.

In Figure 2, the left side represents the system model for the DC, influenced by the workload, weather, grid energy supply components, and the actions of the Carbon Footprint Optimizer. This optimizer considers the current workload, grid CI, and external weather patterns to make real-time decisions on workload, cooling energy optimization, and energy offset using the battery.

To address the dependency problem inherent in this formulation, we map it to a control-simulation framework, illustrated on the right side of the figure. The three control problem MDPs are wrapped using the OpenAI Gym framework and solved using the multi-agent approach, termed "RL Agents for Carbon Footprint Reduction." This approach includes a load shifting agent A_{LS} for MDP_{LS}, an

energy agent A_E for MDP_E, and a battery agent A_{BAT} for MDP_{BAT}. These agents interact with the simulation model through 'rollouts', subsequently training and updating their policies based on the data collected from these rollouts.

In the rollout phase, the agents interact with their respective MDPs, collecting and sharing information as shown in Figure 3. Agent A_{LS} determines the load shifting actions based on its state variables. The effects of A_{LS}'s decisions are then relayed to A_E, determining cooling setpoint parameters for the next time interval. Subsequently, the agent A_{BAT} decides battery actions, considering the net DC IT and cooling energies, the current battery charge, and the grid CI. The weighted reward formulation (Table 1) ensures that each agent is aware of the effects of its actions across the three individual problems. The policy update phase uses the rollout data at intervals to refine agent policies.

5 Experiments

We utilized EnergyPlus (Crawley et al., 2000), an open-source building energy simulation software, to simulate a two-zone DC with the HVAC system. Integration with Python was achieved through the Sinergym framework (Jiménez-Raboso et al., 2021), adopting the OpenAI Gym interface, which facilitates the development of DRL control algorithms. This setup enables dynamic adjustments to cooling set points and DC workload. The load shifting and battery models were borrowed from Acun et al. (2023). The flexible workload was set at 10% for the agent A_{LS}. The three models were integrated into a multi-agent environment wrapper using RLLib (Liang et al., 2018). The battery, A_{BAT}, was set at 50% of the DC's peak hourly consumption, referencing the uninterrupted power supply (UPS) standards.

We explored two MARL methods: the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) (Lowe et al., 2017), which employs decentralized agents that learn from a centralized critic, and the Independent Proximal Policy Algorithm (IPPO) (de Witt et al., 2020) designed for collaborative yet independent agent actions. With a 15-minute action time interval, we achieved granular system control and swift adaptability to changes in the DC. We utilized real-world IT load data from the Alibaba (Cheng et al., 2018) dataset for a more accurate simulation. The external data for training was sourced from New York weather and CI data, incorporating Ornstein-Uhlenbeck (OU) (Espen et al., 2005) noise to enhance generalizability.

We validated the performance of our agents using weather and CI data from diverse locations: Arizona, New York, and Washington. These locations represent a range of climatic conditions and clean energy availabilities. Time-of-use rate plans were adopted to account for energy cost considerations. This multi-location validation highlights the adaptability of our model to various environmental contexts.

6 Results

In this section, we compare the carbon footprint savings achieved by using the individual agents (LS, EO, BAT), their combinations (LS+EO, LS+BAT, EO+BAT), and finally, the $\mathcal{DC}$-$\mathcal{CFR}$ approach against the industry-standard ASHRAE controller for DC cooling. For robustness, each experiment was conducted 20 times with distinct random seeds and weather data variations.

Tables 2, and 3 display the annual reductions in Carbon Footprint across three locations using IPPO and MADDPG, respectively. We also include the energy savings from different approaches in Table 4 using IPPO. The energy consumption results for MADDPG mirrored those with IPPO, making

	Percentage Reduction of Carbon Footprint with IPPO compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	$\mathcal{DC}$-$\mathcal{CFR}$ (Our proposal)
Arizona	7.72 ± 0.18	8.16 ± 0.05	0.25 ± 0.08	13.26 ± 0.07	7.98 ± 0.10	8.46 ± 0.05	**14.36 ± 0.09**
New York	7.13 ± 0.19	8.02 ± 0.06	0.41 ± 0.03	14.39 ± 0.08	7.68 ± 0.20	8.21 ± 0.07	**15.08 ± 0.11**
Washington	4.27 ± 0.20	7.54 ± 0.11	0.46 ± 0.05	13.62 ± 0.08	4.53 ± 0.17	7.78 ± 0.08	**13.96 ± 0.06**

Table 2: **Carbon Footprint Reduction Percentages** using IPPO versus industry-standard ASHRAE. Values represent the mean ± standard deviation, based on 20 varied experiments over one year.

	Percentage Reduction of Carbon Footprint with MADPPG compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	$\mathcal{DC}$-$\mathcal{CFR}$ (Our proposal)
Arizona	8.76 ± 0.50	5.81 ± 2.09	0.24 ± 0.44	11.87 ± 1.36	8.96 ± 0.50	7.21 ± 1.98	**13.40 ± 0.48**
New York	8.02 ± 0.13	5.09 ± 0.09	0.17 ± 0.04	11.32 ± 0.05	8.27 ± 0.11	6.64 ± 0.13	**13.01 ± 0.12**
Washington	8.21 ± 0.05	7.19 ± 0.03	0.32 ± 0.05	12.21 ± 0.12	8.54 ± 0.07	7.68 ± 0.07	**13.27 ± 0.06**

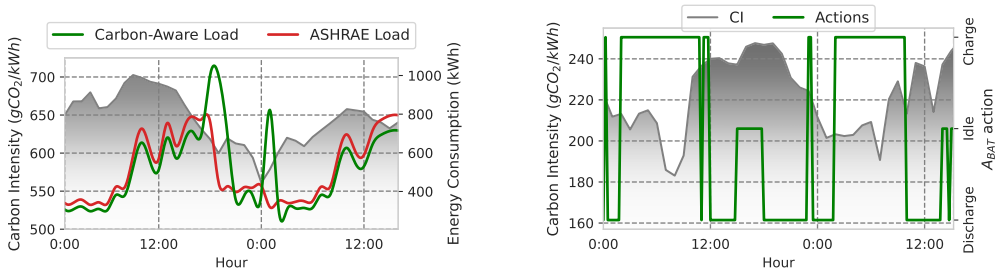
Table 3: **Carbon Footprint Reduction Percentages** using MADDPG compared to industry standard ASHRAE. Values represent the mean ± standard deviation, based on 20 varied experiments over one year.

	Percentage Reduction of Energy Consumption with IPPO compared to ASHRAE Data Center Max Load 1.2MWh Experiment with EnergyPlus for a period of 1 year; Lookahead N = 4 hours						
	Algorithms						
	LS	EO	BAT	LS+EO	LS+BAT	EO+BAT	$\mathcal{DC}$-$\mathcal{CFR}$ (Our proposal)
Arizona	7.11 ± 0.17	8.32 ± 0.04	0.00 ± 0.00	14.28 ± 0.07	7.15 ± 0.09	8.41 ± 0.05	**14.54 ± 0.33**
New York	7.05 ± 0.18	8.07 ± 0.06	0.00 ± 0.00	14.35 ± 0.08	7.12 ± 0.20	8.28 ± 0.08	**14.62 ± 0.07**
Washington	4.38 ± 0.21	7.42 ± 0.11	0.00 ± 0.00	13.78 ± 0.06	4.46 ± 0.18	7.31 ± 0.04	**13.85 ± 0.07**

Table 4: **Energy Reduction Percentages** using IPPO compared to industry standard ASHRAE. Values represent the mean ± standard deviation, based on 20 varied experiments over one year.

a separate table for MADDPG's energy results redundant. The $\mathcal{DC}$-$\mathcal{CFR}$ approach outperforms individual strategies, optimizing both energy and cost. Results highlight substantial savings across all evaluated metrics. It is important to note that agent A_{BAT} does not directly impact power consumption.

The effectiveness of load shifting and battery supply in reducing the carbon footprint is demonstrated in Figure 4. Figure 4(a) illustrates how the $\mathcal{DC}$-$\mathcal{CFR}$ A_{LS} shifts flexible IT loads to hours with low grid CI. Meanwhile, Figure 4(b) shows that A_{BAT} optimizes carbon-aware workload assignment by charging during periods of low CI and discharging during high CI periods.



(a) Load Shift

(b) Energy Storage

Figure 4: Snapshot of: (left) Carbon-Aware Load (Our proposal) against ASHRAE load; (right) Actions taken by the A_{BAT} based on CI.

7 Conclusions

This paper introduces the $\mathcal{DC}$-$\mathcal{CFR}$ framework, a Deep Reinforcement Learning (DRL)-based solution that optimizes data center operations in real-time for energy consumption, load shifting, and battery decision-making. By coordinating three specialized agents, $\mathcal{DC}$-$\mathcal{CFR}$ effectively reduces both carbon emissions and energy use. Unlike traditional methods that depend on extended forecast models, our approach utilizes short-term grid carbon intensity data for agile decision-making. Tested across various data centers in different locations, $\mathcal{DC}$-$\mathcal{CFR}$ consistently outperformed industry benchmarks, including the ASHRAE rule-based controller, regarding carbon reduction, energy efficiency, and cost-effectiveness. We plan to open-source the $\mathcal{DC}$-$\mathcal{CFR}$ framework to encourage broader community adoption. We intend to enhance it with additional data center optimization strategies, further promoting sustainability in large-scale computational environments.

References

D. B. Crawley, L. K. Lawrie, C. O. Pedersen, F. C. Winkelmann, Energy plus: energy simulation program, ASHRAE journal 42 (2000) 49–56.

P. Scharnhorst, B. Schubnel, C. Fernández Bandera, J. Salom, P. Taddeo, M. Boegli, T. Gorecki, Y. Stauffer, A. Peppas, C. Politi, Energym: A building model library for controller benchmarking, Applied Sciences 11 (2021). URL: https://www.mdpi.com/2076-3417/11/8/3518. doi:10.3390/app11083518.

J. R. Vázquez-Canteli, J. Kämpf, G. Henze, Z. Nagy, Citylearn v1.0: An openai gym environment for demand response with deep reinforcement learning, in: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, BuildSys '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 356–357. URL: https://doi.org/10.1145/3360322.3360998. doi:10.1145/3360322.3360998.

T. Moriyama, G. D. Magistris, M. Tatsubori, T. Pham, A. Munawar, R. Tachibana, Reinforcement learning testbed for power-consumption optimization, CoRR abs/1808.10427 (2018). URL: http://arxiv.org/abs/1808.10427. arXiv:1808.10427.

B. Acun, B. Lee, F. Kazhamiaka, K. Maeng, U. Gupta, M. Chakkaravarthy, D. Brooks, C.-J. Wu, Carbon explorer: A holistic framework for designing carbon aware datacenters, in: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2, ACM, 2023. URL: https://doi.org/10.1145/3575693.3575754. doi:10.1145/3575693.3575754.

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, W. Cirne, Carbon-aware computing for datacenters, IEEE Transactions on Power Systems 38 (2023) 1270–1280. URL: https://doi.org/10.1109/tpwrs.2022.3173250. doi:10.1109/tpwrs.2022.3173250.

L. Buşoniu, R. Babuška, B. De Schutter, Multi-agent reinforcement learning: An overview, Innovations in multi-agent systems and applications-1 (2010) 183–221.

J. Jiménez-Raboso, A. Campoy-Nieves, A. Manjavacas-Lucas, J. Gómez-Romero, M. Molina-Solana, Sinergym: A building simulation and control framework for training reinforcement learning agents, in: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Association for Computing Machinery, New York, NY, USA, 2021, p. 319–323. URL: https://doi.org/10.1145/3486611.3488729. doi:10.1145/3486611.3488729.

E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, I. Stoica, RLlib: Abstractions for distributed reinforcement learning, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of *Proceedings of Machine Learning Research*, PMLR, 2018, pp. 3053–3062. URL: https://proceedings.mlr.press/v80/liang18b.html.

R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, I. Mordatch, Multi-agent actor-critic for mixed cooperative-competitive environments, 2017. URL: https://arxiv.org/abs/1706.02275. doi:10.48550/ARXIV.1706.02275.

C. S. de Witt, T. Gupta, D. Makoviichuk, V. Makoviychuk, P. H. S. Torr, M. Sun, S. Whiteson, Is independent learning all you need in the starcraft multi-agent challenge?, 2020. URL: https://arxiv.org/abs/2011.09533. doi:10.48550/ARXIV.2011.09533.

Y. Cheng, Z. Chai, A. Anwar, Characterizing co-located datacenter workloads: An alibaba case study, in: Proceedings of the 9th Asia-Pacific Workshop on Systems, APSys '18, Association for Computing Machinery, New York, NY, USA, 2018. URL: https://doi.org/10.1145/3265723.3265742. doi:10.1145/3265723.3265742.

F. Espen, Benth, J. Šaltytė-Benth, Stochastic modelling of temperature variations with a view towards weather derivatives, 2005. URL: https://doi.org/10.1080/1350486042000271638. doi:10.1080/1350486042000271638. arXiv:https://doi.org/10.1080/1350486042000271638.