Team Control Number

14976

Problem Chosen

B

2024

(HiMCM)

Summary Sheet

Energy Consumption and Carbon Impact of HPC: Insights from Modeling, Forecasting and Optimization

High-Powred Computing (HPC) are increasingly crucial in advancing fields like artificial intelligence, big data analysis, and scientific research, but their growing energy consumption and carbon emissions pose significant environmental challenges. Our team has studied HPC energy consumption, predicted carbon emissions, and proposed strategies to reduce its environmental impact. Using mathematical models, forecasts, and optimizations, we provide insight to help make the HPC industry more sustainable.

In task 1, we estimate global HPC energy consumption to be 520 TWh in 2023, which represents 1.89% of total global electricity consumption. For task 2, we estimate carbon emissions from the HPC sector at 221 million tons in 2023, accounting for 0.67% of global CO2 emissions. They introduce a regionally weighted method to calculate HPC carbon intensity, which avoids biases associated with global averages. Furthermore, they estimate the annual economic and health costs of these emissions to be 48.62 billion USD.

In task 3, using the ARIMA model, we forecast that global HPC energy consumption will increase to 1,219 TWh by 2030 under the baseline scenario, with carbon emissions increasing by 91.69% to 423.99 Mt. We develop a decomposition model to reveal that the primary drivers of growth in energy consumption are compute demand and improvements in energy efficiency.

In task 4, we first use regression analysis to study the effect of increasing renewable energy sources on carbon intensity and HPC carbon emission. We then develop an optimization model to illustrate the trade-off between the environmental benefits of increasing renewable energy adoption and the associated costs of maintaining electricity supply stability.

Finally, we propose six measures to reduce carbon emissions from the HPC sector, including technical and policy measures. We also built an optimization model to demonstrate that prioritizing low-carbon HPC nodes for workload allocation can significantly reduce emissions.

Keywords: Energy Consumption, ARIMA, Carnon Emission, Carbon Intensity, High-Powered Computing

Letter to the Officials

To: United Nations Advisory Board, United Nations Headquarters

Subject: Inclusion of Environmental Impacts of HPC and Developmental Goals and Suggestions for 2030

Dear Members of the United Nations Advisory Board,

We notice that the growing reliance on High-Performance Computing (HPC) for applications like AI, data science, and engineering has significant environmental implications. A key concern is the extensive energy required for HPC operations, which contributes to carbon emissions and exacerbates climate challenges, particularly in regions with limited access to renewable energy sources.

HPC's environmental impact includes high energy use, water for cooling, electronic waste, rare earth depletion, land conflicts, and air and chemical pollution. Addressing these challenges requires balancing HPC's benefits with sustainability. The goal is to drive innovation by processing vast datasets and simulations efficiently while using energy-conscious systems that minimize environmental harm.

To promote sustainable HPC usage, we recommend a combination of technical and policy measures. On the technical front, increasing the share of renewable energy in powering HPC facilities can significantly reduce their environmental impact. Based on our analysis, reducing the U.S. and China's carbon intensity of electricity to Europe's level could cut global HPC emissions by 18.9%. Improving energy efficiency through advanced hardware design, alongside enhancing Power Usage Effectiveness (PUE) by adopting innovative cooling technologies, will further optimize resource utilization. Additionally, leveraging computational networks for efficient workload allocation can minimize energy wastage and operational costs.

On the policy side, introducing carbon taxation for HPC facilities will incentivize operators to adopt greener practices, align with environmental goals, and drive investment in renewable energy and efficient technologies. These measures together can balance the growing demand for HPC with environmental sustainability.

Just as Mahatma Gandhi said, "The earth provides enough to satisfy every man's needs, but not every man's greed," the earth's resources are not enough to satisfy the greed of every HPC. We would be honored if you considered adopting our proposed methods. Your September 2024 report, "The Governing AI for Humanity," addressed AI development, and we suggest including a detailed section on HPC's environmental impact in your 2030 goals. Incorporating our suggestions would be a privilege, and we hope all sectors can collaborate to make HPC more sustainable.

Let's go GREEN with HPC: High Performance, Low Emissions!

Sincerely,

Team #14976

Contents

1 Introduction

High-Powered computing (HPC) is a technology that utilizes supercomputers or data centers to deliver large-scale computational capabilities. Data centers serve as the critical infrastructure for HPC, providing computing, storage, and networking resources. In recent years, the HPC and data center markets have experienced significant growth, driven by demand from scientific research, industrial applications, cloud computing, cryptocurrency mining, and the increasing demand for generative AI models.

However, with the rapid growth of computational capabilities, the energy consumption of HPC systems has increased significantly. Leading supercomputers worldwide now consume tens of megawatts, comparable to the energy demand of large industrial facilities, and most still rely on traditional fossil fuels. This exacerbates energy pressures, significantly increases carbon emissions, and poses critical challenges to environmental sustainability. Therefore, conducting a comprehensive assessment of HPC's energy consumption and its environmental impact, particularly on climate change and other ecological dimensions, is essential.

To address this challenge, our team has performed mathematical modeling and data analysis to describe and forecast HPC energy consumption, carbon emission, and environmental impacts. The workflow of our study is illustrated in Figure 1.

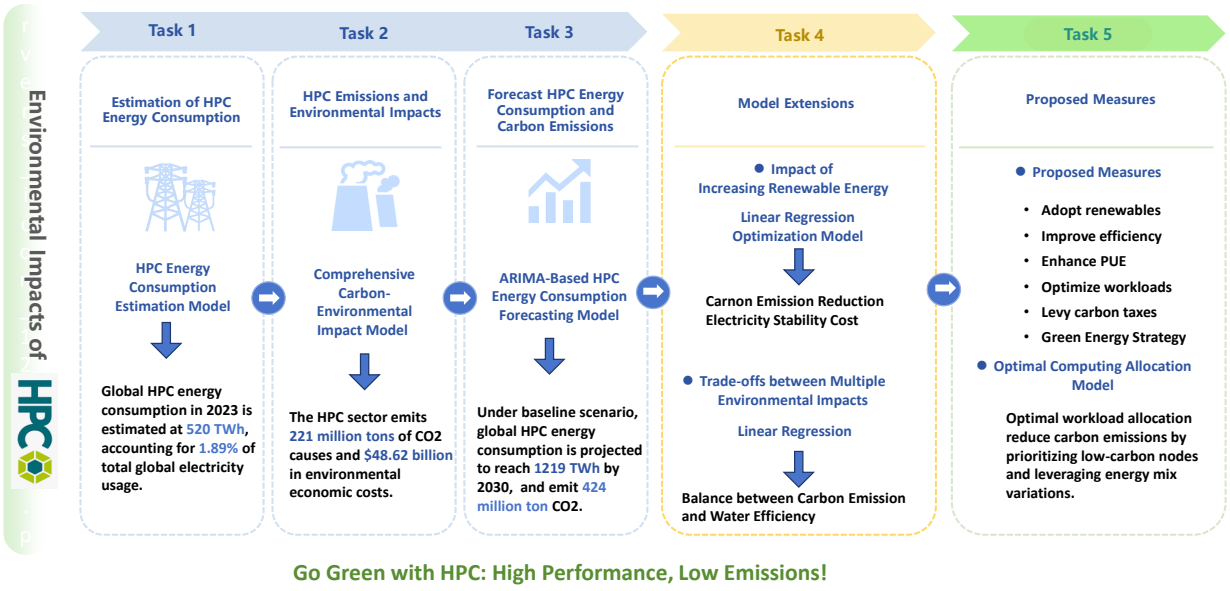

Figure 1 Workflow of Our Work

2 Assumptions and Notations

2.1 General Assumptions

- **Assumption 1: IEA's estimate of global HPC energy consumption for 2022 is reliable.** The IEA is the most authoritative international organization for statistics and research in the energy and electricity sectors. We also assume that from 2022 to 2023, it grows at the same rate as the compound annual growth rate (CAGR) observed from 2015 to 2022.
- **Assumption 2: Google's energy consumption is assumed to maintain a constant proportion with respect to the global HPC sector.** Since Google's energy consumption time series data is publicly available while the IEA has not disclosed the global HPC time series data, this assumption is necessary to derive a proxy series for the latter.
- **Assumption 3: The proportion of renewable energy generation in a country or region**

has a linear relationship with its carbon intensity. The results of our country-level cross-sectional linear regression model and the fitted data support our assumption.

- **Assumption 4: The cost of ensuring power supply stability associated with increasing the proportion of renewable energy is assumed to be a quadratic function of the proportion.** As the proportion of renewable energy increases, the variability and intermittency in power supply also increase, resulting in rapidly accelerating costs for power storage and grid stabilization.

2.2 Notations

Table 1 presents the key concepts and notations with their descriptions.

Table 1 Concepts and Notations for HPC Consumption and Carbon Emissions

Notation	Name	Description	Unit
CAGR	Compound Annual Growth Rate	The yearly growth rate of a variable over a specified period, compounded annually.	%/year
AEC	Annual Energy Consumption	The total amount of energy consumed annually for operations.	TWh
CEC	Compute Energy Consumption	The portion of energy consumption directly used for computing tasks in HPC systems.	TWh
PUE	Power Usage Effectiveness	A metric of energy efficiency representing the ratio of total facility energy to computing energy.	Scalar (0-1)
EE	Energy Efficiency	The efficiency of energy utilization, typically defined as useful output per unit of energy input.	FLOPs/W
UR	Utilization Ratio	The proportion of total computing resources actively used relative to total capacity.	%
CC	Compute Capacity	The overall computational capability of an HPC system, often measured in terms of operations.	FLOPS
CI	Carbon Intensity	The amount of carbon dioxide emissions per unit of energy consumed.	gCO_2/kWh
α	Proportion of Renewable Energy	The share of total energy consumption met by renewable energy sources.	Scalar (0-1)

3 Task 1: Estimation of HPC Energy Consumption

Estimating the annual energy consumption of global HPC systems is not straightforward due to the lack of clear definitions for the HPC sector, as well as the limited availability of energy consumption data for specific HPC facilities. HPC is widely applied in data centers, cryptocurrency mining, and AI, with electricity being its primary energy source. Statistical data provided by governments, consulting firms, and companies mainly cover data centers around the world, as they are the core infrastructure of HPC operations and have relatively readily available energy consumption data. Globally, there are over 8,000 data centers, with approximately 33% located in the United States, 16% in Europe, and almost 10% in China.

According to the IEA's 2024 report, global electricity consumption for data centers, cryptocurrency mining, and AI in 2022 was estimated at 460 TWh, with 120 TWh used for cryptocurrency mining [7]. This indicates a compound annual growth rate (CAGR) of 13.13% from 194 TWh in 2015-2022, as calculated using the formula below.

$$CAGR = \left(\frac{460}{194}\right)^{\frac{1}{2022-2015}} - 1 = 13.13\% \tag{1}$$

In leading large economies such as the United States, China, and the European Union, HPC energy consumption currently contribute approximately 2% to 4% of total electricity consumption.

However, their tendency to be geographically concentrated amplifies their localized effects. For instance, data centers now exceed 10% of electricity use in at least five US states. In particular, the 'Big Five' in the tech industry, Google, Amazon, Meta, Apple, and Microsoft, were alone responsible for an estimated energy consumption of over 90 TWh. This indicates that these five companies made up almost 20% of the total energy used by the HPC industry [7].

Neither the EIA nor other major international organizations have provided an official estimate of HPC energy consumption in 2023. However, assuming a growth rate of 13.13% from 2022 to 2023, we estimate that global HPC energy consumption will be approximately **520 TWh in 2023**, representing 1. 89% of the total world's electricity consumption.

4 Task 2: HPC's Carbon Emissions and Environmental Impact

4.1 Estimation of HPC Carbon Emission

Given the energy consumption and carbon intensity data for N regions, we can apply the following formula to estimate the annual carbon emissions of HPC systems, which accounts for regional differences in electricity consumption and carbon intensity:

$$\begin{aligned} \text{HPC Carbon Emissions} &= \sum_{i=1}^{N} AEC_i \times CI_i = \sum_{i=1}^{N} (w_i \times AEC) \times CI_i \\ &= AEC \times \sum_{i=1}^{N} (w_i \times CI_i) = AEC \times CI_{HPC}, \end{aligned} \tag{2}$$

where, N represents the number of regions considered in the analysis, and AEC_i denotes the annual energy consumption of HPC in region i, expressed as a proportion of the global HPC energy consumption (AEC). The term CI_i corresponds to the carbon intensity of electricity generation in region i (measured in g CO_2/kWh), while w_i reflects the share of HPC energy consumption attributed to region i. Together, these components allow for the calculation of CI_{HPC}, the weighted average carbon intensity specific to HPC operations. By incorporating regional variations in energy consumption and carbon intensity, this formula provides a more accurate estimate of HPC carbon emissions.

Simply using the global average carbon intensity to calculate HPC carbon emissions can lead to significant estimation bias. For example, as demonstrated in the figures below, the proportion of total global electricity consumption [1] and the share of HPC electricity consumption [2] differ markedly by region. In the United States, HPC energy consumption constitutes 51% of the global total, while the country's overall energy consumption represents only 14% of the global total. In contrast, in China, HPC accounts for 16% of global HPC energy consumption, but its overall energy share is as high as 32%. These differences demonstrate why directly applying global average electricity proportions and carbon intensity results in a misrepresentation of the actual HPC carbon intensity.

[1] https://ourworldindata.org
[2] https://www.datacenterdynamics.com

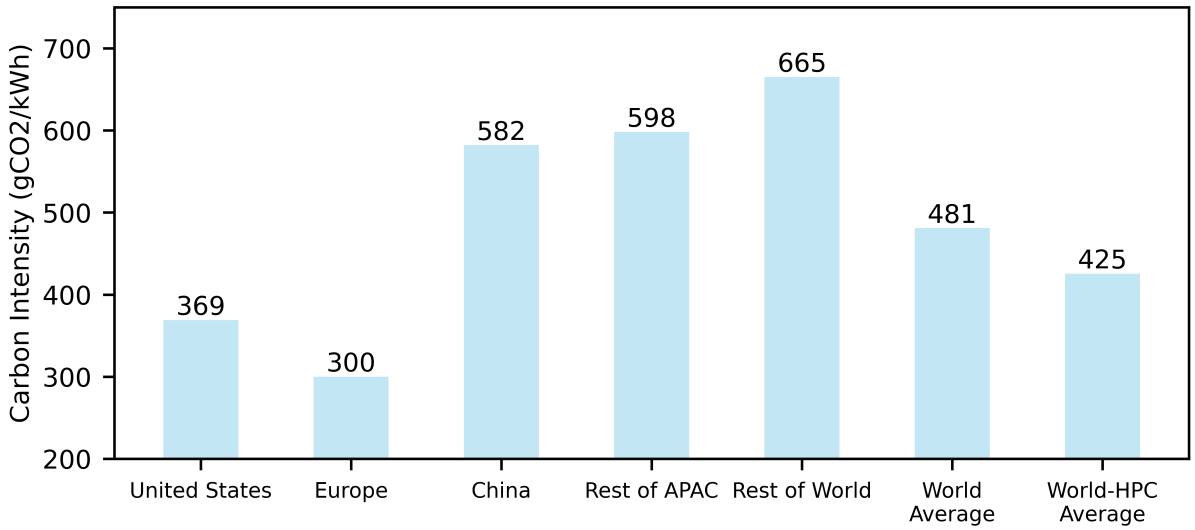

Figure 2 Electricity Carbon Intensity for Different Contries and Regions

Figure 3 Shares of HPC and Overall Energy Consumption of Different Areas

Given an estimated global HPC energy consumption of 520 TWh and a regionally weighted carbon intensity of 425 gCO_2eq/KWh (CI_{HPC}), the annual carbon emissions of global HPC systems are approximately 221 million tons of CO_2. This accounts for 0.67% of global energy-related CO_2 emissions (33 billion tons annually, IEA 2022). If we directly use the global average carbon intensity (481 gCO_2/kWh) to calculate the carbon emissions from HPC, it would result in an overestimation of approximately 13%. This discrepancy arises because the global average carbon intensity fails to account for regional differences in electricity consumption and HPC deployment density, which significantly influence the actual carbon intensity of HPC operations. Therefore, adopting a regionally weighted calculation based on the share of HPC energy consumption provides a more accurate estimation of HPC carbon emissions and avoids the bias introduced by overlooking these regional variations.

4.2 HPC's Environmental Impact

This section evaluates the environmental impact associated with the estimated 221 million tons of CO_2 emissions from the High-Powered Computing (HPC) sector annually. The analysis incorporates the impacts of climate change, economic consequences, and health and ecosystem degradation to provide a comprehensive understanding of the sector's contribution to environmental challenges. The environmental impacts of HPC can be illustrated in Figure 4.

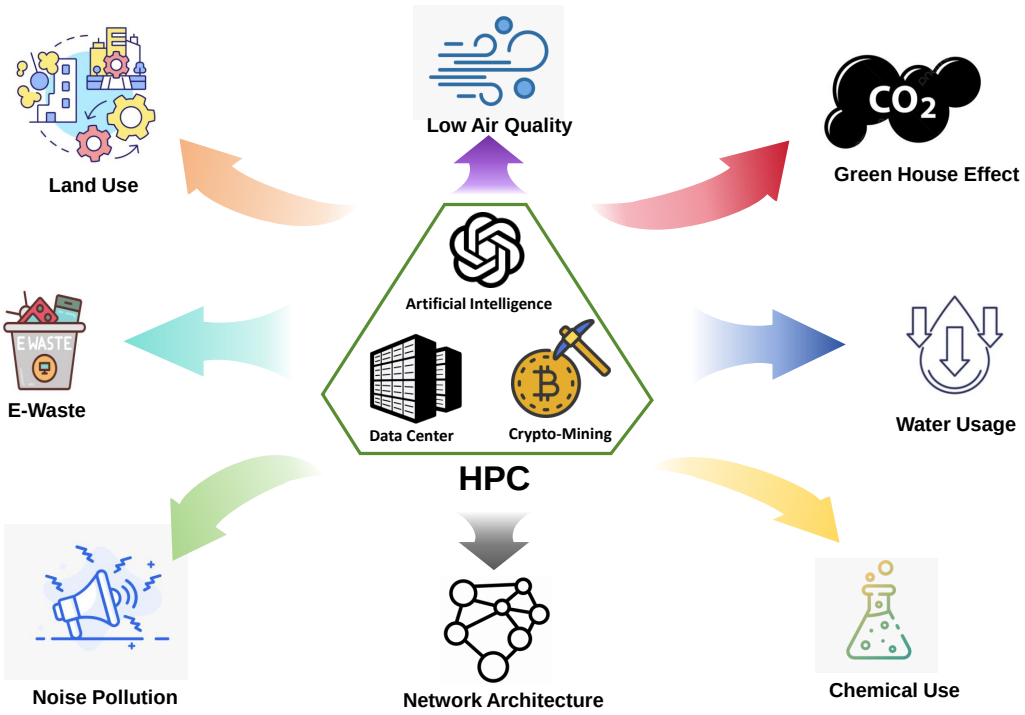

Figure 4 HPC's Environmental Impact

4.2.1 Impact on Global Temperature

Based on findings from the Intergovernmental Panel on Climate Change (IPCC), it is established that every 100 billion tons (100,000 Mt) of CO_2 emissions results in a global temperature increase of approximately 0.45°C [8]. For 221 million tons of CO_2, the contribution to global warming is:

$$\Delta T = \frac{221\,\text{Mt CO}_2}{100,000\,\text{Mt CO}_2} \times 0.45\,^\circ\text{C} \approx 0.0009945\,^\circ\text{C} \tag{3}$$

Thus, the annual CO_2 emissions from the HPC sector are estimated to contribute a marginal but measurable increase of approximately 0.0009945°C to global average temperatures. While this contribution appears small on its own, cumulative emissions from the sector, when combined with emissions from other industries, have a compounding effect on global warming over time.

Research indicates that every 1°C rise in global temperature increases the frequency of extreme weather events by approximately 7%[5]. While 0.0009945°C is a minor increase, cumulative emissions over time will exacerbate such impacts.

4.2.2 Economic Costs of Carbon Emission

The Economic Costs represents the long-term economic damage caused by the emission of one additional ton of CO_2. According to recent estimates from the U.S. Environmental Protection Agency (EPA), the average EC value is approximately 190 USD/ton CO_2[4]. Using this value, the economic cost of the 221 million tons of CO_2 emissions from the HPC sector can be calculated as follows:

$$\text{Economic Cost} = 221\,\text{Mt CO}_2 \times 190\,\$/\text{t CO}_2 = 41.99\,\text{billion USD} \tag{4}$$

The potential economic loss due to 221 million tons of CO_2 emissions is approximately 41.99 billion USD.

4.2.3 Impacts on Health and Ecosystem

Beyond the SCC, CO_2 emissions are also responsible for additional costs arising from their detrimental effects on human health and natural ecosystems. These costs are estimated to range between 10 and 30 USD/ton CO_2, based on methodologies established by the World Bank [20]. Taking the upper limit of this range for a more conservative estimate, the health and ecosystem costs attributable to the HPC sector can be calculated as follows:

$$\text{Health and Ecosystem Cost} = 221\,\text{Mt CO}_2 \times 30\,\$/\text{t CO}_2 = 6.63\,\text{billion USD} \tag{5}$$

Therefore, the annual health and ecosystem costs associated with the HPC sector's emissions are estimated at approximately 6.63 billion USD. This encompasses adverse health outcomes (e.g., respiratory diseases caused by air pollution) and the degradation of critical ecosystem services such as clean water, biodiversity, and carbon sequestration.

Combining the Economic Costs of carbon and the additional Health and Ecosystem Costs provide a comprehensive estimate of the total environmental cost associated with the HPC sector's annual CO_2 emissions. The calculation is as follows:

Table 2 Environmental Impact Parameters and Data Sources

Parameter	Value	Source
Global Temperature Increase	0.45°C/100 billion tons CO_2	IPCC [8]
Extreme Weather Frequency Increase	7%/1°C	Fischer, 2015 [5]
Economic Costs	190 USD/ton CO_2	EPA [4]
Health and Ecosystem Costs	10–30 USD/ton CO_2	World Bank [20]

Table 3 Summary of Environmental Impacts

Impacts	Value
Increased Temperature	0.0009945°C
Increased Extreme Weather Frequency	0.007%
Economic Costs	41.99 billion USD
Health and Ecosystem Costs	6.63 billion USD
Economic Costs + Health and Ecosystem Costs	48.62 billion USD

Thus, the total estimated environmental cost of the HPC sector's annual emissions is approximately 48.62 billion USD. This represents a substantial financial burden, highlighting the need for improved emission reduction strategies and sustainable practices within the sector.

5 Task 3: Forecasting the Global HPC Energy Consumption

5.1 The ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model is commonly used for time series forecasting, especially for non-stationary data [2]. By applying differencing, ARIMA transforms non-stationary data into a stationary form, making it suitable for analysis. In this study, we apply the ARIMA model to forecast global HPC energy consumption and HPC carbon intensity from 2024 to 2030. The general form of an ARIMA (p, d, q) model is expressed as:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{6}$$

where Y_t is the observed value at time t, α is a constant term, $\phi_1, \phi_2, \ldots, \phi_p$ are coefficients of the autoregressive (AR) terms, $\theta_1, \theta_2, \ldots, \theta_q$ are coefficients of the moving average (MA) terms, and ϵ_t is the white noise error term.

5.1.1 Stationarity and Differencing

The first step in ARIMA modeling is to ensure the time series is stationary, meaning its statistical properties, such as mean and variance, remain constant over time. Non-stationary data often exhibit trends or seasonality, violating this assumption. To test for stationarity, we use the Augmented Dickey-Fuller (ADF) test [3]. If the series is non-stationary, differencing is applied until stationarity is achieved, with the number of differencing steps determining the d parameter in the ARIMA model. Once stationarity is confirmed, the next step is to identify the values of p and q values, which represent the orders of the autoregressive and moving average components, respectively.

5.1.2 Identifying p and q Using ACF and PACF

In line with standard practice [2], we use the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the autoregressive (p) and moving average (q) orders of the ARIMA model.

The ACF measures the correlation between the series and its lagged values. Specifically, the autocorrelation at lag k is the correlation between Y_t and Y_{t-k}, adjusted for the series mean. The ACF is crucial for identifying the MA component, and is calculated as:

$$\rho_k = \frac{\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^{T}(Y_t - \bar{Y})^2} \tag{7}$$

where $\bar{Y}$ is the mean of the time series, T is the total number of observations, and k is the lag at which the autocorrelation is calculated.

The Partial Autocorrelation Function (PACF) measures the correlation between Y_t and Y_{t-k}, controlling for the influence of all intervening lags. The PACF helps identify the autoregressive (AR) component of the model. The PACF at lag k is obtained from the regression:

$$Y_t = \sum_{k=1}^{T} \phi_k Y_{t-k} + \epsilon_t \tag{8}$$

where ϕ_k is the partial autocorrelation at lag k, and ϵ_t is the white noise error term. If the PACF cuts off after a specific lag, it suggests an AR model of that order.

5.1.3 Model Estimation

After identifying the values of p, d, and q, the ARIMA model parameters ϕ_p and θ_q are estimated using Maximum Likelihood Estimation or Least Squares Estimation. MLE is commonly used for its efficiency and accuracy. Tools like Stata automate this process, providing quick and reliable parameter estimates. Once the ARIMA model is estimated, it can be used to forecast future values based on historical data, allowing us to predict future trends in global HPC energy consumption.

5.2 Forecasting Future HPC Energy Consumption

The International Energy Agency (IEA) predicts that global HPC electricity demand will increase from 460 TWh in 2022 to 590 TWh by 2026, providing a trend chart for 2019 to 2026. However, specific time series data has not been disclosed, making it challenging to forecast electricity demand beyond 2026. Statistical models typically require historical time series data to estimate model parameters and make future predictions. To address this limitation and conduct forecasts through 2030, we propose a method that assumes global HPC energy consumption is a constant multiple of Google's energy consumption. This assumption is reasonable, as Google is one of

the largest HPC consumers, accounting for approximately 5% of global HPC energy usage in 2023. While this approach may introduce some margin of error, it provides a practical proxy and valuable insights for future research and policy development in the absence of detailed global data. We sourced data on Google's energy consumption from 2011 to 2023 from Statista. In 2023, Google's energy consumption was 25.911 TWh, while global HPC consumption was estimated at 520 TWh, yielding a ratio of 20.07. Using this ratio, we projected Google's energy consumption for 2024–2030 and scaled the projections by 20.07 to estimate global HPC energy consumption over the same period.

5.2.1 Model Identification

Using Stata 17.0's `arima` command, we followed a series of steps. First, we ensured the time series was stationary using the Augmented Dickey-Fuller (ADF) test. For Google's energy consumption data from 2011 to 2023, the ADF test initially indicated non-stationarity. After first-order differencing, the test statistic was -0.154 with a p-value of 0.9438, confirming non-stationarity. However, after applying second-order differencing, the test statistic dropped to -4.019, which is smaller than the 1% critical values, and the p-value of 0.0013 is well below 0.05, as shown in Figure 5. Thus, the use of $d = 2$ was justified.

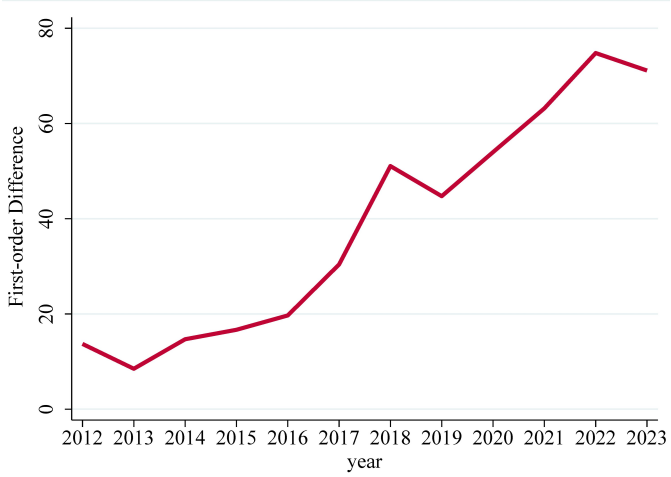

Figure 5 First Order Difference

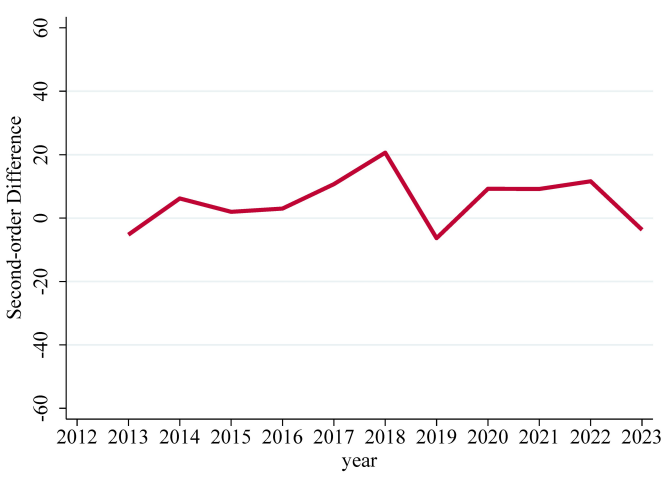

Figure 6 Second Order Difference

With the series stationary, we used the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify the appropriate values of (p) and (q). Figure 6 shows that the ACF plot exhibits autocorrelations tapering off after lag 1, suggesting $q = 1$, while the PACF plot displays a sharp cutoff after lag 1, indicating $p = 1$. Based on these observations, we selected $p = 1$, $d = 2$, and $q = 1$, resulting in a specification of the ARIMA (1,2,1) model for Google's energy consumption data.

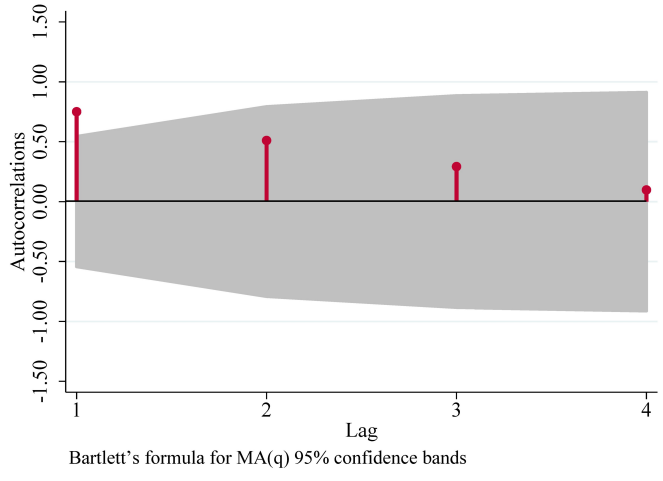

Figure 7 ACF

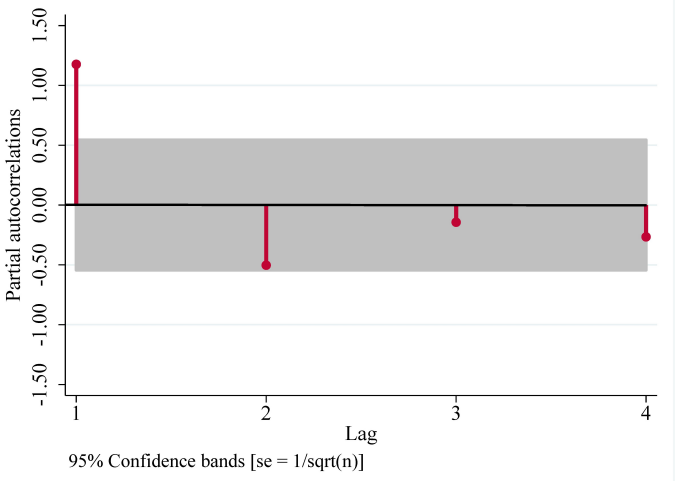

Figure 8 PACF

5.2.2 ARIMA (1,2,1) model estimation

The ARIMA (1,2,1) model is specified as follows:

$$\Delta^2 Y_t = \alpha + \phi_1 \Delta^2 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{9}$$

where $\Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$ is the second difference of Google's energy consumption at time t, α is a constant, ϕ_1 is the coefficient for the first-order autoregressive term (AR(1)), θ_1 is the moving average (MA(1)) coefficient, and ϵ_t is the white noise error term.

The parameters ϕ_1 and θ_1 are estimated using Maximum Likelihood Estimation (MLE). After estimation, diagnostic checks were conducted: the ACF plot of the residuals showed no significant autocorrelations, and the Ljung-Box Q test returned a p-value of 0.9455, confirming that the residuals are white noise, indicating the model adequately fits the data.

5.2.3 Forecasting

The estimated ARIMA (1,2,1) model is represented by the equation:

$$\Delta^2 \hat{Y}_{t+1} = 315.49 + 0.1925 \Delta^2 Y_t - 1.000 \epsilon_t \tag{10}$$

For multi-step forecasting at $t + h$ (where $t \geq h$), the recursive formula is given as:

$$\Delta^2 \hat{Y}_{t+h} = 315.49 + 0.1925 \Delta^2 Y_{t+h-1} \tag{11}$$

for $h \geq 2$, where the future error terms ϵ_{t+h-1} are assumed to be zero.

Figure 9 presents the historical and forecasted global HPC energy consumption from 2011 to 2030. Following the methodology of Masanet et al. (2020), we used the ARIMA (1,2,1) model to generate the baseline forecast[12]. For the low and high case scenarios, we adjusted the baseline by decreasing and increasing it by 67.24%, respectively. The blue line represents historical data from 2011 to 2023, while the orange dashed line illustrates the baseline forecast for 2024 to 2030. Additionally, two alternative scenarios were modeled: the low case and high case, reflecting lower and higher growth rates, respectively.

By 2030, under the baseline scenario, HPC energy consumption is projected to increase by 134.48%, reaching approximately 1219 TWh. In the low case, energy consumption could rise by 67.24% to 870 TWh, while in the high case, it is expected to grow by 201.73% to 1569 TWh. These forecasts provide valuable insights into the potential range of global HPC energy consumption, supporting future energy planning and sustainability strategies.

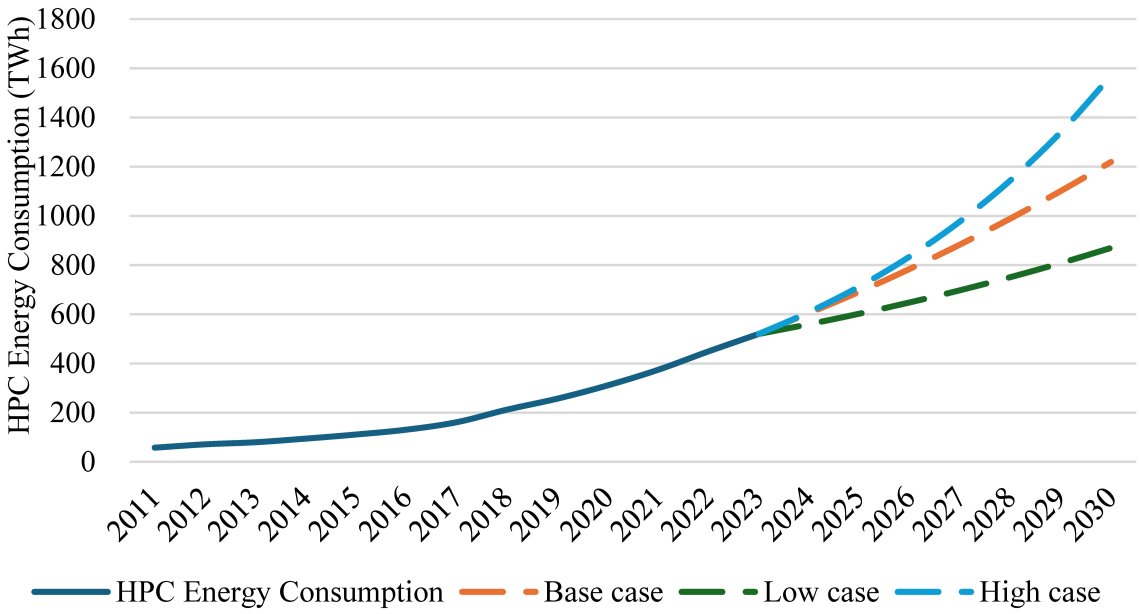

Figure 9 HPC Energy Consumption Forecasting

5.3 Driving Factors: New Computing Demand and Energy Efficiency Improvement

5.3.1 Analysis of driving factors

To understand the driving factors of HPC energy consumption, we developed a model that breaks it down into distinct components based on HPC characteristics.For a certain HPC facility, we calculate both the **Compute Energy Consumption (CEC)** and the **Annual Energy Consumption (AEC)**. The CEC represents the direct energy consumed by HPC equipment for computational tasks, while the AEC accounts for the facility's total energy use, including overheads such as cooling and auxiliary systems.

$$\text{AEC} = \text{CEC} \times \text{PUE} = \frac{\text{Compute Capacity} \times \text{Utilization Ratio} \times \text{Time}}{\text{Energy Efficiency}} \times \text{PUE} \tag{12}$$

where Compute capacity refers to the facility's ability to perform computational tasks, typically measured in terms of FLOPS, Energy Efficiency is the performance of the HPC system, measured in FLOPs/W) and PUE (Power Usage Effectiveness) accounts for the energy consumption overhead of the facility. The decomposition equation shows that AEC variation is driven by four factors: Compute Capacity, Utilization Ratio, Energy Efficiency, and PUE, with Time fixed at one year. Using the approximation that $d\ln x \approx \Delta x/x$ for small relative changes in x, and applying logarithms to both sides of the equation followed by differentiation, we derive the relationship between the annual growth rates of each component as shown below:

$$\frac{\Delta\text{AEC}}{\text{AEC}} = \frac{\Delta\text{Compute Capacity}}{\text{Compute Capacity}} + \frac{\Delta\text{UR}}{\text{UR}} + \frac{\Delta\text{PUE}}{\text{PUE}} - \frac{\Delta\text{Energy Efficiency}}{\text{Energy Efficiency}} \tag{13}$$

A large PUE indicates lower energy efficiency, meaning that a significant portion of the energy is consumed by non-computing operations, such as cooling and facility management, rather than powering the computing infrastructure[1]. Figure 10 presents the PUE data for the global HPC market, which dropped dramatically from 2.5 in 2007 to 1.65 in 2015, reflecting significant improvements in energy efficiency. However, from 2015 to 2023, the PUE remained relatively stable, fluctuating around 1.55 to 1.67. We fit a cubic curve to the historical PUE data to forecast values for 2024–2030, with the predicted trend displayed in the figure below. The PUE is expected to decrease by approximately 2.6% per year from 2023 (1.58) to 2030 (1.4).

Publicly available data on HPC Utilization Ratio are limited, but fragmented reports from various sources indicate that the average HPC utilization rate ranges between 0.5 and 0.8. We predict that this rate is unlikely to change significantly due to constraints such as workload variability and system scheduling limitations.

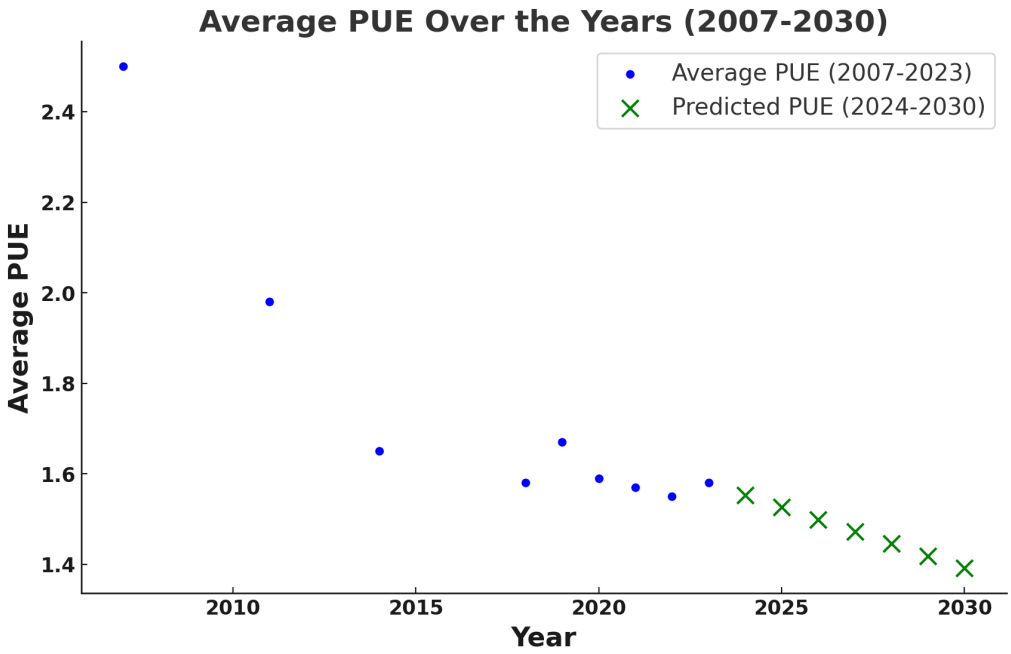

Figure 10 PUE

As changes in PUE and UR tend to be small, the main drivers of future HPC energy consumption growth will be the increasing market demand for computing and improvements in HPC energy efficiency, with the difference in their growth rates determining the overall growth rate of HPC energy consumption. The micro-level data from the HPC top500 list (https://www.top500.org) also confirm our conclusion. The figure below presents the growth curve of total compute capacity and average energy efficiency from the world's 500 largest HPC facilities. During the past decades, despite the rapid improvement in energy efficiency of HPC facilities, total compute capacity has grown at an even faster rate, driving the rapid increase in HPC energy consumption.

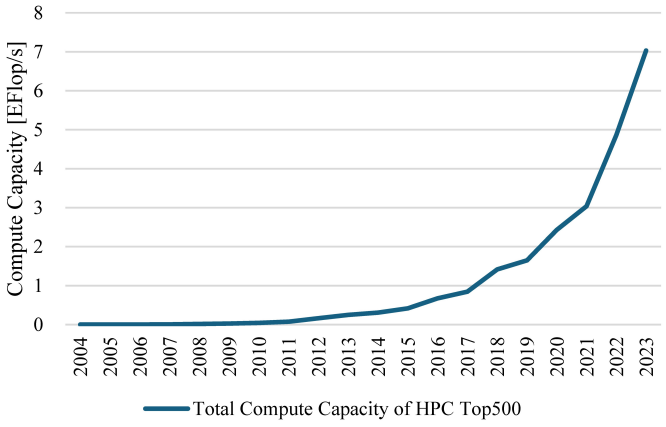

Figure 11 Total HPC Compute Capacity

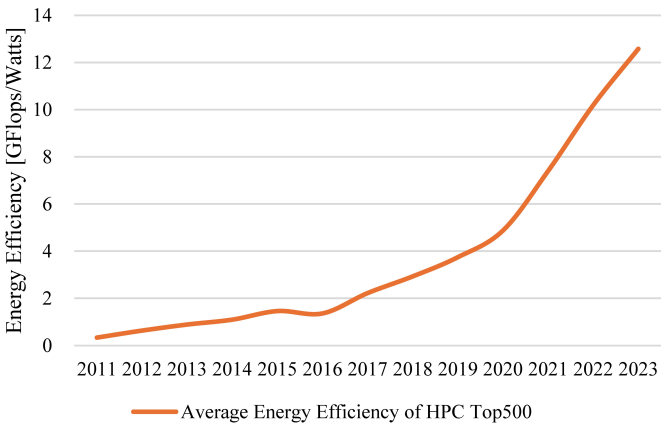

Figure 12 Average HPC Energy Efficiency

5.3.2 LLM growth and energy Consumption

Rapid growth in artificial intelligence, machine learning, big data analytics, and scientific simulations is driving an unprecedented demand for HPC systems [17, 13, 9, 16]. As shown in

Figure 13, a single ChatGPT query consumes nearly 10 times more electricity (2.9 Wh) than a Google search (0.3 Wh) [3], according to the International Energy Agency. Additionally, ChatGPT's monthly visits surged sixfold, from 250 million in December 2022 to 1.6 billion in December 2023 [4].

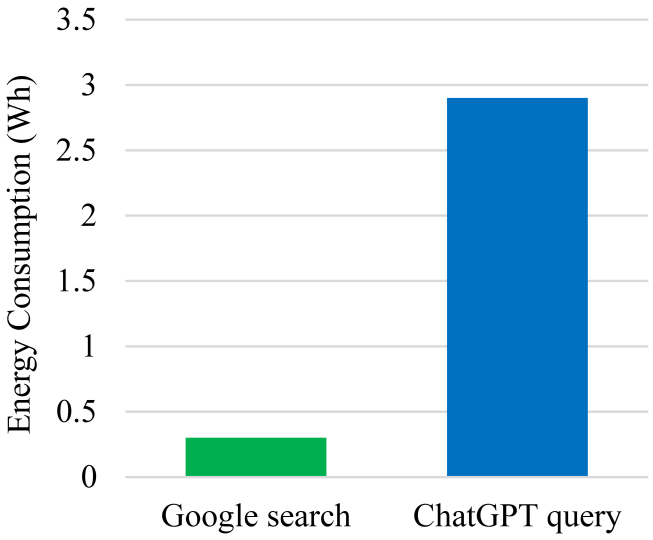

Figure 13
Energy Consumption (Google vs ChatGPT)

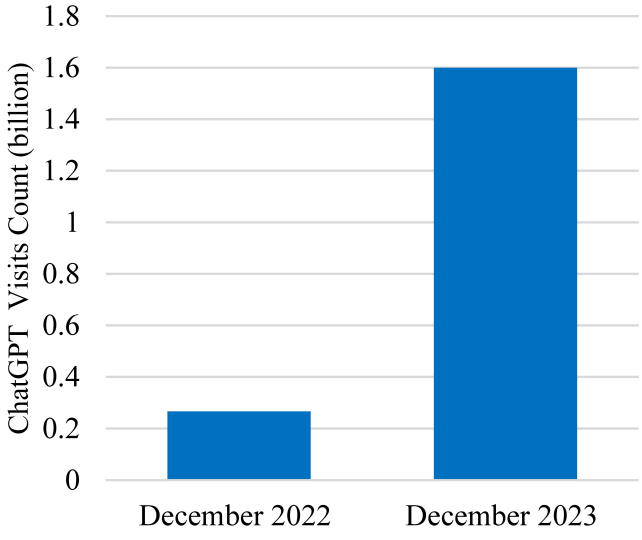

Figure 14
ChatGPT Visits

5.4 Outlook for HPC Carbon Emissions in 2030

Before forecasting global carbon emissions, it is essential to first predict the future global HPC carbon intensity. Using the time-series data on the global HPC carbon intensity constructed in the previous section, we employ the ARIMA model and follow similar procedures. Figure 15 illustrates the historical trend and future projections of global HPC carbon intensity in electricity generation. The orange line shows that from 2000 to 2023, carbon intensity remained relatively stable with slight fluctuations but started to decline after 2020. The projections for 2024 to 2030 are represented by three scenarios: in the Base case, the global intensity of HPC carbon is projected to decrease by 18.25%, reaching 347.72 gCO_2/kWh by 2030. In the Low case, the intensity decreases by 9.13% to approximately 386.54 gCO_2/kWh, while in the High case, it drops by 27.38%, reaching around 308.91 gCO_2/kWh.

[3] https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand
[4] https://seo.ai/blog/how-many-users-does-chatgpt-have

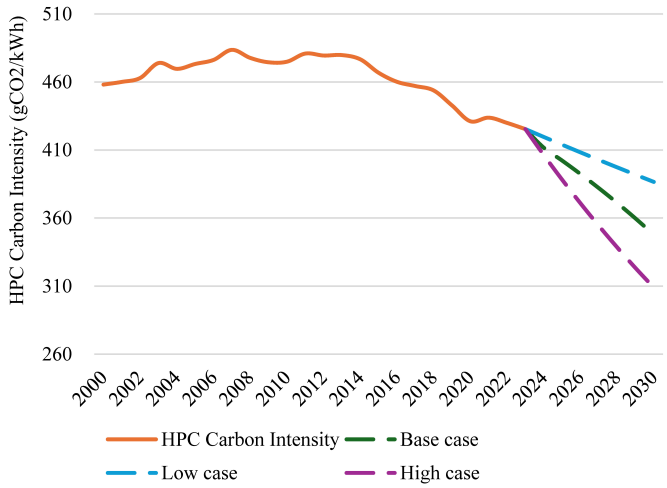

Figure 15 Future HPC Carbon Intensity

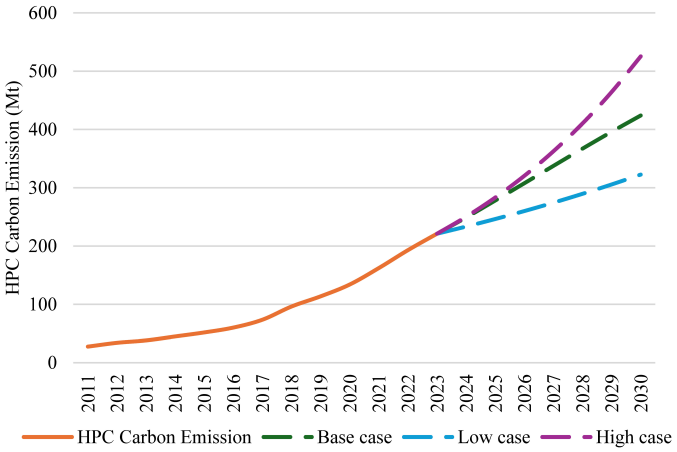

Figure 16 Future HPC Carbon Emission

Finally, according to the formula $CE = EC \times CI$, where CE represents global HPC carbon emissions, EC is HPC energy consumption, and CI is HPC carbon intensity, we can estimate global HPC carbon emissions. Figure 16 illustrates the historical (2011–2023) and forecasted (2024–2030) global HPC carbon emissions. The solid orange line shows historical data, while the dashed lines represent three forecast scenarios. In the base case, emissions are projected to increase by 91.69%, reaching approximately 423.99 Mt by 2030. The low case forecasts a slower increase to 322.59 Mt, a 45.84% increase, while the high case anticipates emissions rising to 525.39 Mt, a 137.53% increase.

6 Task 4: Model Extensions

6.1 Effects of Increasing Renewable Energy

We extend the model to analyze the impact of increasing the portion of renewable energy and carbon emission reductions. We first investigates how increasing the share of renewable energy in electricity generation affects electricity carbon intensity. The goal is to quantify the relationship between renewable energy penetration and the reduction of carbon emissions in the power sector.

We conducted an Ordinary Least Squares (OLS) regression to examine the relationship between electricity carbon intensity and the share of renewable energy. The results of the regression indicate that when renewable energy accounts for 0% of electricity generation, carbon intensity is 640.35 g CO_2/kWh. Furthermore, for every 1% increase in renewable energy share, carbon intensity decreases by an average of 6.32 g CO_2/kWh. The final estimated regression equation is:

$$\text{Carbon Intensity} = 640.35 - 6.32 \cdot \text{Renewables Proportion} \tag{14}$$

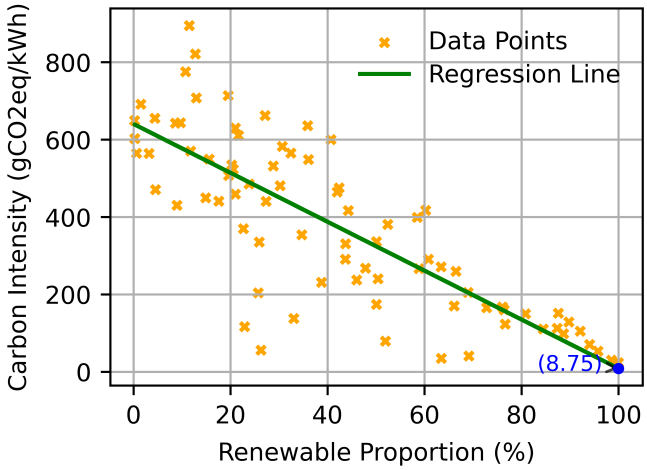

Figure 17
Carbon Intensity and Renewable Proportion

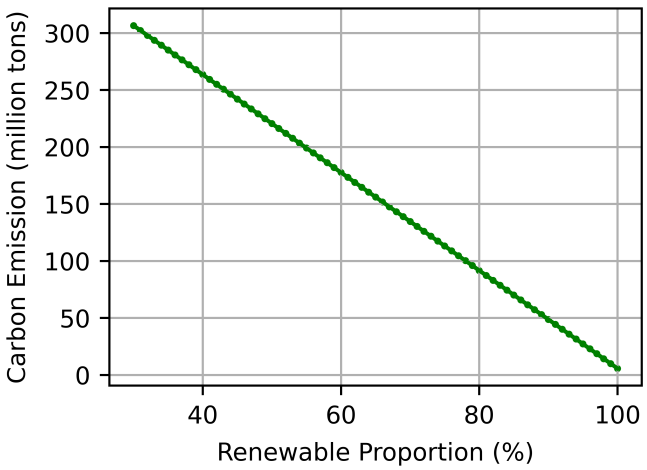

Figure 18
Carbon Emission and Renewable Proportion

The negative slope demonstrates a significant inverse relationship between the share of renewable energy and the intensity of carbon. For example, increasing the share of renewable energy from 30% to 40% would reduce the intensity of carbon by approximately 63.2 gCO_2/kWh. This result highlights the critical role of renewable energy in reducing emissions, as renewable energy replaces high-carbon fossil fuel sources. The findings emphasize the importance of policies that promote the adoption of renewable energy to achieve the de-carbonization goals.

6.2 Trade-off between Renewable Energy Proportion and Energy Supply Stability Cost

For the HPC industry, the widespread adoption of renewable energy can significantly reduce its carbon emission intensity. However, renewable energy sources such as wind and solar are inherently intermittent, which conflicts with the stringent requirements of HPC systems for power supply stability. To address this issue, energy storage systems must be introduced to store surplus renewable energy for use during periods of insufficient generation. This energy configuration requires a balance between the carbon reduction benefits of renewable energy and the costs of energy storage, to determine the optimal proportion of renewable energy. To this end, we have designed and applied an optimization model aimed at minimizing the total monetized costs of the HPC system. This model comprehensively accounts for the social monetization cost of carbon emissions, energy costs, and storage costs.

To balance economic costs and environmental benefits, we develop an optimization model to determine the optimal proportion of renewable energy α. The objective is to minimize the total cost consists of three components $(\text{Cost}_{\text{carbon}} + \text{Cost}_{\text{energy}} + \text{Cost}_{\text{storage}})$:

- **Social monetized cost of carbon emissions:** The environmental damage caused by carbon emissions from fossil fuel generation. To simplify the model, we assume the carbon emission for renewable energy is zero.

$$\text{Cost}_{\text{carbon}} = (1-\alpha) \cdot C_{\text{intensity}} \cdot \eta \cdot E_{\text{demand}} \tag{15}$$

- **Energy costs:** The cost of generating electricity from renewable and fossil fuels.

$$\text{Cost}_{\text{energy}} = \alpha \cdot P_{\text{renewable}} \cdot E_{\text{demand}} + (1-\alpha) \cdot P_{\text{fossil}} \cdot E_{\text{demand}} \tag{16}$$

- **Storage costs:** Assumes storage costs are proportional to the square of the proportion of renewable energy.

$$\text{Cost}_{\text{storage}} = \mu_{\text{storage}} \cdot \alpha^2 \cdot E_{\text{demand}} \tag{17}$$

The objective function can be written as a quadratic function of the proportion:

$$\text{Total Cost: } y = A \cdot \alpha^2 + B \cdot \alpha + C \tag{18}$$

in which $A = \mu_{\text{storage}} \cdot E_{\text{demand}}$, $B = (P_{\text{renewable}} - P_{\text{fossil}} - C_{\text{intensity}} \cdot \eta) \cdot E_{\text{demand}}$, and $C = (P_{\text{fossil}} + C_{\text{intensity}} \cdot \eta) \cdot E_{\text{demand}}$. Then the optimal proportion of renewable energy can be calculated as $\alpha^* = -B/2A$.

We assign a group of hypothetical values to the parameters in the model. The values are not based on real-world data, but the logic of the model is valid, with the aim of demonstrating the trade-off between the proportion of renewable energy and the stability cost of the electricity supply. The optimal proportion of renewable energy, derived from the given parameters, is $\alpha^* = 0.8333$.

Table 4 Parameter Settings

Notation	Description	Unit	Value
$P_{\text{renewable}}$	Renewable energy cost per unit	/MWh	100
P_{fossil}	Fossil fuel energy cost per unit	/MWh	60
$C_{\text{intensity}}$	Carbon intensity of electricity from fossil energ	kg CO_2/KWh	0.7
η	Carbon social monetization factor		200
μ_{storage}	Cost factor associated with electricity storage facilities	/MWh	60
E_{demand}	Total electricity demand	MWh	1000

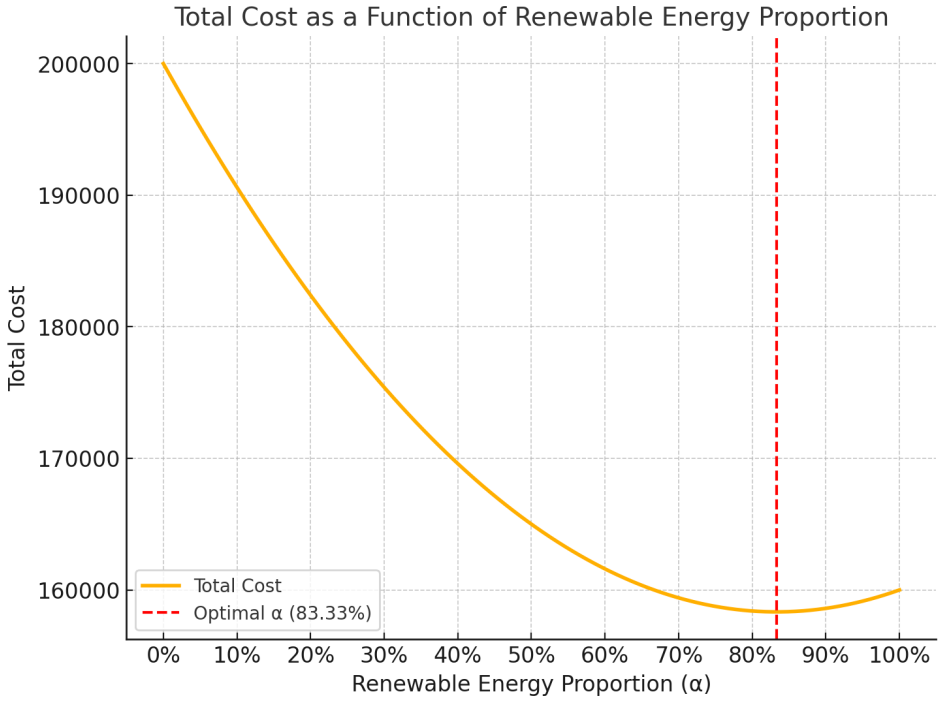

Figure 19 Optimal Proportion of Renewable Energy

The graph above shows the relationship between the total cost and the proportion of renewable energy (α) based on the given parameters. The optimal proportion of renewable energy (α^*) is marked at approximately 83.33%, where the total cost is minimized.

Through the analysis, the optimal proportion of renewable energy under current parameters is determined to be 83.33%. In this proportion, the total cost is minimized, including the socially monetized cost of carbon emissions, energy costs, and storage costs. Although high proportions of renewable energy generation incur higher storage costs, this configuration significantly reduces

carbon emissions while ensuring power supply stability. This demonstrates the potential and feasibility of adopting high proportions of renewable energy in the HPC industry.

6.3 The Trade-off between Impacts of HPC on Carbon Emission and Water Efficiencies

To fully grasp the environmental impact of HPC, it is essential to recognize the potential trade-offs between optimizing different environmental metrics. For instance, focusing solely on reducing carbon emissions may inadvertently increase water consumption, especially in regions reliant on water-intensive energy sources like hydropower. This underscores the need for energy efficiency strategies in data centers to consider both carbon and water footprints. Without such a balanced approach, optimizations aimed at reducing carbon emissions could exacerbate water scarcity issues, particularly in regions already facing significant water stress.

Li et al. (2023) identified a weak negative correlation between carbon intensity (gCO_2/kWh) and electricity water intensity (L/kWh) across various regions[10]. Building on this finding, we utilized electricity water intensity data from Li et al. (2023) and combined it with 2022 carbon intensity data from eGRID subregions in the U.S. and other countries. For the U.S., carbon intensity data were sourced from Electricity Maps, while country-level data were obtained from the Statistical Review of World Energy. First, we use a linear regression model to estimate the relationship between electricity water intensity and carbon intensity, resulting in the following equation:

$$\text{Water intensity}_t = 4.0676 - 0.0021\text{Carbon intensity}_t \tag{19}$$

The detailed results, as illustrated in Figure 20, indicate a negative relationship between carbon intensity and water consumption. Specifically, regions with higher carbon intensity, such as Wyoming and India, generally exhibit lower water consumption. Conversely, regions like Sweden, which rely extensively on hydropower, show low carbon emissions but higher water consumption. In summary, regions reliant on hydropower tend to have a lower carbon footprint but higher water usage, while those dependent on fossil fuels display the opposite trend.

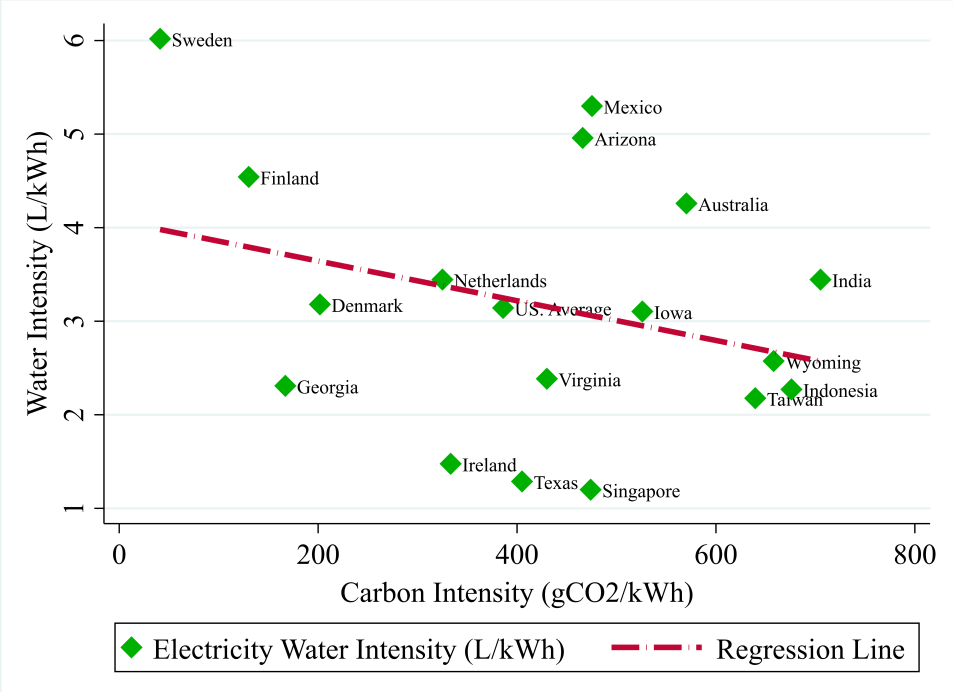

Figure 20 Carbon Intensity/Water Efficiency

The misalignment between carbon and water efficiencies suggests that optimizing solely for carbon reduction may inadvertently lead to increased water consumption. Therefore, HPC work-

loads should be dynamically scheduled based on regional energy profiles, shifting tasks to locations or times where both carbon and water footprints are minimized. This could include regions with a balanced energy mix or periods when renewable energy sources, such as solar or wind, are abundant. These findings highlight the need for comprehensive frameworks that integrate carbon and water metrics, allowing data centers and AI systems to mitigate their overall environmental impact without exacerbating water scarcity or carbon emissions.

7 Task 5: Measures to Reduce HPC Carbon Emissions

Rapid growth of the HPC industry has contributed significantly to global carbon emissions, which requires a concerted effort to mitigate its environmental impact. As one of the most energy-intensive sectors, HPC requires innovative strategies to achieve sustainability. Here, we propose six key measures to address this challenge, including four technical measures and two policy measures.

7.1 Technical Measures

1. **Increase the Proportion of Renewable Energy in Power Supply** Shifting the energy supply to renewable sources, especially for major HPC energy-consuming regions such as the United States, China, and Europe, is critical to decarbonizing the HPC sector. Data centers can adopt solar, wind, and other green energy sources to reduce reliance on fossil fuels. Energy storage solutions such as batteries or hydrogen can stabilize supply and address intermittency. Based on our previous analysis in Section 4.1, if the carbon intensity of electricity in the United States and China were reduced to the same level as that of Europe, global carbon emissions from HPC would decrease by 18.9%.
2. **Improve the Energy Efficiency (EE) of HPC Facilities** Enhancing the energy efficiency of HPC systems is a fundamental approach to reducing their carbon footprint. This can be achieved by adopting more energy-efficient hardware, such as low-power processors and advanced storage technologies.
3. **Enhance Power Usage Effectiveness (PUE)** Improving the PUE of HPC facilities involves reducing non-computational energy consumption. Advanced cooling technologies, such as liquid cooling or natural cooling methods, can significantly reduce energy usage in data center operations. Non-essential energy consumption, such as lighting and uninterruptible power supply inefficiencies, must also be minimized.
4. **Efficient Workload Allocation through Computational Networks** Through a computing power network, computational tasks can be optimally allocated based on the energy efficiency of HPC servers and the intensity of carbon emission from electricity generation in different regions and time periods. This approach prioritizes the most carbon-efficient HPCs for processing tasks, effectively reducing overall carbon emissions. In the subsequent analysis, we will demonstrate the significant potential of this method for carbon reduction through an optimization model.

7.2 Policy Measures

1. **Implement Carbon Taxation for HPC Facilities** A carbon tax provides financial incentives to influence economic behavior, encouraging the HPC industry to reduce its carbon emissions. The tax can be levied on the basis of the carbon emissions generated by different HPC systems, ensuring a fair and targeted approach. By internalizing the societal costs of emissions, as we have analyzed in the previous section, it mitigates negative externalities and drives the adoption of cleaner technologies and investments in renewable energy, fostering greater sustainability in the HPC sector.

2. **Advocate Large Tech Companies Implementing Green Energy and Zero-Carbon Strategies** Large tech companies, as primary users of HPC, should actively implement green energy and net-zero emission strategies. Given that the top five U.S. tech companies account for approximately 20% of the HPC market, their adoption of green power initiatives could significantly reduce the sector's energy consumption and carbon emissions. Furthermore, these efforts would set a strong example for other market participants, accelerating the industry's transition toward sustainability.

7.3 Quantifying the Effect of Optimal Computing Workload Allocation

To demonstrate the potential of the method for carbon reduction we mentioned above, we integrated data from the TOP500 database and Electricity Maps, linking HPC systems to regional carbon intensities. This enabled an optimization model to align computational tasks with carbon efficiency. A US case study highlights actionable strategies to minimize emissions and improve HPC sustainability.

7.3.1 Model

This section aims to minimize the total carbon emissions of HPC systems by strategically distributing computational loads across systems and time periods. Using a real-world dataset derived from HPC systems in the United States, we construct and validate a dynamic optimization model. For a given computational task D, the objective is to minimize total carbon emissions by adjusting the allocation of computational loads across HPC systems in different individuals and time periods. The objective function for total carbon emissions, $TotalCarbon$, is given by:

$$TotalCarbon = \sum_{t=1}^{T}\sum_{i=1}^{N} Carbon_{it} = \sum_{t=1}^{T}\sum_{i=1}^{N} (w_{it}D \times EC_i \times C_{it}) \tag{20}$$

Variable	Description	Unit
T	Number of time periods	
N	Number of HPC systems	
w_{it}	Computational load weight of HPC i at time t	
D	Total computational workload	TFlop
EC_i	Energy consumption per unit of computation for HPC i	kWh/TFlop
C_{it}	Carbon emissions per unit of electricity generated for HPC i	kt/TFlop
$\bar{D}_i$	The maximum computational capacity of HPC i	TFlop

Table 5 Description of Variables

In this formulation, w_{it} serves as the decision variable, determining the share of the total computational load allocated to each HPC i at each time period t, while the other parameters account for the energy intensity and carbon footprint associated with energy generation for each HPC. The goal is to minimize $TotalCarbon$ by adjusting w_{it} values across all HPC systems and time periods. The constraints are as follows: The weights w_{it} are non-negative and their total sum equals 1. Specifically, for each i and t, the condition $0 \leq w_{it} \leq 1$ must hold, and the total sum across all HPC systems and time periods is constrained by $\sum_{t=1}^{T}\sum_{i=1}^{N} w_{it} = 1$. Additionally, for each HPC i and each time period t, the computational allocation should not exceed the HPC system's maximum computational capacity $\bar{D}_i$. This condition is expressed as $w_{it} \cdot D \leq \bar{D}_i$, ensuring that the allocation does not surpass the HPC's capacity.

$$\min_{w_{it}} \quad TotalCarbon \quad \text{s.t.} \quad 0 \leq w_{it} \leq 1 \quad \forall\, i,\, t; \quad \sum_{t=1}^{T}\sum_{n=1}^{N} w_{it} = 1; \quad w_{it} \cdot D \leq \bar{D}_i. \tag{21}$$

7.3.2 Dataset and samples

For our analysis, we focused on HPC systems located in the United States from the TOP500 list [5] and evaluated their carbon emissions optimization using carbon intensity data from Electricity Maps[6]. This sample allowed us to examine the optimization problem across different time intervals, highlighting temporal variations in carbon emissions.

7.3.3 Algorithm to find the optimal solution

In our analysis, we first used a linear programming solver, specifically the *scipy.optimize.linprog* function in Python, to determine the optimal solution for the problem. This solver efficiently minimized the objective function under the given constraints, providing a reliable benchmark for evaluating alternative approaches. Interestingly, we found that the algorithm we proposed achieves an equivalent outcome to the solver's optimal solution. To better understand the essence of the problem, we present this algorithm, demonstrating its ability to uncover the underlying structure and equivalence in the optimization process.

We propose an a greedy-based algorithm to minimize carbon emissions by strategically distributing workloads across HPC nodes. It begins by sorting nodes based on their carbon emissions per unit of computation ($EC_i \times C_{it}$), prioritizing those with lower carbon emissions per unit of computation. Workloads are then allocated iteratively, calculating cumulative capacity until the total workload is met, with any remaining load distributed proportionally to the final node. The allocation weights are optimized to ensure compliance with constraints, including non-negativity, capacity limits, and full workload distribution.

7.3.4 Results of optimization

To evaluate the performance of the proposed algorithm, we analyzed its results using visualizations of load distribution and carbon emission reduction, see figure below.

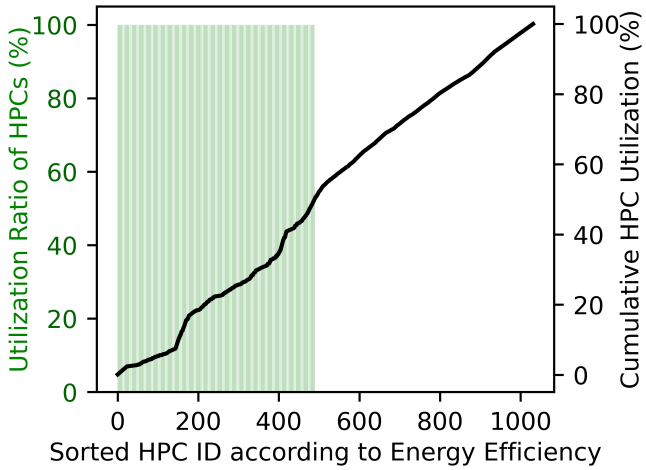

Figure 21
Optimized workloads (Total Utilization = 50%)

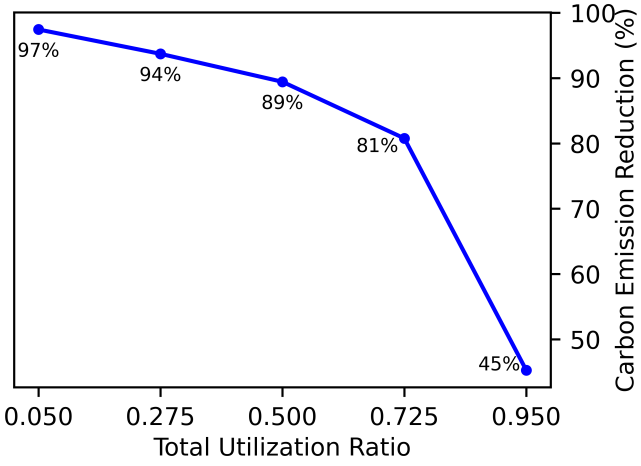

Figure 22
Total Utilization & Carbon Emission Reduction

The figure on the left demonstrates the optimized distribution of computational loads under a utilization ratio of 50%. The bar chart indicates the percentage of workload allocated to each HPC node, while the line chart shows the cumulative percentage of HPC capacity utilized. The results reveal that the algorithm allocates higher loads to nodes with lower carbon emissions per

[5] https://www.top500.org
[6] https://app.electricitymaps.com

unit of computation ($EC_i \times C_{it}$), ensuring efficient use of available capacity while minimizing environmental impact.

The algorithm supports iterative adjustment of the total utilization ratio (R) to explore trade-offs between system utilization and emission reduction. By varying R, it generates a spectrum of feasible solutions, allowing decision-makers to balance operational efficiency with sustainability. The figure on the right hand side highlights the relationship between total utilization ratios and carbon emission reductions. The curve indicates a significant decrease in emissions as the utilization ratio increases, demonstrating the effectiveness of the carbon-aware strategy. Notably, the reduction rate diminishes at higher utilization levels, reflecting the diminishing returns of load distribution optimization as computational capacities are saturated.

The results demonstrate that the proposed algorithm effectively reduces carbon emissions by prioritizing low-intensity nodes while adhering to system constraints. Higher utilization ratios yield substantial emission reductions, though with diminishing returns as capacity limits are approached.

7.3.5 Explanation for carbon emission reduction

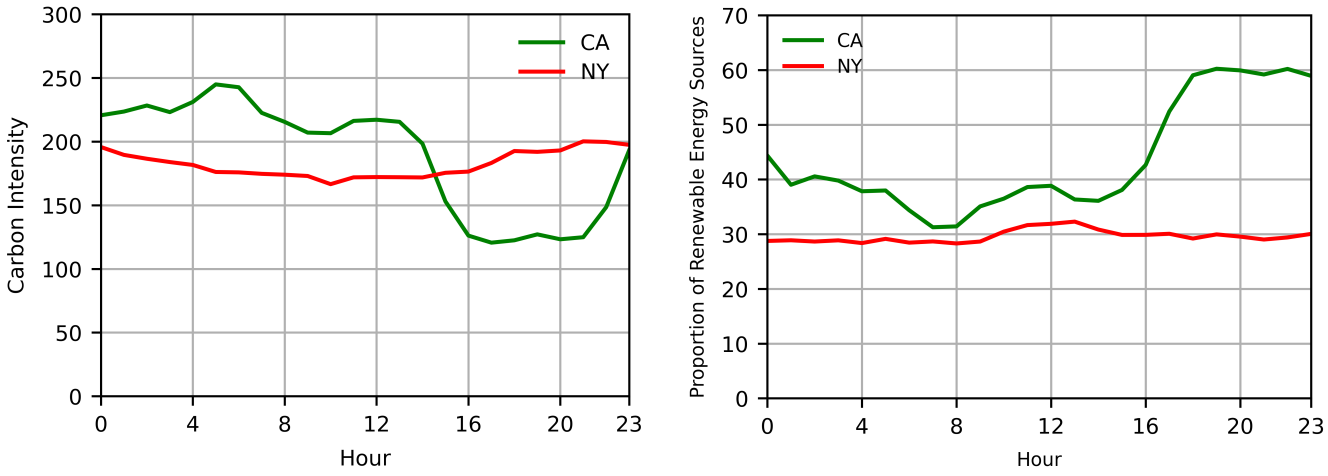

Figure 23 Carbon Intensity and Proportion of Renewable Energy Sources for CA and NY

The results presented in the two figures emphasize the temporal variability of carbon emission intensity and the percentage of renewable energy used in California (CA) and New York (NY), highlighting the necessity of dynamically allocating computational workloads across both time and space to minimize environmental impact.

Carbon emissions in CA peak midday and decline sharply in the evening, reflecting fluctuations in its energy mix and renewable energy availability, while NY maintains a more stable emission profile due to its consistent energy sources. Using CA and NY as examples, this analysis underscores the value of carbon-aware workload optimization. Temporal shifting can leverage periods of high renewable energy availability, such as midday in CA, while spatial redistribution can shift tasks from high-carbon periods or regions, such as evening in CA, to more stable profiles like NY, achieving significant emission reductions.

8 Strengths and Weaknesses

Strengths

- **ARIMA Forecasting for HPC energy consumption.** We use ARIMA model to predict global HPC energy consumption and carbon intensity from 2024 to 2030.

- **Decomposition model to understand Energy Consumption**. we present a decomposition model that breaks down HPC energy consumption into key drivers: Compute Capacity, Utilization Ratio, Power Usage Effectiveness (PUE), and Energy Efficiency.
- **Scenario-Based Analysis**. We incorporate multiple growth scenarios (low, base, high) for HPC energy consumption and emissions, considering varying rates of renewable energy adoption and efficiency improvements.
- **Optimization Framework**. We develop an optimization model to dynamically allocate computational workloads across regions and time zones based on carbon intensity and renewable energy availability.
- **Statistical and Regression Analyses**. We applied linear regression to analyze the relationship between renewable energy proportion and carbon intensity, using the results to assess the impact of increasing renewable energy adoption.

Weaknesses

- **Lack of High-Quality Public Data**. The analysis is limited by the scarcity of publicly available high-quality data, especially for time series data on key variables such as global HPC energy consumption.
- **Simplified Assumptions**. The study uses simplified assumptions, such as constant ratios and linear relationships between variables, which may oversimplify the complexities of real-world dynamics and introduce potential inaccuracies in the models.

9 Conclusion

This paper presents a comprehensive analysis of the energy consumption and carbon emissions of the world. Our findings indicate that in the baseline scenario, HPC energy consumption is projected to increase to 1,219 TWh by 2030, driven primarily by the increasing demand for compute capacity.

Using a decomposition model, we identified four key factors: Computing Capacity, Utilization Ratio, PUE, and Energy Efficiency, which collectively determine the growth of HPC energy consumption. ARIMA modeling also allowed us to forecast the global intensity of HPC carbon through 2033, providing robust information on future trends. Furthermore, we developed an optimization framework to dynamically allocate workloads based on renewable energy availability and carbon intensity, demonstrating significant potential to reduce emissions.

The strength of this paper lies in its rigorous mathematical modeling and scenario-based analysis, offering actionable strategies for mitigating emissions in the HPC sector. By combining detailed forecasts, optimization models, and a holistic assessment of driving factors and environmental impacts, this study provides valuable guidance for policymakers and industry stakeholders to achieve sustainability goals while supporting the continued growth of HPC sector.

References

[1] M. Azarifar, M. Arik, and J.-Y. Chang, *Liquid cooling of data centers: A necessity facing challenges*, Applied Thermal Engineering, 247 (2024), pp. 123112–123112.

[2] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA, 1970.

[3] D. A. Dickey and W. A. Fuller, *Distribution of the estimators for autoregressive time series with a unit root*, Journal of the American Statistical Association, 74 (1979), pp. 427–431.

[4] Environmental Protection Agency (EPA), *Social cost of greenhouse gases: Report*, tech. rep., Environmental Protection Agency, 2023.

[5] E. M. Fischer and R. Knutti, *Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes*, Nature Climate Change, 5 (2015), pp. 560–564.

[6] International Data Corporation (IDC), *Worldwide high-performance computing market forecast, 2023-2027*, tech. rep., International Data Corporation, Framingham, USA, 2023.

[7] International Energy Agency (IEA), *Electricity 2024: Analysis and forecast to 2026*, tech. rep., International Energy Agency, 2024.

[8] IPCC, *Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva, Switzerland, 2023.

[9] A. Jo, *The promise and peril of generative ai*, Nature, 614 (2023), pp. 214–216.

[10] P. Li, J. Yang, M. A. Islam, and S. Ren, *Making ai less "thirsty": Uncovering and addressing the secret water footprint of ai models*, (2023).

[11] G. M. Ljung and G. E. P. Box, *On a measure of lack of fit in time series models*, Biometrika, 65 (1978), pp. 297–303.

[12] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, *Recalibrating global data center energy-use estimates*, Science, 367 (2020), pp. 984–986.

[13] McKinsey & Company, *The state of ai in 2023: Generative ai's breakout year*, tech. rep., McKinsey Global Institute, New York, USA, 2023.

[14] NVIDIA, *Hpc and ai: A new era of accelerated computing*, tech. rep., NVIDIA, Santa Clara, USA, 2023.

[15] D. Patterson et al., *Carbon emissions and large neural network training*, arXiv, (2021).

[16] P. Samuelson, *Generative ai meets copyright*, Science, 381 (2023), pp. 158–161.

[17] Stanford University, *Ai index report 2023*, tech. rep., Stanford Human-Centered AI Institute, Stanford University, Stanford, USA, 2023.

[18] S. S. V. Varsha, B. S. Thomas, C. Kundu, A. K. Vuppaladadiyam, H. Duan, and S. Bhattacharya, *Can e-waste recycling provide a solution to the scarcity of rare earth metals? an overview of e-waste recycling methods*, Science of the Total Environment, 924 (2024), pp. 171453–171453.

[19] R. Widmer, H. Oswald-Krapf, D. Sinha-Khetriwal, M. Schnellmann, and H. Böni, *Global perspectives on e-waste*, Environmental Impact Assessment Review, 25 (2005), pp. 436–458.

[20] World Bank and Institute for Health Metrics and Evaluation, *The cost of air pollution: Strengthening the economic case for action*, tech. rep., World Bank, Washington, DC, 2016.