Mitigating the Environmental Impact of High-Performance Computing Data Centers

Zhaojie Lu, Rhibull Tang, Zijia Wan, Zetao Ye

Shenzhen Middle School, Shenzhen, Guangdong, 518000, China

Abstract. While High Performance Computing (HPC) is critical for scientific breakthroughs and AI development, its growing energy use (often dependent on fossil fuels) requires a detailed environmental impact analysis. This study addresses the ever-increasing ecological challenges posed by high-performance computing (HPC) data centers by developing a model to quantify and predict their energy consumption and carbon emissions. Our model estimates the annual energy consumption of HPC data centers globally using data on the size and capacity of small and massive data centers. It distinguishes between theoretical maximum and actual energy consumption based on average utilization. The results show an upward trend in carbon emissions associated with HPC, which are projected to reach 5.81 x 10^{17} grams of carbon dioxide by 2030. As a result, we propose a series of recommendations to reduce the environmental impact of HPC at both the technical and policy levels in favor of promoting a healthy and sustainable environment.

Keywords: High-Powered Computing; Data Center; Electricity Consumption; Generalized Linear Model; Polynomial Regression.

1. Introduction

High-performance computing (HPC) is a cluster of computers utilizing high-performance processors to solve large-scale scientific problems and process massive amounts of data. Due to its wide range of applications and efficient computational speed, it plays an increasingly important role in finance, government informatization, education, enterprise, and online games. However, at the same time, its high energy consumption characteristics seriously impact the environment. In recent years, the environmental impact of high-performance computing (HPC) has gradually received attention from researchers. Since the high-power input of HPC is usually provided by non-renewable energy sources, most commonly fossil fuels, carbon emissions are generated during the combustion process, which leads to climate warming. In addition, HPC systems consume large quantities of water and chemical coolants, contributing to wastewater pollution. In addition to water pollution, components often need to be promptly replaced due to component aging, which tends to generate large amounts of e-waste, further exacerbated by the manufacturing process. In addition, the real estate construction of data centers and the associated dust and air pollution from power plants contribute to the environmental legacy of HPC. The impact of data center ground usage on local populations and air pollution from power plants on human health are also among the common HPC environmental impacts. Modeling these ecological threats is therefore critical to creating sustainable 'green' HPC solutions in the future.

If sustainable and inclusive artificial intelligence and other computationally demanding new technologies must be developed and manufactured, the environmental impact of high-performance computing must be considered. This study combines energy consumption predictions with carbon emission factors for various energy sources, including coal, natural gas, renewable energy, and nuclear energy. Models are used to quantify and predict the energy consumption and carbon emissions of High-Performance Computing (HPC) data centers. An ARIMA model predicts data center expansion up to 2030 and assesses the impact on future energy use. The results of the projections can be used to make valuable recommendations for energy use and environmental sustainability in the current world, contributing to the achievement of the 2030 Sustainable Development Goals (SDGs).

2. Problem Restatement

As HPC hardware in data centers accelerates technological progress, it imposes a heavy environmental price. This paper aims to build a model to determine and predict the ecological consequences of expanding HPC-based data centers. The model will:

(1) Estimate the environmental impacts of data center power consumption in terms of annual electricity use at full and average operational loads while considering carbon output.
(2) Project the environmental impacts to 2030, including predicted increases in HPC hardware adoption.
(3) Discuss the implications of the shift toward alternative energy sources and the environmental consequences of cooling systems (both sources of emissions and those of waste), respectively.
(4) Illustrate how the model can capture the positive effect of current and future mitigation measures to ease the environmental load of data center infrastructure.

The work will produce an advisory report to be submitted to the UN Advisory Board, which will suggest that the model's results be included in their report on the environmental impact of HPC leveraged for AI computation.

3. Methods and Results

3.1 Annual Energy Consumption Estimate

We must consider two situations in this question: (1) full capacity and (2) average utilization rate. Therefore, to better estimate the total annual energy consumption, we need to define the number of data centers worldwide (N) and estimate the power of high-powered computing (HPC) under two conditions: (1) full capacity and (2) averaged utilization rate.

To obtain the average power of HPC data centers at full capacity, we identified the online database and found the Data sources: Brightlio, Data Centers World Map, etc.

Table 1. Summary of data centers worldwide (type, number, and average full capacity)

Type	Number (N_i)	Power (P_i)	Averaged full capacity ($\dot{P}_i$)
Small data centers	992	20-100 MW	60 MW
Large or hyper-scale data centers	9986	1-5 MW	3 MW

The annual energy consumption of HPC capacity is defined as:

$$EC_{total} = \sum_{i=1}^{n} N_i * \dot{P}_i * time \quad (1)$$

According to the online database, there are 10978 data centers in 158 countries worldwide. Under the full capacity, the annual energy consumption is calculated as:

$$EC_{total} = (992*60+9986*3)*24*365 MWh = 7.84*10^8 MWh \quad (2)$$

Then, we found the average utilization rate between 12% and 18%. Therefore, considering the average utilization rate, the actual annual energy consumption is equal to the following:

$$EC_{average} = R_{utilize} * EC_{total} = 12\%*7.84*10^8 MWh = 9.408*10^7 MWh \quad (3)$$

3.2 Total Carbon Emission Estimate

We must consider the energy structure to estimate the total carbon emission. For electricity consumption, we mainly use:

The average carbon emission considering the energy mixture is calculated as:

$$C_{mean} = \sum_{i=1}^{n} p_i C_i \quad (4)$$

Here, C_{mean} is defined as the averaged carbon emission per unit energy generation, and $C_i (i = 1,2,\ldots,n)$ is defined as the carbon emission per unit energy generation for n different sources. Therefore, we can calculate the C_{mean}:

$$C_{mean} = (34\% * 820 + 2\% * 490 + 32\% * 740 + 2\% * 48 + 3\% * 38 + 5\% * 24 + 20\% * 12 + 2\% * 12)g/kWh = 456.34g/kWh$$

Table 2. Carbon emission per unit of energy consumed by different sources

Energy Source	carbon emission (g/per kWh)
Coal	820
Biomass	740
Natural gas	490
Solar energy	48
Geothermal	38
Hydropower	24
Nuclear	12
Wind	12

Table 3. Percent of energy source for electricity generation

Energy Source	Percentage of energy source for electricity generation
Coal	34%
Biomass	2%
Natural gas	32%
Solar energy	2%
Geothermal	3%
Hydropower	5%
Nuclear	20%
Wind	2%

Therefore, the actual total carbon emissions due to the electricity consumption of HPC data centers are equal to:

$$\mathrm{C_{total}} = \mathrm{C_{mean}} * \mathrm{EC_{average}} = 456.34\mathrm{g/kWh} * 9.408 * 10^{10}\mathrm{kWh} = 4.293 * 10^{13}\mathrm{g} \quad (5)$$

The total annual carbon emissions are estimated as **$4.293*10^{13}$**g.

3.3 Model Application

Here, we need to study three significant factors that may affect the estimation of total carbon emissions: (a) the growth of HPC, (b) the increasing demand for energy in other sectors, and (c) the potential for energy sources and mixes.

3.3.1 The Growth of HPC

To estimate HPC's growth, we consider the annual increase of HPC and the annual power increase.

Table 4. Number of small and hyperscale data centers worldwide from 2015 to 2023.

year	small	hyper-scale	total
2015	10325	259	10584
2016	9657	338	9995
2017	11676	386	12062
2018	11718	448	12166
2019	12161	504	12665
2020	12177	597	12774
2021	12092	700	12792
2022	12055	846	12901
2023	12035	992	13027

Therefore, we can estimate the number of small and hyper-scale data centers.

$$N_{small} = 26908396 * year - 66.57 * year^2 - 271886108.50$$

$$N_{hyperscale} = -29555.37 * year + 7.34 * year^2 + 29748619.47$$

Therefore, we can calculate the future averaged carbon emission using the formula proposed in (2).

$$C_{future_{mean}} = \sum_{i=1}^{n} p_i C_i \quad (6)$$

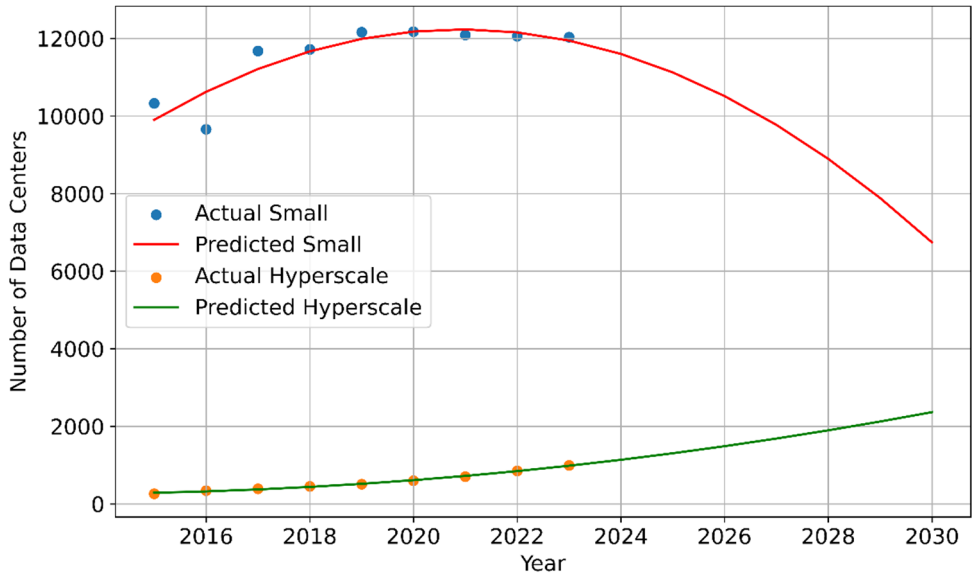


Fig 1. The number of data centers by 2030.

Therefore, using a similar method, we can estimate the total energy consumption of HPC data centers:

$$EC_{total} = \sum_{i=1}^{n} N_i * \dot{P}_i * time \quad (7)$$

Then, the total energy consumption of the HPC data center is equal to:

$$EC_{total} = R_{utilize} * \sum_{i=1}^{n} N_i * \dot{P}_i * time$$
$$= \{(-29555.37 * year + 7.34 * year^2 + 29748619.47) * 60 + (26908396 * year - 66.57 * year^2 - 271886108.50) * 3\} * 24 * 365MWh$$
$$= 8760 * \{80712852.63 * year - 132.37 * year^2 + 969258842.7\}MWh$$

3.3.2 The Increasing Demand for Energy in Other Sectors

There are primarily four sectors: residential, commercial, industrial, and transportation. Here, we investigated the percentage of each industry using historical data from 2013 to 2023. The percentages of residential, commercial, and industrial sectors are defined as:

Table 5. Annual energy consumption by different sectors.

Year	Residential	Commercial	Industrial	Transportation	Total
2013	1,394,812	1,337,079	985,352	7,625	3,724,868
2014	1,407,208	1,352,158	997,576	7,758	3,764,700
2015	1,404,096	1,360,752	986,508	7,637	3,758,992
2016	1,411,058	1,367,191	976,715	7,497	3,762,462
2017	1,378,648	1,352,888	984,298	7,523	3,723,356
2018	1,469,093	1,381,755	1,000,673	7,665	3,859,185
2019	1,440,289	1,360,877	1,002,353	7,632	3,811,150
2020	1,464,605	1,287,440	959,082	6,548	3,717,674
2021	1,470,487	1,328,439	1,000,613	6,334	3,805,874
2022	1,509,233	1,390,873	1,020,464	6,599	3,927,169
2023	1,450,025	1,408,109	1,009,256	6,864	3,874,253

A time series model was developed for each sector to predict the demand to study the increasing demand for electricity in other sectors (residential, commercial, industrial, and transportation). A non-seasonal autoregressive integrated moving average (ARIMA) model was developed from the

combination of differencing, autoregression, and moving average. Then, the whole model is written as:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (8)$$

p denotes the differenced time series, which may have undergone multiple differencing operations. d denotes the degree of first differencing involved. q denotes the order of the moving average part.

Table 6. ARIMA time series model to predict the electricity demand for other sectors.

Sectors	Model
Residential Demand	ARIMA (1,0,2)
Commercial Demand	ARIMA (2,0,2)
Industrial Demand	ARIMA (1,0,1)
Transportation Demand	ARIMA (2,0,0)

3.3.3 The Potential for Different Energy Sources and Mixes

Therefore, we can develop the regression model of each energy source below:

$$S_{coal} = -91512.54 * year + 185786177.47$$
$$S_{Biomass} = -424.64 * * year + 870660.18$$
$$S_{Natural\ gas} = 65157.17 * * year - 130030574.84$$
$$S_{Solar} = 15299.17 * * year - 30801866.75$$
$$S_{Geothermal} = -123.55 * * year + 261700.07$$
$$S_{Hydropower} = -964.97 * * year + 2216606.33$$
$$S_{Nuclear} = -2300.01 * * year + 5434730.25$$
$$S_{Wind} = 26730.14 * * year - 53579775.36$$

According to the number of different energy sources, we were able to calculate the percentage ($\dot{p}_i$) of each energy source using the formula:

$$\dot{p}_i = \frac{S_i}{\sum_{i=1}^{n} S_i} \quad (9)$$

For example, the percentage of coal energy consumption is equal to:

$$p_{coal} = \frac{S_{coal}}{\sum_{i=1}^{n} S_i} = \frac{-91512.54 * year + 185786177.47}{11860.77 * year - 19842342.65}$$

Similarly, we can calculate the percentage of other energy sources.

Then, we can calculate the weighted carbon emission per unit of energy consumption using the same formula above:

$$C_{mean} = \sum_{i=1}^{n} p_i C_i = C_{mean}$$
$$= \left(\frac{-91512.54 * year + 185786177.47}{11860.77 * year - 19842342.65} * 820 \right.$$
$$+ \frac{-424.64 * * year + 870660.18}{11860.77 * year - 19842342.65} * 490$$
$$+ \frac{65157.17 * * year - 130030574.84}{11860.77 * year - 19842342.65} * 740$$
$$+ \frac{15299.17 * * year - 30801866.75}{11860.77 * year - 19842342.65} * 48 + \frac{-123.55 * * year + 261700.07}{11860.77 * year - 19842342.65}$$
$$* 38 + \frac{-964.97 * * year + 2216606.33}{11860.77 * year - 19842342.65} * 24$$
$$+ \frac{-2300.01 * * year + 5434730.25}{11860.77 * year - 19842342.65} * 12 + \frac{26730.14 * * year - 53579775.36}{11860.77 * year - 19842342.65}$$
$$\left. * 12 \right) g/kWh = \frac{-26032383.1 * year + 54555576641}{11860.77 * year - 19842342.65} g/kWh$$

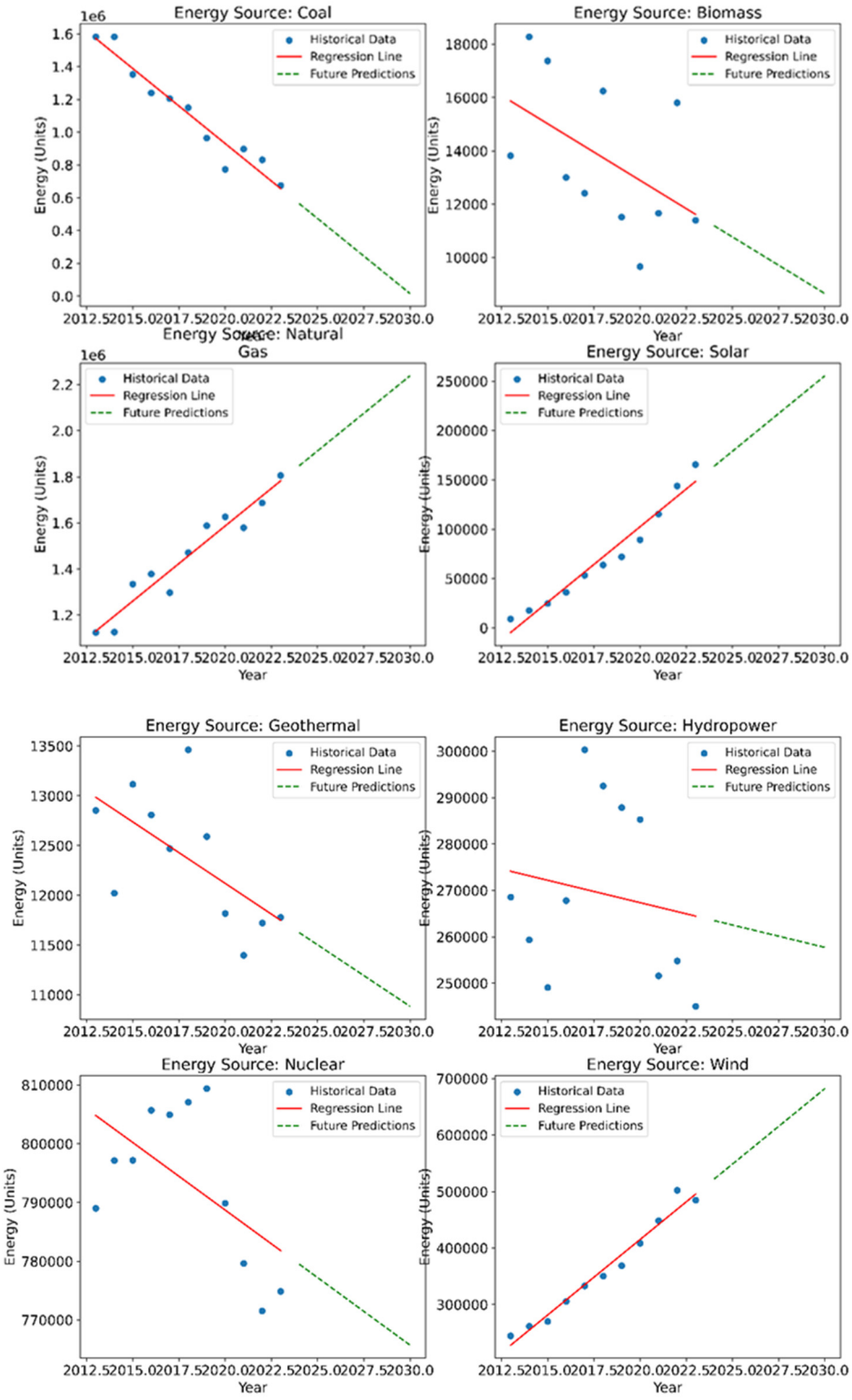


Fig 2. Regression model of different energy sources

Therefore, we can calculate the updated total carbon emission due to the growth of data centers, development of renewable energy, and change in energy mixture:

$$C_{total} = C_{mean} * EC_{average}$$
$$= 8760 * \{80712852.63 * year - 132.37 * year^2 + 969258842.7\} MWh$$
$$* \frac{-26032383.1 * year + 54555576641}{11860.77 * year - 19842342.65} g/kWh$$

Based on the above formula, we can estimate the total carbon emissions from 2024 to 2030, as shown in Table 7.

Table 7. Total carbon emissions from 2024 to 2030.

Year	Total carbon emission
2024	$6.43 * 10^{17}$
2025	$6.33 * 10^{17}$
2026	$6.22 * 10^{17}$
2027	$6.12 * 10^{17}$
2028	$6.01 * 10^{17}$
2029	$5.91 * 10^{17}$
2030	$5.81 * 10^{17}$

Therefore, the realistic bound for the total carbon emission by 2030 is $5.81*10^{17}$g CO_2.

3.4 Model Expansion

A polynomial regression model was developed to study the impact of increasing the portion of renewable energy (x_i) and a corresponding reduction in carbon emissions (y_i):

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i (i = 1,2,\ldots,n) \tag{10}$$

That can be expressed in the matrix form below:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ 1 & x_3 & x_3^2 & \ldots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{11}$$

Or it could be written as:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \tag{12}$$

The estimated polynomial regression coefficients are:

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \vec{y} \tag{13}$$

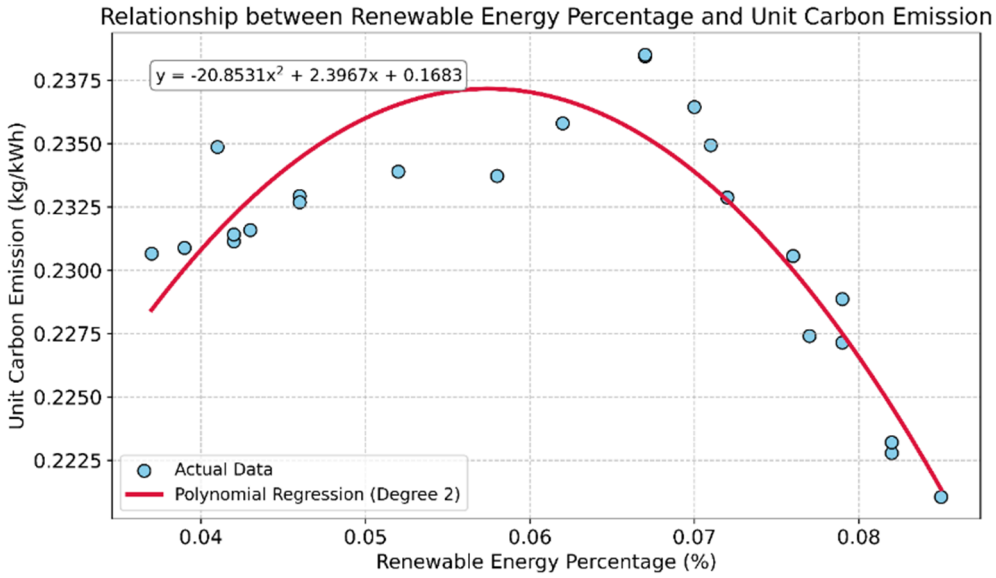


Fig 3. Relationship between renewable energy percentage and unit carbon emission

$$CE_{unit} = -20.85 * P_{renew}^2 + 2.40 * P_{renew} + 0.17 \tag{14}$$

Therefore, we can see the declining trend of unit carbon emission when increasing the renewable energy percentage.

It is not feasible to entirely switch to 100% renewable energy; there are a couple of potential challenges below:

(a) Reliability

(1) Solar and wind power are inherently intermittent, meaning their output fluctuates with weather conditions. However, this variation is a significant issue regarding grid stability since supply is always required to equalize demand.

(2) Large-scale energy storage solutions (e.g., batteries, pumped hydro, compressed air) are needed to address intermittency and ensure a reliable power supply when renewable sources are unavailable. These technologies are still developing and can be expensive.

(3) Sophisticated grid management, including grid and smart grids, is needed to achieve real-time supply and demand balance, incorporate heterogeneous renewables, and account for variability.

(b) Infrastructure and Land Use

(1) Substantial investments are needed to modernize and expand transmission infrastructure to bring remote renewable energy generation (e.g., offshore wind farms and solar farms in desert regions) to urban areas.

(2) Renewable energy installations consume large land parcels, affecting ecosystems, agriculture, and surrounding communities. Careful site planning and location are critical to keep environmental impacts minimal.

(3) Renewable energy technologies involve significant raw materials (i.e., rare earth minerals and silicon). Sustainably and ethically sourcing these materials is an issue that is becoming increasingly important.

(c) Economic and Financial Considerations

(1) Upfront costs: Moving to 100% renewables requires significant upfront capital expenditures to build the new generation capacity, energy storage, and grid capability.

(2) Shifting away from fossil fuels can lead to job losses in traditional energy industries. New jobs and retraining are needed in renewable energy as part of a just transition.

(3) Affordable energy prices during the transition are essential for equal access to clean energy.

(d) Social and Political Factors

(1) Public acceptance: Obtaining public support for large-scale renewable energy installations is difficult, mainly because of concerns about aesthetic implications, impacts on land use, and environmental consequences.

(2) Stable and supportive government regulations, policies, and incentives are essential to stimulating investment in renewable energy and smoothing the transition.

(3) The shift towards renewables can alter geopolitical power relations, especially for nations whose economy is based on fossil fuel exports. International cooperation is needed to manage this transition effectively.

(e) Technological Development

(1) Technological developments in energy storage, smart grids, and grid management facilities are needed to overcome intermittency and enhance grid reliability.

(2) Further development of new - generation renewable energy resources (e.g., high - temperature geothermal energy, wave, tidal energy) may extend the renewable energy portfolio and decrease the dependence on current - generation energy resources.

4. Model Sharing and Response

4.1 Actionable Recommendations

4.1.1 Technical Solutions

(a) Optimize Hardware and Software

(1) Prioritize the procurement of hardware with high performance-per-watt metrics, such as ARM-based processors or specialized accelerators designed for energy efficiency.
(2) Design and implement energy-efficient algorithms and software. Encourage code profiling and optimization practices among HPC users.

(b) Improve Data Center Efficiency

(1) Develop advanced cooling systems (e.g., liquid, free-air, evaporative cooling) to decrease the energy needed for temperature control.
(2) Investigate the possibility of collecting and using waste heat produced by HPC systems for building heating or other applications.
(3) Where available, HPC data centers should be powered with renewable energy (solar, wind, or hydropower). On-site generation or Power Purchase Agreements (PPAs) can be valuable tools.

(c) Promote Green HPC Practices

(1) Optimize job scheduling algorithms to maximize resource utilization and minimize idle time, reducing energy consumption.
(2) Utilize virtualization and cloud-based high-performance computing (HPC) resources to consolidate workloads and optimize resource utilization, eliminating the need for dedicated hardware.
(3) Develop energy-aware workload management systems focusing on energy efficiency during job scheduling and running.

4.1.2 Policy-Oriented Solutions

(a) Incentivize Green HPC Procurement

(1) Authorize or encourage energy-efficient HPC hardware and software procurement at government-funded research centers and agencies.
(2) Provide tax relief or financial incentives to organizations seeking green HPC technologies and practices.

(b) Promote Research and Development

(1) Further, provide government funding to develop energy-efficient HPC hardware, software, and data center technologies.
(2) Foster cooperation to speed up the development and implementation of green HPC approaches.

(c) Establish Energy Efficiency Standards

(1) Develop standardized metrics and rating systems for evaluating the energy efficiency of HPC systems and data centers. Energy Star for HPC could be explored.
(2) Implement mandatory energy efficiency goals for HPC sites, encouraging operators to adopt best practices.

(d) Raise Awareness and Education

(1) Provide education courses for HPC users and operators on energy-optimal techniques and hardware.
(2) Present and disseminate green HPC research and best practices through workshops, conferences, and web resources.

(e) International Collaboration:

Embrace international teamwork for green HPC research, development, and policy to speed progress and share expertise globally.

4.2 Model Response

When one of the recommendations is acted upon, i.e., when the portion of renewable energy increases by 5% from the current 8.5% of all electricity generation worldwide, the carbon emission per unit of electricity generation would reduce from 0.22 pound/kWh to 0.135 pound/kWh, which is 38.6% reduction in carbon emissions. Considering the vast amount of electricity consumption by the HPC data center, there is a significant need to promote the transition to renewable energy to protect the environment.

5. Conclusion

By combining Autoregressive Integrated Moving Average (ARIMA), capacity forecasting, and various machine learning models (linear regression, polynomial regression, support vector regression, random forests, gradient boosting), this paper investigates capacity forecasting and utilization forecasting using daily observations from 2021 to 2023 for five major countries. The study results show that the energy consumption of HPC infrastructure will increase significantly based on growing capacity and utilization. If left unchecked, this growth path will play a key role in global energy and greenhouse gas trade, jeopardizing the achievement of the 2030 Sustainable Development Goals (SDGs), particularly those related to climate action and responsible lifestyles and production. The increasing reliance on HPC for artificial intelligence development, scientific research, and other computationally intensive tasks necessitates a proactive and comprehensive strategy to mitigate its environmental impact. The current trajectory of HPC energy consumption is unsustainable and requires immediate attention. In response, this paper proposes measures such as developing energy efficiency standards, integrating renewable energy sources, developing sustainable cooling technologies, and international cooperation. By incorporating these recommendations into the 2030 Development Goals, the United Nations can drive HPC towards a more sustainable model.

References

[1] 2024 Global Data Center Market Comparison. (n.d.). Page 30. https://cushwake.cld.bz/2024-Global-Data-Center-Market-Comparison/30/.

[2] AI is poised to drive a 160% increase in data center power demand. (2024, May 14). Goldman Sachs. https://www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand.

[3] Author, N. (2023, June 2). The huge carbon footprint of large-scale computing – Physics World. Physics World. https://physicsworld.com/a/the-huge-carbon-footprint-of-large-scale-computing/.

[4] Bautista, D. (2022, March 24). Water usage for data center cooling and electricity. AKCP Remote Sensor Monitoring | Data Center Monitoring. https://www.akcp.com/blog/water-usage-for-data-center-cooling-and-electricity-generation/.

[5] National Knowledge Infrastructure. (n.d.). https://www.cnki.net/.

[6] Energy.gov. (n.d.). The Department of Energy's Energy.gov. https://www.energy.gov/.

[7] Global e-Waste Monitor 2024: Electronic Waste Rising Five Times Faster than Documented E-waste Recycling. (n.d.). UNITAR. https://unitar.org/about/news-stories/press/global-e-waste-monitor-2024-electronic-waste-rising-five-times-faster-documented-e-waste-recycling.

[8] High-Performance Computing data center power usage effectiveness. (n.d.). Computational Science | NREL. https://www.nrel.gov/computational-science/measuring-efficiency-pue.html.

[9] Santos, D. C. D. (2022, November 14). Understanding the energy consumption of HPC scale Artificial intelligence. arXiv.org. https://arxiv.org/abs/2212.00582.

[10] Wengel, G. (n.d.). A Guide to Carbon-Aware Computing | Insights & Sustainability | ClimatIQ. https:// www. climatiq.io/blog/a-guide-to-carbon-aware-computing.