Team Control Number

15225

Problem Chosen

B

Summary

In particular, **High Powered Computing (HPC)** is increasingly providing support to scientific and industrial development; thus raising questions about its associated **energy consumption** and related **environmental impacts** which have become significant global concerns. To quantify HPC energy consumption and its resulting carbon emissions, and to provide actionable insights toward mitigation, we develop a multifaceted modeling framework in this study. We develop an integrated geospatial, predictive modeling and scenario analysis approach to assess the environmental footprint of the HPC systems now and in the future.

We start with an analysis based on our **Energy Consumption Model**. The model accounts for the total energy use of HPC systems based on computational demand, regional energy sources and technological advancements of energy efficiency. Our careful addition of **latitude and longitude** allows for adjustments for local climatic conditions, energy grid composition, and renewable energy potentials. This level of **geospatial precision** enables us to do better estimates of energy consumption than available generalized models.

Based on this, the **Carbon Dioxide Emission Model** combines energy consumption data with emission factors that are country and energy specific. The regional **energy mixes** characterized by this model allow for much more detailed analysis of carbon emissions in various locations. For example, regions with a high reliance on coal exhibit higher emissions than regions with a renewable predominant energy grid. This approach offers a microcosmic view of how global HPC deployment decisions and energy policies combine to impact carbon footprints.

In order to extend our findings into the future, we use an **ARIMA Forecasting model** to predict future trends in energy consumption and emissions. By modeling through historical data, this time series analysis tool outputs estimates of the extent by which future HPC related emissions may be influenced by changes in computational demand, energy efficiency improvements or renewable energy adoption. Finally, we can explore scenarios and examine what might happen if we assume accelerated renewable energy integration or technological stagnation, with the ARIMA model.

We conclude by calculating the current **environmental impacts** of HPC systems and provide future forecasting and **mitigation methods**. We provide a scalable policy driven solution by combining geospatial precision with predictive modeling. To have a sustainable future for HPC, we suggest focusing on adoption of renewable energy, increased efficiency based designs, and targeted investments in regions where HPC growth will make the smallest environmental impact.

Keywords: High Performance Computation (HPC); Data Center; Energy Consumption Model; Carbon Emission Model; ARIMA Method; Water Usage.

Contents

1 Introduction

1.1 Background

High-Performance Computing systems have become indispensable in driving even more advanced fields such as climatology, artificial intelligence, genetic data analysis, and the development of new materials. Paradoxically, this steep growth in HPC capability has ushered in considerable environmental concerns. Recent estimates indicate that global data centers, including HPC facilities, are calculated to consume about 1% of the world's electricity [2]. This is expected to rise even more as the HPC infrastructure continues expanding, putting further demands on energy resources and increasing carbon emissions.

The energy demands of HPC facilities are very high. A single supercomputer may consume up to 22.7 MW annually, equivalent to the energy usage of a few thousand households [3]. Assuming complete-year operation, estimates suggest that between 30 and 50 TWh of electricity are consumed annually in HPC centers around the world. Such huge quantities of energy put immense stress on already overburdened supplies, and as long as fossil fuels remain the predominant energy source, the effect on carbon emissions and climate change will only worsen.

The cooling mechanisms employed in large HPC facilities further amplify their environmental impact. Analyses predict that 30% to 50% of the total energy consumption in HPC systems is dedicated to cooling processes, which often involve substantial water usage and generate electronic waste. This has heightened the drive toward developing affordable, greener, and more sustainable cooling technologies, as well as searching for alternative energy sources. Additionally, HPC applications require increasingly higher power for CPUs and GPUs. Each successive generation of processors has significantly higher power consumption. Improvements in HPC infrastructure to support artificial intelligence and machine learning models are expected to drive computational power demands by approximately 10-15% annually, further increasing energy consumption.

1.2 Question Restatement

The following work covers the environmental aspect of the problem—the ecological impact of High-Powered Computing—motivated by its extremely high energy use for high-performance computation. In addition, our analysis will consider a wide range of energy sources: from traditional ones, like fossil fuels, to renewable energy and nuclear power. While the principal focus lies on the total CO_2 emissions associated with HPC operations, the investigation extends to other ecological concerns such as electronic waste and water consumption. Subsequently, the study aims to project future scenarios by considering the anticipated growth of HPC capabilities, policy developments, and shifts in energy resource utilization.

First, the study constructs a global model to estimate the annual energy consumption of HPC systems, accounting for both full capacity operations and average utilization rates.

Second, it develops an environmental model to calculate the total CO_2 emissions generated during typical HPC operations, based on energy mixes that include fossil fuels, renewable energy, and other sources.

Third, the study incorporates additional variables into the CO_2 emission model to account for the projected growth of HPC infrastructure and the anticipated evolution of energy portfolios, enabling the prediction of emission ranges for the year 2030.

Fourth, it investigates the environmental impact of transitioning to a 100% renewable energy-powered HPC infrastructure by modifying the model to assess the potential reduction in emissions.

Finally, the study enhances the model by introducing an additional variable to represent another significant environmental impact, such as water consumption or e-waste generation, and evaluates this aspect's importance and its interrelation with energy consumption and other key areas.

Through these objectives, we aim to provide a comprehensive framework for understanding and mitigating the environmental consequences of HPC systems.

The Figure 1 is the final model we developed in flow chart form:

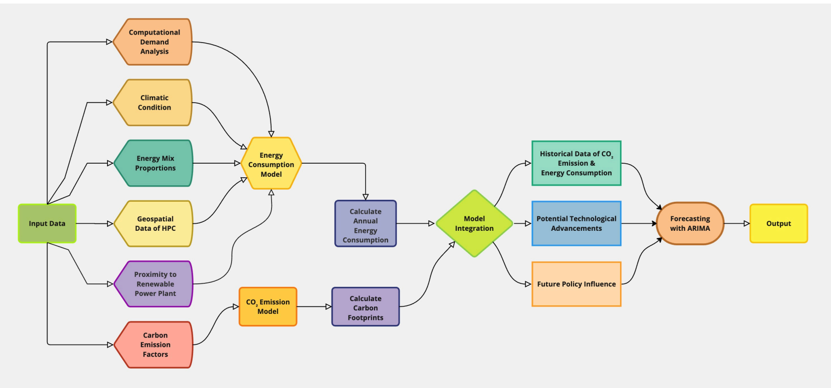


Figure 1: Flow Chart of Our Model

2 Model I: HPC Energy Consumption Model

In this essay, we define High-Powered Computing (HPC) as encompassing data centers worldwide, which serve as the primary target for assessment in terms of data availability and generalizability. The focus on large-scale data centers within the HPC ecosystem is due to their pivotal role as the core infrastructure for executing HPC's computationally intensive tasks, ranging from scientific simulations to machine learning. Moreover, data related to these facilities is comparatively more accessible. By examining these massive data centers, we aim to evaluate the primary factors driving HPC's energy consumption and the sustainability challenges they present [1].

2.1 Assumptions and Justifications

- **Assumption 1**: The annual average temperature for regions at similar latitudes remains stable over time.

  Justification: It is based on the consistency observed in long term climate patterns across latitude bands. While there are seasonal fluctuations, the annual average temperature at specific latitude intervals are relatively constant over time due to the Earth's axial tilt and solar radiation distribution, which create predictable temperature zones [4].

- **Assumption 2**: The average energy consumption of each Data Center could be estimated out of per unit floor area within a range.

  Justification: Estimating the energy consumption of each data center based on per-unit floor area is a feasible approach, as demonstrated by the ENERGY STAR methodology [5]. This approach uses typical values for energy use per square foot to approximate the energy demands of data centers, considering their high density and specific cooling needs. By leveraging reliable industry data on energy use intensity (EUI) per unit area, this assumption allows for a straightforward and standardized calculation of energy consumption across various data center sizes, making it both practical and widely applicable.

- **Assumption 3**: The energy consumption estimated using the per-unit floor area model is influenced by the local annual average temperature.

  Justification: Cooling systems are typically the largest energy consumers in data center operations, accounting for approximately 50% of total energy usage, as reported by Infotech Group [7]. Consequently, data centers located in cooler climates require less energy for cooling, leading to a reduction in overall energy consumption while maintaining high-performance operations. This relationship highlights the critical role of local climate in shaping energy demands.

2.2 Model Overview

To evaluate the annual energy consumption of data centers globally, it is crucial to account for both full capacity operations and average utilization rates. Accordingly, we aim to develop a model that estimates actual energy consumption using available data.

The foundation of our approach begins with an equation that defines the total energy consumption for a single unit within a data center. The equation is as follows:

$$E_{individual} = \bar{P} \times \Delta t \tag{1}$$

For modeling purposes, the elementary unit of analysis can be selected as processing units, nodes, or entire data centers. Regardless of the chosen unit, the model maintains consistency by leveraging the corresponding average power per device to establish accurate relationships.

Since this estimation is conducted on a yearly basis, the time interval is fixed at one year. Thus, the primary focus of the model lies in calculating the annual power usage per data center. Energy consumption for a single data center is determined by two primary components: energy used by IT equipment (e.g., networks, servers, and storage) and energy used by infrastructure facilities (e.g., power conditioning systems and cooling). Furthermore, advancements in HPC-related technologies and resulting improvements in energy efficiency are also integrated into the model.

Based on these principles, we will develop a comprehensive framework to estimate power usage for data centers. The input parameters considered in our model are outlined in Table 1.

To estimate the energy consumption of HPC systems, we will proceed as follows:

- Construct a temperature model in Section 2.3
- Model the impact of surrounding climate on HPC energy consumption in Section 2.4
- Develop an Energy Factor Model for HPC in Section 2.5
- Introduce an energy use per unit floor area model in Section 2.6

Table 1: Notations in Model I

Symbol	**Description**	**Unit**
A	Data Center gross floor area	ft^2
ε	Average energy use per unit floor area	kWh/ft^2
k	Climate impact factor on energy consumption	%
n	Number of data centers in the world	Number (#)
T	Annual average temperature	$^\circ C$
L_1	Latitude of the data center	Degrees
L_2	Longitude of the data center	Degrees
$\overline{P}$	Average power per data center	kW
t	Time interval	Year
ρ	Total power consumption factor	Unitless (#)
NC	Computational demand	Number of computations
EE	Energy efficiency	Computations/Unit Energy
$\overline{U}$	Average utilization rate of data center	Percentage (%)
E_i	Energy consumption for data center i	kWh
E_{total}	Total energy consumption across all data centers	kWh
$C_{Full,i}$	Full capacity of data center i	kWh
$C_{Full,total}$	Total full capacity across all data centers	kWh

2.3 Temperature Model

To calculate the annual average temperature at data center locations, we assume a linear relationship between latitude and temperature. As latitude decreases, the average annual temperature generally decreases, driven by the Earth's axial tilt and the distribution of solar radiation. This assumption aligns with statistical findings from prior research [8].

In a study involving 2,086 weather stations, researchers derived a statistically significant regression model that explains temperature variations across different locations. Notably, latitude and longitude were identified as the two most significant predictors of temperature, rendering other variables negligible for this analysis. Thus, we simplify our model by including only latitude and longitude as inputs.

Based on these findings, we define the local average annual temperature as a function of longitude and latitude:

$$T(L_1, L_2) = 39.2 - 0.56|L_1| - 0.04L_2 \tag{2}$$

The visualization of temperature as a function of longitude and latitude is presented below:

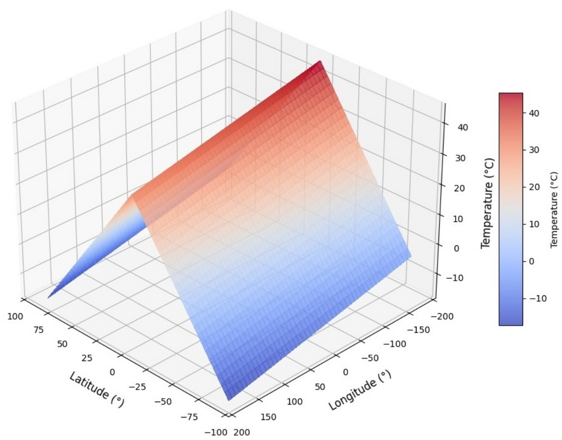


Figure 2: 3D Visualization of Temperature Based on Latitude and Longitude

2.4 Climate Impact on Energy Consumption of Data Center

Local temperature significantly impacts the energy consumption of data centers, primarily due to its effect on cooling systems. Lower ambient temperatures reduce the energy required to cool the equipment, while higher temperatures increase the demand for cooling. Therefore, we include local climate as a critical factor influencing energy consumption in our model.

Based on the findings of Ramachandra and Vikas, energy consumption increases by approximately 5% for every 1°F rise in temperature [9]. This relationship can be mathematically expressed as the following function:

$$Energy_{new} = Energy_{old} \times (1 + 0.05(T_{Fahrenheit})) \tag{3}$$

To ensure the applicability of the previously established relationship between energy consumption and climate to real-world data, we converted the temperature scale from Fahrenheit to Celsius:

$$Celsius = \frac{1}{1.8}(Fahrenheit - 32) \tag{4}$$

$$Energy_{new} = Energy_{old} \times (1 + 0.05 \times (\frac{1}{1.8}T_{Celsius})) \tag{5}$$

Thus, our climate impact factor can be represented as:

$$k(T) = 1 + \frac{1}{36}\Delta T \tag{6}$$

This implies that, relative to the baseline temperature, each 1 degree Celsius increase will result in a 1/36 increment in energy consumption.

To establish the baseline temperature(T), we account for the relationship between lower environmental temperatures and reduced energy consumption in HPC systems. We set the typical annual average temperature of a location midway between Calgary and Miami, with a latitude of 38.40584°North and an average temperature of 17.6927°C, as the baseline.

Thus, our model is rewritten as:

$$k(T) = 1 + \frac{1}{36}(T - 17.6927) \tag{7}$$

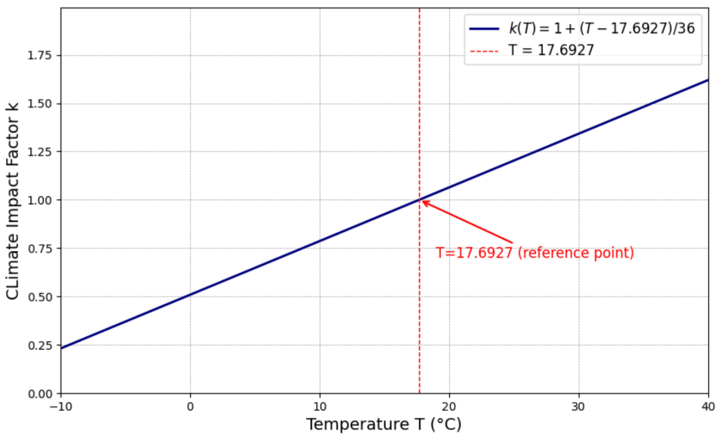


Figure 3: Climate Impact Factor as Function of Temperature

2.5 Energy Usage Factor Model for Data Center

To accurately estimate the overall trend in power consumption of high-performance computing (HPC) systems, we developed a predictive model incorporating computational demand and energy efficiency trends. Using the release of the official Energy Star report in August 2023 as a baseline, we define the energy efficiency factor for 2023 as 1. The model accounts for exponential growth in computational demand alongside improvements in energy efficiency, enabling projections of future power consumption.

According to Castro's research on HPC, the 12 times improvement in energy efficiency achieved between 2013 and 2022 is overshadowed by a 13 times increase in computational power demand [10].

Similarly, for data centers, increased demand has been offset by efficiency gains, resulting in a net power increase of 6% between 2010 and 2018 [11].

The total power consumption factor $\rho(t)$ of a data center system at a given year t can be expressed as:

$$\rho(t) = \frac{NC(t)}{EE(t)} \tag{8}$$

The computational demand for HPC systems grows exponentially, doubling approximately every 1.85 years, as evidenced by historical performance trends. This exponential growth can be mathematically represented as:

$$NC(t) = 2^{\frac{t-2023}{1.85}} \tag{9}$$

Conversely, energy efficiency—defined as the number of computations performed per unit of energy—also follows an exponential growth trend, doubling approximately every 2.29 years. This relationship can be expressed as:

$$NC(t) = 2^{\frac{t-2023}{2.29}} \tag{10}$$

By substituting the expressions for $NC(t)$ (computational demand) and $EE(t)$ (energy efficiency) into the base equation, we derive the formula for overall power consumption:

$$\rho(t) = \frac{2^{\frac{t-2023}{1.85}}}{2^{\frac{t-2023}{2.29}}} = 2^{0.10386(t-2023)}$$

Which indicates that power consumption per HPC system increases at an annual compounded rate of approximately $2^{0.10386} - 1 \approx 0.0746$, assuming current trends persist.

This value is consistent with Sifflez's regressive analysis of the top 100 systems in the TOP500 list, which found that the average power usage of individual HPC systems increased by a factor of 2.55 from 2010 to 2021 [10]. This corresponds to a mean annual increase in power usage of:

$$\textit{Annual Growth in Energy Demand} = 2.55^{\frac{1}{12}} - 1 \approx 0.08113106$$

Although the value differs slightly from our data—likely due to being derived from only 100 of the TOP500 supercomputers—it is more appropriate to use the factor derived from our model, as it considers a broader dataset. For this reason, we will employ the factor calculated by our model in the subsequent analysis.

2.6 Energy Use Per Unit Floor Area Model

The energy consumption of individual components can vary significantly across different data centers. While tracking IT energy usage is considered a best practice, it is often challenging to gather comprehensive global data simultaneously. Based on limited data from reference surveys and anticipated difficulties in collecting detailed operating conditions from Portfolio Manager users, it has been concluded that employing a single source Energy Utilization Intensity (EUI) value to estimate data center energy consumption is more practical than developing a more complex estimation method.

To estimate the total energy consumption of a data center, the EUI value is multiplied by the facility's floor area. However, this approach does not account for one of the most critical external factors influencing energy consumption: local temperature. Data centers must operate within an optimal temperature range to achieve peak performance, resulting in cooling systems being the dominant source of energy use. To address this limitation, our model incorporates a temperature factor, detailed in Section 2.4.

In our analysis, the typical energy use per unit floor area is derived from the Energy Star Technical Reference [5]. In the United States, this value is approximately $2,000kBtu/ft^2$, while in Canada, it is $20.35GJ/m^2$. For consistency and comparability, all values are converted into $kBtu/ft^2$. The estimated data, recalculated in these standardized units, are as shown in Table 2:

Table 2: Data Center Energy Usage Per Unit Floor Area

	United States	**Canada**
Source Energy	$2,000\ kBtu/ft^2$	$20.35\ GJ/m^2$
Conversion Value	$586.1421404\ kWh/ft^2$	$525.1602408\ kWh/ft^2$

From our analysis, we observed that the smaller value corresponds to Canada, while the larger value pertains to the United States. This disparity is likely due to climatic differences, as Canada is situated farther from the equator, whereas the United States is closer. Our model accounts for this climatic effect, as detailed in Section 2.4.

To minimize estimation errors stemming from the variation in energy use per unit floor area, we adopt the mean of these two countries' typical values as a representative metric for global data centers.

Thus, the final value used in this model is calculated as follows:

$$\varepsilon = \frac{\varepsilon_{US} + \varepsilon_{Canada}}{2} = \frac{586.1421404kWh/ft^2 + 525.1602408kWh/ft^2}{2} = 555.6511906kWh/ft^2$$

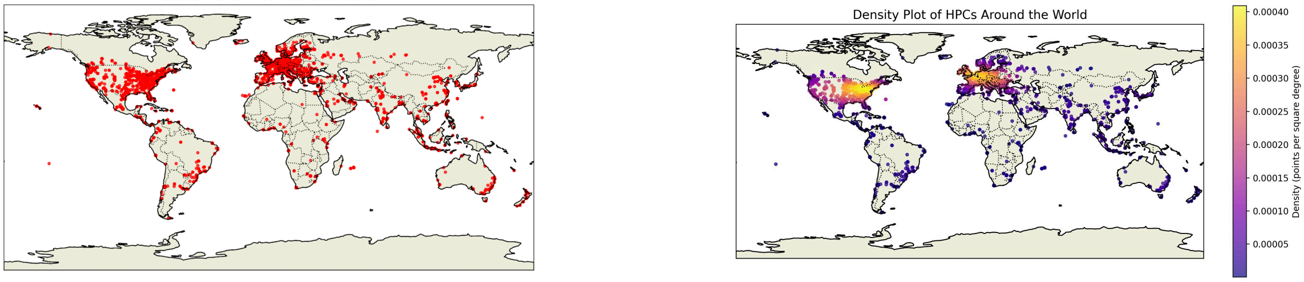
2.7 Annual Energy Consumption Model for Data Centers

Combining all the factors, the annual average energy consumption can then be expressed as:

$$E_{total} = \sum_{i=1}^{n} E_i(A, \varepsilon, L_1, L_2, t) = \sum_{i=1}^{n} \varepsilon \times \rho(t) \times A_i \times k_i(T(L_1, L_2)) \tag{11}$$

2.8 Annual Energy Consumption Results

Using data from Data Center Map [6], we compiled a dataset of global data centers, including details such as the data center's name, latitude, longitude, and floor area, as shown in Figure 4.



(a) Distribution of Data Centers Around the World

(b) Density Plot for HPC s Around the World

Figure 4: Geological Information of Data Centers Worldwide

2.8.1 Annual Energy Consumption Under Average Utilization Rate

Since our model is derived from approximations provided by ENERGY STAR, based on reports of actual data center utilization, we assert that the results generated by our algorithm represent the global energy consumption of data centers under average utilization rates.

For this analysis, we assume that the 3,216 data centers in our dataset are representative of typical data centers, accurately reflecting the general distribution patterns of data centers worldwide. By scaling the calculated values in proportion to the ratio of these 3,216 data centers to the estimated total number of global data centers, we can approximate the overall energy consumption of all data centers.

By apply the model here to the data set to all 3216 data centers, we get the following data:

$$E_{3216} = 3.74165 \times 10^{11} kWh$$

Based on the hypothesis that this value represents the global energy consumption of all data centers, the final global energy consumption can be estimated by adjusting it according to the proportional representation of the 3,216 data centers in the dataset relative to the total number of data centers worldwide.

$$E_{total} = E_{3216} \times \frac{8119}{3216} = 3.74165 \times 10^{11} kWh \times \frac{8119}{3216} = 9.44 \times 10^{11} kWh$$

The value of the estimation is equivalent to 944 TWh. This result aligns with the IEA report, which estimates that data centers globally consumed approximately 460 terawatt-hours (TWh) of electricity in 2022, with projections indicating that total consumption could exceed 1,000 TWh by 2026 [12].

2.8.2 Annual Energy Consumption Under Full Capacity

To estimate the energy consumption of data centers operating at full capacity, it is necessary to determine the maximum energy usage under optimal conditions. This requires identifying the average utilization rate in relation to total capacity.

To facilitate this analysis, data centers are categorized into three tiers: Internal, Midtier, and Hyperscale. The categorization criteria are defined in Table 3:

Table 3: Data Centers Categorization Standard

	Internal	**Midtier**	**Hyperscale**
Typical Size of Data Center	$< 20,000 ft^2$	$20,000 - 400,000 ft^2$	$> 400,000 ft^2$
Average Utilization rate from 2020 to 2030	15%	25%	50%

The rationale behind this categorization and the assumptions regarding average utilization rates is drawn from a U.S. data center energy usage report and additional research [13, 14].

Mid-tier data centers, which are primarily service providers, are assumed to operate at higher utilization rates compared to internal data centers, as their servers are often configured for specialized and predictable operations. Hyperscale data centers are expected to achieve even higher utilization rates than mid-tier providers and internal data centers due to the greater efficiency and utilization of servers in cloud-based environments compared to non-cloud environments.

From these assumptions, we can calculate the average utilization rate of each data center type, which can then be used to estimate the full capacity of all data centers based on the relationship:

$$\text{Full Capacity Energy Consumption} = \frac{\text{Energy Consumption}}{\text{Average Utilization Rate}} \tag{12}$$

As such, the full capacity can be measured by:

$$C_{Full,i} = \frac{E_i}{\bar{U}_i(A)} \tag{13}$$

In which, average utilization rate ($\bar{U}$) can be modeled as:

$$\bar{U}(A) = \begin{cases} 15\%, & \text{if } A < 20000 ft^2 \\ 25\%, & \text{if } A \in [20000, 400000] ft^2 \\ 50\%, & \text{if } A > 400000 ft^2 \end{cases} \tag{14}$$

In consequence, we measure the global data center energy consumption by:

$$C_{Full,total} = \sum_{i=1}^{n} C_{Full,i} = \frac{E_i}{\bar{U}_i(A)} \tag{15}$$

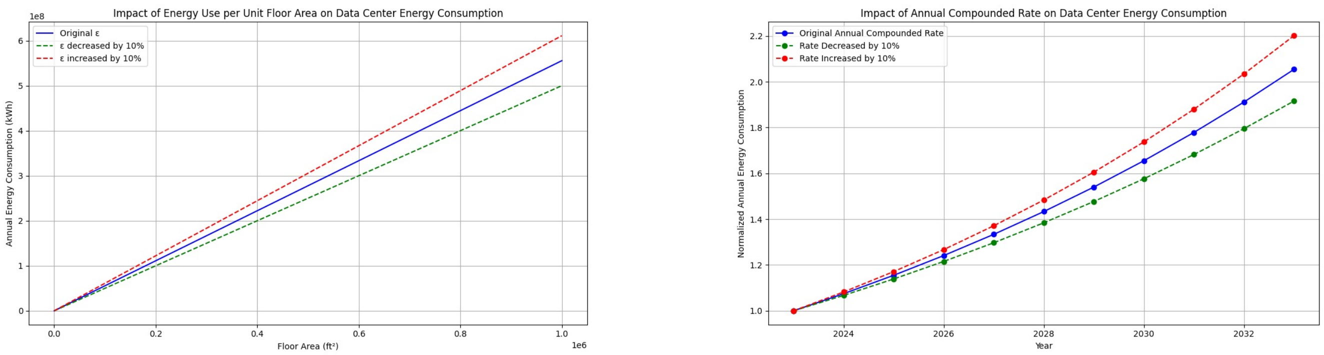
Applying this model to the previous dataset, we get the following result:

$$C_{Full,3216} = \sum_{i=1}^{n} C_{Full,i} = 2.49443 \times 10^{12} kWh$$

$$C_{Full,total} = C_{3216} \times \frac{8119}{3216} = 2.49443 \times 10^{12} kWh \times \frac{8119}{3216} = 6.297 \times 10^{12} kWh = 6297 TWh$$

2.9 Sensitivity Analysis

Lastly, we analyze the sensitivity of our model by changing Average energy use per unit floor area (ε) and total power consumption factor ($\rho(t)$) by 10%. After applying these changes, we obtained the results in Figure 5:



(a) Impact of Changing 10% of Energy Use per Unit Floor Area on Data Center Energy Consumption

(b) Impact of Changing in Annual Demand Growth Rate on Data Center Energy Consumption

Figure 5: Sensitivity Analysis of Changing ε and ρ by 10%

After incorporating the changes, we recalculated the energy consumption estimation using our model, as shown in Table 4:

Table 4: Sensitivity Test on Energy Consumption

Change	-10%	Original	-10%
Energy Consumption (kWh)	7.65×10^{11}	9.44×10^{11}	1.14×10^{12}

This represents an approximate 19% reduction and 21% increase in energy consumption for the decreased and increased growth rates, respectively. These substantial changes indicate that our model is highly sensitive to variations in the annual compounded growth rate, emphasizing the critical importance of accurately estimating growth rates in computational demand and energy efficiency. Small percentage changes in (t) and can lead to significant impacts on projected energy consumption, which has profound implications for energy planning and policy development in the context of high-powered computing infrastructures.

3 Model II: Environmental impact of Carbon Emission Caused by Energy Consumption of Data Center

3.1 Assumptions and Justifications

- **Assumption 1**: All relevant historical data concerning data centers and energy mixes is sufficiently accurate, and historical trends will persist over the next decade.

  Justification: We assume that the data available is reliable, even if some are derived from estimates or secondary research. While our model accounts for potential changes in energy production and industrial competition, we assume no sudden or revolutionary disruptions, such as pandemics or unforeseen global events, will occur in these areas.

- **Assumption 2**: Only seven types of energy sources—solar, hydro, wind, gas, nuclear, coal, and oil—are considered in our energy mix.

  Justification: These energy sources constitute the majority of the global energy mix, representing a balanced spectrum of renewable and non-renewable options with varying carbon intensities. This selection ensures the model captures key environmental and sustainability impacts while maintaining computational manageability (Citation).

- **Assumption 3**: The growth of High-Performance Computing (HPC), variations in energy mixes, and increasing energy demand across sectors occur on an annual basis rather than continuously.

  Justification: Companies and institutions typically report data annually, and predictive models rely on these updates. As such, changes in the above three aspects are assumed to happen yearly, in line with available reporting cycles.

- **Assumption 4**: The impact of various electricity sources on each data center's energy consumption can be estimated based on the percentage contribution of power plants within a defined radius.

  Justification: Energy transmission losses increase significantly over long distances, so we assume each data center primarily relies on energy sources within a proximate range (Citation: Average optimal electricity transmission distance). This assumption allows for a practical model to estimate each data center's specific energy mix and associated carbon emissions.

3.2 Model Overview

In this model, we evaluate the environmental impact of carbon emissions resulting from the energy consumption of data centers, with particular emphasis on the composition of energy mixes.

The input parameters considered in this analysis are outlined in Table 5.

Table 5: Notations in Model II

Symbol	Description	Unit
$CE_{total,i}$	Total carbon emission at time interval t caused by one data center	tonCO_2
CE_{total}	Total carbon emission at time interval t	tonCO_2
E_i	Total energy consumption of one data center	kWh
f_j	The proportion of contribution of energy source j to one data center	%
e_j	Carbon usage effectiveness of j energy source	$kgCO_2 \times kWh^{-1}$
p_k	Power generated by a single power plant of a specific energy source	kWh
d	Distance between the data center and a power plant	km
P_k	Available power to data center by one power plant.	kWh
P_s	All available power from all power plants	kWh
P_{total}	All available power to data center from all 7 energy sources	kWh
m	Number of power plants of energy type s	Dimensionless(#)
k	Index for power plants of a specific energy source	Dimensionless(#)
η	Transmission efficiency	Decimal
ΔT_G	Global mean surface temperature	$^{\circ}C$
E_{CO_2}	CO_2 emissions	tonCO_2
κ	Transient Climate response to cumulative carbon emissions	$^{\circ}C/10^{12}$tonCO_2
E_{total}	Total energy consumption by data centers	kWh
R_W	Water consumption rate	$liters/kWh$
WU	Water usage	$liters$

3.3 Carbon Usage Effectiveness Model

The Carbon Emission Factor (CEF) represents the amount of carbon emissions generated per unit of energy consumed, depending on the energy source type. Table 6 outlines the CEF values for several common electrical energy sources, highlighting that fossil fuels have the highest carbon emission factors.

Table 6: Carbon Emission Factor [15]

Energy Type	**Carbon Emission Factor** / $\text{tonCO}_2e \times kWh^{-1}$
Coal	0.968
Oil	0.890
Natural Gas	0.440
Solar Energy	0.053
Wind Energy	0.029
Nuclear Energy	0.015
Water	0.0135

3.4 Proportional Contribution of Electricity Model

In this model, the proportion of a specific energy type contributing to a data center's energy mix is determined by the contribution of power plants within an effective radius. We assume that the closer a power plant is to the data center, the greater the proportion of its generated power will be consumed by that data center.

To reflect this, we use the distance between data centers and power plants to determine whether a power plant contributes to a data center's energy consumption. This approach is based on the principle that institutions prioritize electricity from nearby sources to optimize transmission efficiency and minimize energy losses.

Thus, the power delivered to a data center by a specific power plant is calculated as:

$$P_k = \begin{cases} p_k \cdot \eta^{d_k}, & \text{if } d_k \leq 1500\,\text{km}, \\ 0, & \text{if } d_k > 1500\,\text{km}. \end{cases} \tag{16}$$

We assume the value of η (transmission efficiency) is set to 0.995, based on the standard performance of high-voltage AC transmission systems. Typical high-voltage AC transmission maintains efficiency for distances up from 300 to 500 km before significant energy losses occur, while Ultra High Voltage AC systems can extend this range from 1000 to 1500 km [16]. For power plants located beyond 1500 km from the data center, we assume they contribute no energy to the data center's consumption due to excessive transmission losses. The total power contribution from all power plants of energy type s to the data center is calculated as:

$$P_s = \sum_{k=1}^{m} P_k = \sum_{k=1}^{m} p_k \cdot \eta^{d_k}, \quad \text{where } d_k \leq 1500\,\text{km}. \tag{17}$$

The total power delivered to the data center from all energy sources is expressed as:

$$P_{\text{total}} = \sum_{s=1}^{7} P_s \tag{18}$$

Finally, the contribution of a specific energy type s to the data center's power consumption is given by:

$$f_j = \frac{P_s}{P_{\text{total}}} \tag{19}$$

3.5 Total Carbon Emission Model

To estimate the total carbon emissions caused by data centers, the following formula is used:

$$CE_{\text{total}}(t) = \sum_{i=1}^{n} CE_{\text{total},i}(t) = \sum_{i=1}^{n} \sum_{j=1}^{7} E_i(t) \cdot e_j(t) \cdot f_j(t) \tag{20}$$

3.6 Total Carbon Emission Result

We currently have data for only 3,216 data centers. Therefore, we assume that the emissions profile of these data centers is representative of all data centers globally. Using this assumption, we can estimate the total carbon emissions by scaling the data proportionally. Remind that since the numbers are huge, we converted them to $tonCO_2$ instead.

By applying the model described in Section 3.5, we derive the following results from 3216 data centers, as shown in Table 7:

Table 7: Total Carbon Emission From 7 Different Energy Sources

Energy Type	**Carbon Dioxide Emission** (Unit: *ton* CO_2)
Coal	79,594,471,968
Oil	10,801,844,983
Natural Gas	74,756,079,118
Solar Energy	709,813,394.2
Wind Energy	487,464,239
Nuclear Energy	593,864,273.7
Water	541,476,943.3
Total	1.67485×10^{11}

The total carbon emissions from all 8,119 data centers are calculated as:

$$CE_{total} = CE_{3216} \cdot \frac{8119}{3216} = 1.67485 \times 10^{11} \cdot \frac{8119}{3216} = 4.2282 \times 10^{11} \text{tonCO}_2$$

Thus, the total carbon emissions due to global data centers in the year of 2024 are approximately $4.2282 \times 10^{11}\text{tonCO}_2$.

3.7 Impact of Increasing Proportion of Renewable Energy

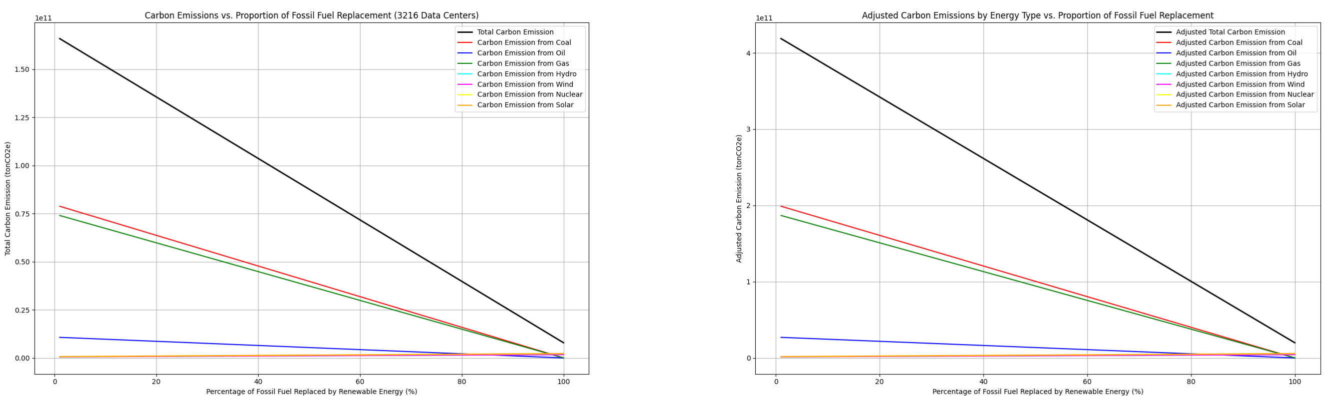
3.7.1 Change in Total Carbon Emission

The transition from fossil fuel-based energy sources to renewable energy presents a significant opportunity to reduce the carbon emissions associated with data center operations. This analysis investigates the reduction in total carbon emissions when the fossil fuel portion of the energy mix is replaced with renewable sources, while maintaining the current proportions of energy contribution from individual sources.

By modeling the impact of increased renewable energy adoption—from incremental changes to a complete transition to 100% renewable energy—we aim to quantify the potential carbon savings and address the technical and operational challenges associated with such transitions. This approach enables an evaluation of the scalability and effectiveness of renewable energy integration in achieving carbon neutrality for data centers.

On average, global data centers rely on fossil fuels for approximately 64.25% of their energy consumption. This figure is based on data from 3,216 data centers. Assuming this proportion is representative of data centers worldwide, we can scale this value by a factor to estimate the global situation.

Figure 6 below illustrates the detailed data on the impact of increasing the proportion of renewable energy on the 3,216 data centers, as well as its extrapolated effects on global data centers:



(a) Carbon Emissions vs. Proportion of Fossil Fuel Replacement (3216 Data Centers)

(b) Adjusted Carbon Emissions by Energy Type vs. Proportion of Fossil Fuel Replacement (8119 Data Center)

Figure 6: Estimation of change in total carbon emission of transmission toward 100% renewable energy

3.7.2 Potential Challenges

Transitioning to a 100% renewable energy system for data centers faces several significant challenges, particularly in the context of infrastructure, energy efficiency, and geographical constraints. These obstacles must be addressed to make such a transition feasible and sustainable on a global scale.

Firstly, the lack of available funds to invest in renewable energy infrastructure during the early stages of development poses a critical barrier. Building renewable energy systems requires substantial upfront capital, including investments in wind farms, solar arrays, and energy storage solutions. According to experts, many governments and private investors hesitate to commit the necessary resources due to the long payback periods and financial risks associated with renewable energy projects [17].

Secondly, renewable energy power plants often face efficiency limitations in energy transmission. High transmission losses can occur when transporting energy from remote renewable energy sources, such as solar farms or wind turbines, to urban centers or data hubs located in central areas. As a result, in some cases, the cost of transmitting renewable energy may exceed the cost of using other energy sources like natural gas, making renewables less economically viable for central operations [18].

Lastly, renewable energy requires spacious areas and specific geographical conditions to reduce loss during the transformation of energy. Solar panels need vast, unobstructed areas with high solar irradiance, while wind farms require locations with stable, high-velocity winds. These geographical requirements often conflict with urban land use, reducing the likelihood of building renewable energy plants near highly populated areas. This geological limitation further increases reliance on long-distance energy transmission and raises questions about land availability and environmental trade-offs [17, 18].

These challenges highlight the complexity of transitioning to a 100% renewable energy system. Bridging funding gaps, improving energy transmission efficiency, and addressing geological hindrances are critical issues necessary for renewable energy to become a practical solution for powering data centers worldwide.

3.8 Environmental Impact of Carbon Emissions Resulting from Energy Consumption of Data Centers

The environmental effects of carbon emissions from data center energy use represent a complex issue with major global implications. A key concern is their contribution to global warming, intensified by the rising demand for computational infrastructure. Data centers, which are largely powered by energy grids still dominated by fossil fuels, release significant amounts of CO_2, further driving climate change. Tackling these issues demands both technological progress in energy efficiency and a deliberate transition to renewable energy sources, alongside adopting more sustainable management practices for data centers.

This section involves one critical aspect of the environmental impact—global warming. By examining this issue, we aim to shed light on the far-reaching consequences of carbon emissions and underscore the pressing need for effective measures to mitigate their effects.

3.8.1 Global Warming Impact

Recent advances in climate modeling have made it clear how cumulative carbon emissions drive global warming. Scientists have demonstrated that the increase in GMST is directly proportional to the total amount of carbon dioxide (CO_2) released into the atmosphere. This relationship is quantified by the measure called Transient Climate Response to Cumulative Carbon Emissions (TCRE), which indicates how much the Earth's temperature rises for every trillion tons of CO_2 emitted [19, 20].

The TCRE value is estimated at 0.45°C per $10^3 Pg\mathrm{CO_2}$ emitted, with a 90% confidence interval ranging from 0.3 °C to 0.6 °C per $10^3 Pg\mathrm{CO_2}$ [21].

Using this value, we can model the relationship between carbon emissions and the GMST increase as follows:

$$\Delta T_G = \kappa \times E_{CO_2} \tag{21}$$

Substitute the global annual carbon emissions from data centers in the year of 2024 ($E_{CO_2} = 4.2282 \times 10^{11}\,\mathrm{tonCO_2}$) into the equation, the GMST is calculated as:

$$\Delta T_G = \kappa \times E_{CO2} = \frac{0.45°C}{10^{12}\ \mathrm{tonCO_2}} \times 4.2282 \times 10^{11}\ \mathrm{tonCO_2} = 0.190269\ °\mathrm{C}$$

This value suggests that the yearly carbon emissions by global data centers cause around a 0.190269 °C rise in GMST. While this may intuitively seem too high, it is important to remember that this is a simplified calculation and does not consider factors that mitigate global warming, such as carbon absorption by forests or oceans. Nonetheless, it highlights the significant impact of data center carbon emissions on climate change and underscores the urgent need to transition to more sustainable energy practices.

3.9 Refined Model of Environmental Impact of Carbon Emission Caused by Data Centers

To refine our model, we incorporated the water usage associated with energy consumption in data centers. Water usage is a critical yet often overlooked factor that significantly impacts both electricity generation and data center cooling.

Our model quantifies water usage by linking it directly to the energy consumption of data centers:

$$WU = E_{total} \times R_W \tag{22}$$

Based on the total energy consumption of global data centers ($E = 9.44 \times 10^{11}$ kWh), Using an average water consumption rate of 7.6 liters per kilowatt-hour (kWh), derived from thermoelectric and hydroelectric generation processes [22], we can estimate the water usage for global data centers, the water usage is calculated as follows:

$$WU = 9.44 \times 10^{11} kWh \times 7.6 \text{ liters} = 7.144 \times 10^{12} \text{ liters}$$

Meaning the global data centers along will consume 7.144×10^{12} liters of usable water on a yearly basis.

This aspect was chosen due to its dual relevance to energy consumption and sustainability concerns, particularly in regions where water resources are limited or stressed. By highlighting the interdependence between energy and water, our refined model emphasizes the broader environmental implications of HPC operations, advocating for solutions such as renewable energy integration and advanced cooling systems to mitigate water usage without compromising computational performance.

4 Prediction of Data Centers

4.1 Assumptions and Justifications

- **Assumption 1**: While the evolution of essentially all real-life issues are arguably piecewise and interwoven, we are going to approach the development of data centers, their energy consumption, as well as environmental impacts as isolated events, thereby ignoring the potential noise term due to complexities of the physical universe in our model.

  Justification: Natural disasters, international politics, technological breakthroughs are but a few among the myriad of events unpredictable by simple math models. Here we attempt to minimize such disturbances in our model by focusing exclusively on specific factors, namely the construction of data centers, temperature changes, and development in technology that improves the energy efficiency. This ought not to deviate our predictions by a significant amount considering how other external factors follow a normal distribution and cancel out the effects of each other.

- **Assumption 2**: The development of technology regarding HPC devices is expected to advance along a smooth trend. In other words, the effect of specific breakthroughs are overlooked in the greater scale as pieces of them add together contributing to a rather gradual process.

  Justification: Specifically in terms of energy efficiency, carbon emission, data center numbers, and other relevant factors contributing to the future scenario of HPC, each and every step along their development is untraceable and therefore unpredictable. We take a step back to look at the general rate of advancement and expect it to obey the same pattern as the past.

- **Assumption 3**: We used the time series model ARIMA to imitate the behavior of each section. For the specific model to function, we assume stationary data indicative growth patterns from the past.

  Justification: This assumption will be mathematically justified in the later section.

4.2 Model Overview

Recall from section 2.6 that the total energy consumption of Data Centers obey the model:

$$E_{total} = \sum_{i=1}^{n} E_i(A, \varepsilon, L_1, L_2, t) = \sum_{i=1}^{n} \varepsilon \times \rho(t) \times A_i \times k_i(T(L_1, L_2)) \tag{23}$$

And the carbon emission resulting from such consumption is given by:

$$CE_{\text{total}}(t) = \sum_{i=1}^{n} CE_{\text{total},i}(t) = \sum_{i=1}^{n} \sum_{j=1}^{7} E_i(t) \cdot e_j(t) \cdot f_j(t) \tag{24}$$

We set off from these models and continue to include the effect of time in necessary terms. Major effects of time come in growth of data centers, increasing demand for energy in other sectors, and potential differences in energy sources and mixes. A general model of Energy consumption and its carbon footprint is therefore a product of the exact same terms, except adding to them the effect of time.

Therefore, many of the the notation is this section is recurrent from previous sections, as shown in Table 8:

Table 8: Notations in Model I & II

Symbol	Description	Unit
A	Data Center gross floor area	ft^2
ε	Average energy use per unit floor area	kWh/ft^2
k	Climate impact factor on energy consumption	%
n	Number of data centers in the world	Number (#)
T	Annual average temperature	$°C$
L_1	Latitude of the data center	Degrees
L_2	Longitude of the data center	Degrees
t	Time interval	Year
ρ	Total power consumption factor	Unitless (#)
NC	Computational demand	Number of computations
EE	Energy efficiency	Computations/Unit Energy
$\overline{U}$	Average utilization rate of data center	Percentage (%)
E_i	Energy consumption for data center i	kWh
E_{total}	Total energy consumption across all data centers	kWh
E_i	Total energy consumption of one data center	kWh
f_j	The proportion of contribution of energy source j to one data center	%
e_j	Carbon usage effectiveness of j energy source	$kgCO_2 \times kWh^{-1}$
$h(t)$	Predicted number of hyperscale data centers	#

4.3 Time Dependent Energy Consumption of Data Center

In order to incorporate the effect of time in the model, the emphasis lies in moderating each term with respect to time. ε and $\rho(t)$ combined provides accurate insights to the use of energy per unit area, where ε is a constant determined by survey and $\rho(t)$ is elaborately modeled to be a function of time in section 2.5, however, only k depends on the temperature at the location, therefore the problem simplifies to rewriting $T(L_1, L_2)$ as $T(L_1, L_2, t)$.

Usually when time is involved in a function of temperature, the focus lies in its seasonal changes. In this case, however, our model approaches the problem on a yearly basis. The function yields the average temperature of a location over the entire year, thus its results are independent of the season as we input the year number as positive integers.

Another factor that may potentially lead to temperature changes is precisely the focus of this essay: global warming as a result of carbon emission. According to climate.gov, the average global temperature has been increasing at the rate of roughly 0.06 °C per decade [23]. This can be incorporated into the model as follows:

$$T(L_1, L_2, t) = 39.2 - 0.56|L_1| - 0.04L_2 + 0.006(t - 2024) \tag{25}$$

However, considering the scope of our investigation and potential variance that may occur over long periods, we consider the effect from global warming too trivial for any significant effect, thereby ignoring it in the model[23].

A, the gross area of HPCs worldwide, is the more interesting focus. In the previous model areas of certain known data centers are averaged and expanded to the entirety of HPCs. In this section for a time dependent version, we turn to data from Srgresearch to perform a time series analysis. We assume the percent growth in hyperscale HPC number to be indicative of the percent growth in HPC area [24]. In short, we can simulate the growth by the follwing function:

$$A(t) = \frac{A(2024) \times h(t)}{h(2024)} \tag{26}$$

ARIMA is specifically selected for this model. Assuming the data stationary, ARIMA is ideal for time series analysis for its easy applicability and customization of parameters.

After tuning with the adf test and running through auto-arima with the python library pmdarima, we select the parameters to be (2,0,0), respectively standing for a lag order of 2, 0 degrees of differentiation, and 0 window of moving average. The parameters align with that of a second order auto-regressive model [25].

Mathematically, the standard model for ARIMA forecasting is [26]:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \tag{27}$$

where $\hat{y}_t$, the predicted function, at t is given by μ, a constant change term; the y value at previous times scaled by a factor ϕ; and e stands for the moving average scaled by another factor theta.

In our model, it is simplified to [27]:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{28}$$

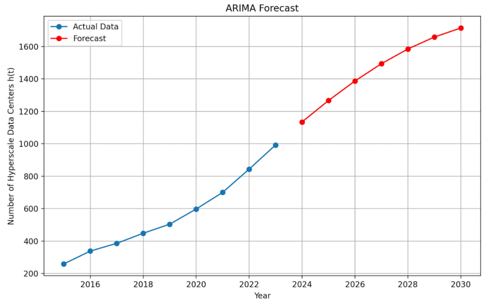
By substituting the former values into this equation, we get:

$$y_t = 18.0616 + 1.94y_{t-1} - 0.9662y_{t-2} + \epsilon_t \tag{29}$$

According to this equation, we ran the ARIMA forcast model, the results are presented in Figure 7:

```
Best model:  ARIMA(2,0,0)(0,0,0)[0] intercept
Total fit time: 10.952 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                    9
Model:               SARIMAX(2, 0, 0)   Log Likelihood                 -46.156
Date:                Tue, 19 Nov 2024   AIC                            100.311
Time:                        17:28:55   BIC                            101.100
Sample:                    12-31-2015   HQIC                            98.609
                         - 12-31-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     18.0616     29.216      0.618      0.536     -39.200      75.323
ar.L1          1.9460      0.155     12.520      0.000       1.641       2.251
ar.L2         -0.9662      0.138     -7.011      0.000      -1.236      -0.696
sigma2       592.2897    446.936      1.325      0.185    -283.689    1468.268
===================================================================================
Ljung-Box (L1) (Q):                   0.64   Jarque-Bera (JB):                 0.42
Prob(Q):                              0.42   Prob(JB):                         0.81
Heteroskedasticity (H):               0.69   Skew:                            -0.32
Prob(H) (two-sided):                  0.77   Kurtosis:                         2.14
===================================================================================
```

(a) ARIMA Model Statistics



(b) ARIMA Forcast Result

Figure 7: ARIMA Forecast on Number of Hyperscale Data Centers

It is worth noting that the dataset of eight elements is relatively small and may lead to misalignments. This reduces the range of reliability of our model to the next 5 years for the precision.

Combining with the general function and comparing with the current value for convenience in calculation, we have:

$$9.44 \cdot 10^{11} \cdot 2^{0.10386(t-2024)} \cdot \frac{18.0616 + 1.9460n(t-1) - 0.9662n(t-2) + \epsilon}{1125} \tag{30}$$

By running the simulation, the prediction results are as shown in Figure 8, the predicted energy consumption in the year of 2030 is $2026TWh$:

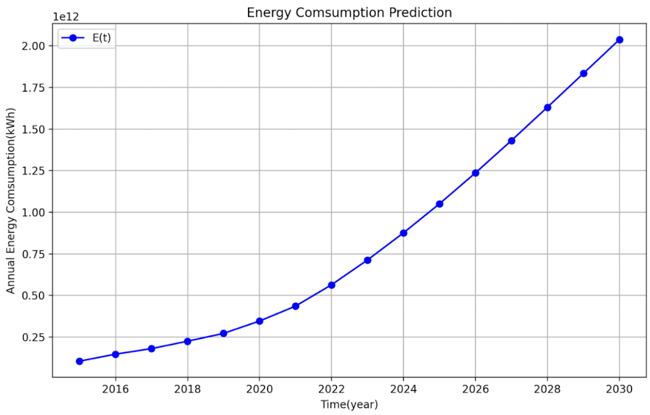


Figure 8: Energy Consumption Prediction Model

4.4 Time Dependent Change in Carbon Emission of Data Centers

While the original model provides an in-depth examination based on individual location of data centers, it is not within a predictive model's capabilities to determine the future location of specific HPCs. Instead, a generalized version is necessary, allowing vital trends and directions to be observed.

Instead of the summation over each and every data center, the revised model averages over given data aiming to find a general relationship. We simply take the results about total energy consumption from the previous model, divide up the energy by sectors, and account for the carbon emission of each. The model is therefore presented in the form:

$$CE_{total} = \sum_{j=1}^{\tau} f_j(t) \times E_{total}(t) \times e_j(t) \tag{31}$$

E_{total} is well-established in the previous section, while e_j the emission from a specific resource is dependent upon multiple interdependent technological requirements, thus is not expected to change by significant amounts in the near future. $f_j(t)$, the future proportion at which each type of resource generates energy is the center of focus. Note that in the previous section this is given by specific power plants located around each data center, while here it stands for an average proportion of energy generation around the globe. In other words, it is assumed that all data centers combined consume energy without preferences, thus the proportion of different sources is equivalent to that of energy generation.

We similarly apply an ARIMA model to each of the sectors to find its growth rate compared to the present by using the historical data from Statista [28, 29, 30, 31, 32, 33, 34]. The ratio is then based on current data to indicate the complete model. Repeating the ARIMA prediction as stated in previous section, we get the results of the ARIMA predictions for the year 2030 are summarized in Figure 9:

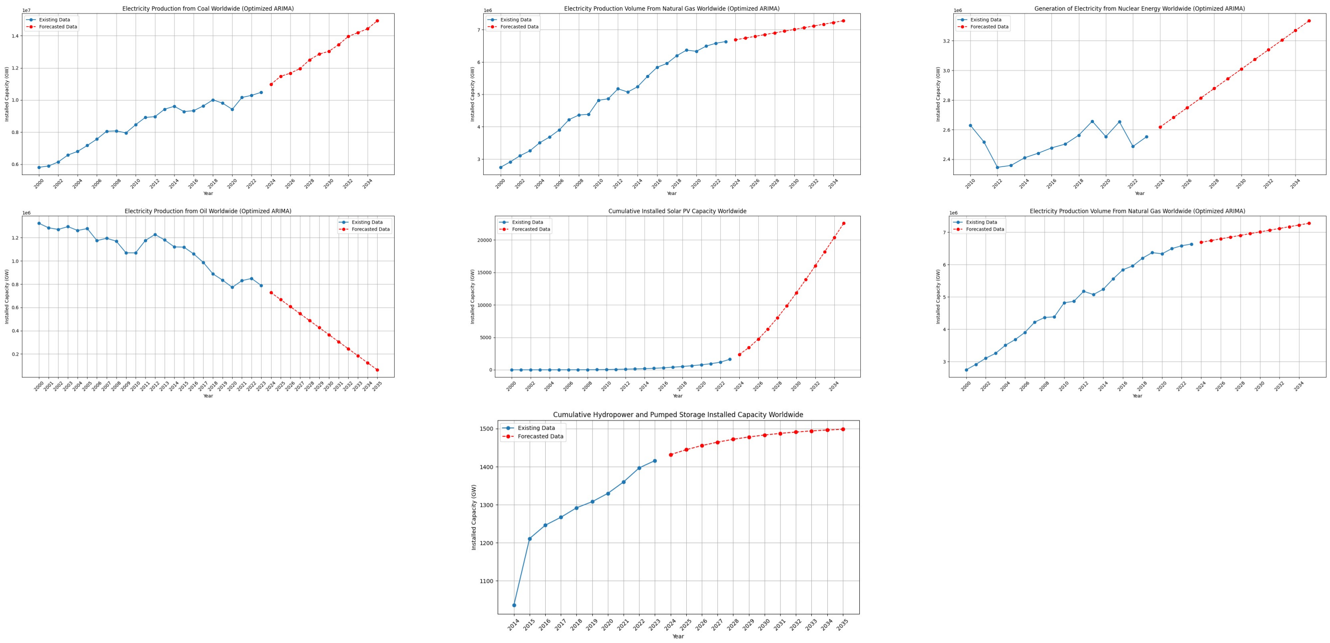


Figure 9: Prediction Using ARIMA Model from year of 2024 to 2035

To gain insights for the year of 2030, we extract the data from out ARIMA prediction from that year, and presenting as Table 9:

Table 9: Predicted Total Carbon Emission From 7 Different Energy Sources in 2030

Energy Type	Global Capacity(GWh)	Proportion to Total Capacity
Coal	13029730.41	55.61359357%
Oil	365970	1.562035913%
Natural Gas	7009480	29.91791538%
Solar Energy	11833.07545	0.050506022%
Wind Energy	1791.852381	0.007647998%
Nuclear Energy	3008750	12.84196943%
Water	1483.451351	0.006331678 %
Total Capacity	23429038.79	100%

By applying the previously developed model to calculate the total carbon dioxide emissions, and combining it with the carbon dioxide emission factors from Table 6, the result is obtained as follows:

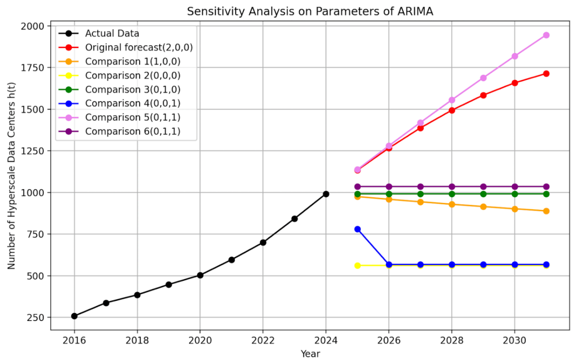
$$CE_{total} = \sum_{j=1}^{\tau} f_j(t) \times E_{total}(t) \times e_j(t) = 1.1228 \times 10^{12} \text{tonCO}_2$$

Thus, we can see that the prediction of carbon dioxide emission in the year of 2030 is estimated as 1.1228×10^{12}tonCO$_2$.

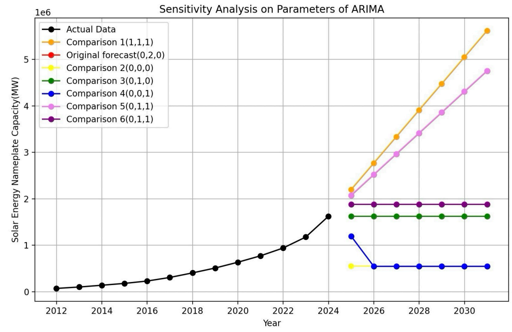
4.5 Sensitivity Analysis on ARIMA Forecasting

As we set off to review the accuracy of our prediction, sensitivity analysis is a vital portion allowing us to cope with potential flaws and determine their effects.

Here we focus on the selection of parameters in the ARIMA model, as it is the only major selection in the process in contrast to other given data. As we examine a few other possible parameters, it is clearly demonstrated in the graph how such a variety of results are possible from the slightest changes, as demonstrated in Figure 10. This, nevertheless, is not a justification for the model's fragility but instead demonstrates its precision.



(a) Sensitivity Analysis on Parameters of ARIMA Model



(b) Sensitivity Analysis on Parameters of ARIMA

Figure 10: Sensitivity Analysis on Parameters of ARIMA

The modified versions are separated into three groups: a growing group, a stationary one, and a group showing a downward trend. Judging from the previous data by common sense, it is obvious that the number of data centers is on an upward trend and is expected to continue increasing in the future. Among these available selections our chosen data is one of the few that correctly represents the relationship, thus confirming our correct selection of parameters. The energy models similarly follow the same trend.

5 Strengths and Weaknesses

5.1 Strengths

- By utilizing real-world data from thousands of data centers globally, our model accurately reflects current industry practices, which enhances the **credibility and applicability** of our findings.
- The model takes into account a broad range of factors that influence energy consumption and environmental impact at a local scale, such as climate variations and different energy mixes. Its framework allows for **application to almost any region**, as it can be easily updated with new data or technological advancements.
- Although comprehensive, the model remains **straightforward**. It simplifies complex calculations into manageable components, enabling **ease of implementation** and encouraging further exploration or extension by incorporating additional factors.

5.2 Weaknesses

- However, the model may **miss certain factors**, such as unexpected technological advancements or shifts in data center utilization, which could affect the accuracy of the predictions.
- Our forecasts assume that **historical trends** in energy consumption, technological development, and electricity generation remain stable, which might not hold true in the face of significant changes.

6 Recommendations on Technical and Policy Oriented Solutions

6.1 Technical Recommendations

6.1.1 Adjusting Inlet Air Temperature

To address the environmental challenges posed by High-Performance Computing (HPC) systems, practical technical solutions have been designed to reduce energy use and emissions while maintaining high efficiency. One effective strategy is to increase the inlet air temperature in data centers. Traditionally, inlet air temperatures were maintained between 20 and 25 °C due to historical norms and manufacturer guidelines [37]. However, advancements in server technology now enable modern equipment to operate reliably at higher temperatures, in line with ASHRAE's updated guidelines.

Studies show that increasing inlet air temperatures can reduce cooling energy consumption by from 2% to 6% per degree, depending on climate conditions. High-temperature data center designs, which target inlet air temperatures of up to 41°C, have demonstrated cooling energy savings of from 13% to 56% through the use of free cooling methods [35]. This adjustment significantly reduces reliance on energy-intensive chiller-based cooling systems while enhancing energy efficiency and sustainability.

By adopting higher inlet air temperature standards, data centers can achieve substantial reductions in cooling energy consumption, aligning with global goals to reduce carbon emissions and environmental impact while maintaining optimal operational performance.

6.1.2 Transition to Renewable Energy Sources

Improve the share of renewable energy in supplying power to the data centers by acting on the preference for on-site or regional solar and wind energy integration, reducing dependency on fossil fuels, and hence shaving the greenhouse gas emissions to improve the sustainability of HPC operations.

6.1.3 Dynamic Power Allocation

Smart power management systems deploy the capacity to optimize energy use through runtime dynamics dependent on workload demand. For example, less urgent computations can be shifted out of peak periods of energy consumption and yet meet operational goals with reduced energy consumption.

6.1.4 Implement Efficient Hardware

The shift to energy-efficient processors, GPUs, and other hardware that maximizes performance per watt is crucial and will cut down substantially on energy use. Modern advances in low-leakage and high-performance components help reduce the baseline energy demand of computations, thereby enabling more sustainable systems with no loss in capability.

6.1.5 Optimize Data Center Design

The renovations of the data center layout can work wonders to enhance airflow, hence reducing hotspots and reliance on power-intensive air cooling mechanisms. Even simple adjustments, like aligning the server racks for better airflow and making better usage of the available airflow, can achieve significant energy economies.

6.2 Policy Oriented Recommendations

In this respect, governments and industrial leaders should address policies that will seriously take into consideration the integration of renewable energy in combination with best energy-efficient practices in data centres, with the aim to reduce the environmental impact of the HPC systems. This would be a serious step towards motivating the shift to 100% renewable energy by offering tax incentives, grants, and subsidies to companies investing in solar, wind, and hydropower. Simplification in the processes for approving the renewable energy projects and access to key resources like land and water for the construction of renewable energy facilities near data centers would further help in this transition. Such initiatives would eventually build a dependable and sustainable supply of energy to data centers, meeting the increasing computational needs while reducing its carbon footprint. Finally, there are regulations that would require the application of high-technology cooling systems, such as high-temperature cooling systems or hydrocooling, that could reduce energy consumption due to cooling by up to 56%.

This can also involve setting targets of energy efficiency by governments, such as reaching a PUE of 1.5 or less, to ensure that data centers become more efficient in their operations. Further steps may include encouraging the phase-out of conventional hard drives in favor of power-saving SSDs, and encouraging the virtualization of servers that utilize full resources available to them while reducing the number of physical servers that consume energy. Coupling these policy drives with ongoing technology advances could potentially bring such HPC system footprints to a minimum. Embedding financial incentives with sustainability goals will not only support the leaps and bounds of growth in industries involving HPC and AI but also ensure such growth occurs in an environmentally non-destructive way.

6.3 Impact of Recommendation on Model I

Section 6.1.1 presents guidelines that enhance cooling systems, such as increasing the intake air temperature or utilizing advanced cooling technologies, which will offer significant energy saving yet feasible to implement. Currently, as much as 50% of total energy consumed in HPC systems is estimated to go into the consumption of cooling systems [7]. This energy consumption for cooling can be reduced, on average, by 34.5% by increasing the intake air temperatures or simply switching to more efficient ways of cooling; this reduces overall energy by 17.25%. These modifications need only limited upgrades to existing facilities and are generally based on existing ASHRAE standards, so they represent a feasible and cost-effective way to increase energy efficiency and promote sustainability.

To understand how this recommendation could influence global energy consumption, let's consider that all data centers worldwide adopt optimized conditions with an intake air temperature of 41°C. If cooling accounts for 50% of the total energy and the proposed measures reduce cooling-related energy usage by 34.5%, the revised energy consumption per unit area (ε_{new}) can be expressed as:

$$\varepsilon_{\text{new}} = \varepsilon_{\text{old}} \cdot (1 - (0.5 \cdot 0.345)) = \varepsilon_{\text{old}} \cdot 0.8275$$

This adjustment contributes to a 17.25% reduction in total energy consumption per unit area. Here, we can see that ε_{new} can be seamlessly integrated into the global energy consumption formula, providing a straightforward method to quantify the practical impact of this recommendation.

Incorporating this reduction into the total global energy consumption model, the updated energy consumption, E_{after}, is calculated as:

$$E_{\text{after}} = E_{\text{before}} \cdot \frac{\varepsilon_{\text{new}}}{\varepsilon_{\text{old}}} = 9.44 \times 10^{11}\,\text{kWh} \cdot 0.8275 = 7.81 \times 10^{11}\,\text{kWh}$$

This result indicates that if all data centers worldwide adopt this technical recommendation, the global energy consumption by data centers would decrease to 7.81×10^{11} kWh, representing a significant improvement in energy efficiency.

7 Letter

HIMCM Team #15225
November 15th, 2024

Advisory Board
The United Nations
New York, NY, United States

Dear Members of the United Nations Advisory Board,

We are reaching out to emphasize the critical need to address the environmental impact of High-Performance Computing (HPC) in an updated version of your report, "*Governing AI for Humanity.*" We strongly urge its inclusion as a key focus area in the United Nations' 2030 Development Goals. The rapid growth of AI model training and large-scale data processing has dramatically increased dependence on HPC systems, bringing about significant environmental challenges that demand immediate attention.

Escalating Environmental Footprint of HPC Systems
Our research utilized comprehensive mathematical models to demonstrate the growing environmental footprint of HPC systems. Currently, the global computing sector accounts for approximately 3% of worldwide electricity consumption, with HPC contributing a significant share due to its intensive operational requirements.

- Energy Consumption Projections: We estimate that HPC's annual energy consumption will increase from 944 terawatt-hours (TWh) in 2024 to approximately 2,026 TWh by 2030, assuming a continued annual growth rate of around 7.5%. This surge is primarily driven by the escalating demands of AI and data processing tasks.
- Carbon Dioxide Emissions: Based on current global averages, this energy consumption translates to an increase in carbon dioxide emissions from 0.423 billion metric tons in 2024 to about 1.123 billion metric tons by 2030. This represents a significant rise that could hinder efforts to meet global climate targets.
- Water Usage and Electronic Waste: HPC's environmental impact extends beyond energy consumption to include substantial water use for cooling—approximately 7.144 trillion liters annually—and increasing volumes of electronic waste due to rapid hardware obsolescence.

Environmental Benefits of Transitioning to Renewable Energy
By modifying our models to explore a shift to 100% renewable energy sources, we found that such a transition could:

- Reduce Carbon Emissions Significantly: Transitioning could lower HPC-related carbon emissions by up to 95%, given that renewable energy sources like wind and solar have minimal associated emissions.
- Alleviate Water Stress: Renewable energy technologies often require less water than conventional power plants, thus decreasing the strain on water resources.

Recommendations to Mitigate HPC's Environmental Impact
Our findings underscore the importance of adopting measures to minimize the ecological footprint of HPC systems. We propose the following approaches:

1. Optimize Cooling Efficiency:
    - Increase Intake Air Temperature: Raising the intake air temperature by 1°C can decrease cooling energy consumption by 2–6%, depending on climatic conditions.
    - Implement Efficient Hardware: A commitment to efficient hardware is by using energy-efficient processors, GPUs, and other hardware that ensures the most performance per watt. Low-leakage, high-performance components can lower the baseline energy needs for computations and help improve efficiency and sustainability at a data center.
2. Transition to Renewable Energy Sources:
    - On-site or Regional Renewable Integration: Data centres should move in the direction of integrating on-site or locally sourced renewable energy resources, such as solar and wind. Companies such as Digital Realty have shown that a shift to 100% renewable energy usage is feasible and sometimes advantageous to reduce the carbon footprint of the data centre operations.

- Policy Incentives: Governments can extend the much-needed impetus through the provision of tax breaks, grants, and subsidies to organizations investing in renewable energy systems. Further streamlining of regulatory procedures and facilitating access to essential resources like land and water resources will definitely catalyze the momentum to switch over to renewable sources of energy when such facilities are built around data centers.

3. Enhance Energy Efficiency Standards:

4. - Set Stringent PUE Targets: Set up strict, compulsory standards for lower PUE ratios that will ensure data centers are optimized in energy use. This would ensure a reduction in grossly wasted energy by making the operations of data centers more efficient and sustainable.
   - Energy-Efficient Hardware: Encourage customers to leverage energy-efficient processors, GPUs, and storage technologies. The evolved hardware solutions assure high performance with minimal energy use inbuilt to decrease the baseline energy use in data center operations.

Call to Action
We strongly stand by the view that the environmental impact of the HPC systems has to be an integral part of technological innovation in order to be compatible with the goals of sustainable development. The inclusion of such a concern, part of the "*Governing AI for Humanity*" report and the UN 2030 Development Goals, allows the Advisory Board:

- Highlight the Urgency: Bring global focus to the pressing environmental challenges created by HPC systems.
- Encourage international collaboration with respect to finding solutions and adopting sustainable HPC practices.
- Guide Policy Development: Help shape policies for sustainable technological advancement.

We call upon the Advisory Board to treat this as a priority issue and incorporate these recommendations in your 2030 development goals. Our future, at best consisting of both AI progress and human knowledge progress with seamless continuity, depends on the action not being postponed in the process of protection.

Thank you for bringing this urgent matter to your attention. Feel free to contact us if you need any more information or wish to assist with developing plans for reducing the environmental impact of HPC systems.

Sincerely,

HIMCM Team #15225

References

[1] Global Data Center Electricity Use to Double by 2026: Report. https://www.datacenterdynamics.com/en/news/global-data-center-electricity-use-to-double-by-2026-report/. Accessed: 2024-11-20.

[2] International Energy Agency, *Data Centres and Data Transmission Networks*. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks. Accessed: 2024-11-20.

[3] TOP500 Project, *TOP500 List*. https://top500.org/lists/top500/2024/06/. Accessed: 2024-11-20.

[4] Fatih Sevgin and Ali Ozturk, "Variation of Temperature Increase Rate in the Northern Hemisphere According to Latitude, Longitude, and Altitude: The Turkey Example," *Scientific Reports*, vol. 14, no. 1, pp. 18207, 2024. https://doi.org/10.3406/ecmed.2014.1269.

[5] ENERGY STAR, *ENERGY STAR Score for Data Center Estimates in the US and Canada*. https://www.energystar.gov/buildings/tools-and-resources/energy-star-score-data-center-estimates-us-and-canada. Accessed: 2024-11-20.

[6] DataCenterMap, "Data Center Map." https://www.datacentermap.com. Accessed: 2024-11-20.

[7] Jawad Haj-Yahya *et al.*, *Energy Efficient High Performance Processors*. Springer Singapore, 2018. https://doi.org/10.1007/978-981-10-8554-3.

[8] Marta Rocha and Francisco Rego, "Climatic Patterns in the Mediterranean Region," *Ecologia Mediterranea*, vol. 40, 2014. https://doi.org/10.3406/ecmed.2014.1269.

[9] Vikas Ramachandra, "Forecasting the Effect of Heat Stress Index and Climate Change on Cloud Data Center Energy Consumption," https://doi.org/10.13140/RG.2.2.18802.86724. Accessed: 2024-11-20.

[10] Sifflez, "Environment and High-Performance Computing". https://sifflez.org/publications/environment-hpc/. Accessed: 2024-11-20.

[11] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, Jonathan Koomey, "Recalibrating Global Data Center Energy-Use Estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020. https://www.science.org/doi/abs/10.1126/science.aba3758.

[12] International Energy Agency, *Electricity 2024*. IEA, Paris, 2024. https://www.iea.org/reports/electricity-2024. Accessed: 2024-11-20. Licence: CC BY 4.0.

[13] Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, and William Lintner, "United States Data Center Energy Usage Report," Lawrence Berkeley National Laboratory (LBNL), June 2016. https://www.osti.gov/biblio/1372902. Accessed: 2024-11-20.

[14] Ben Kepes, "30% of Servers Are Sitting Comatose, According to Research," *Forbes*, 2015. https://www.forbes.com/sites/benkepes/2015/06/03/30-of-servers-are-sitting-comatose-according-to-research/. Accessed: 2024-11-20.

[15] Wei Deng, Fang-Ming Liu, Hai Jin, and Dan Li, "Leveraging Renewable Energy in Cloud Computing Datacenters: State of the Art and Future Research," *Chinese Journal of Computers*, vol. 36, pp. 582–598, 2014. https://doi.org/10.3724/SP.J.1016.2013.00582.

[16] M. Marelli *et al.*, *Overhead Transmission Lines, Gas Insulated Lines and Underground Cables*. https://www.cigre.org/userfiles/files/Publications/Reference_papers/REF%20PAPER%20-%20OHL-UGC-GIL_COReview_2019-10-01.pdf. Accessed: 2024-11-20.

[17] Utility Dive, "Why 100% Renewables Isn't Feasible by 2050," https://www.utilitydive.com/news/why-100-renewables-isnt-feasible-by-2050/560918/. Accessed: 2024-11-20.

[18] Inspire Clean Energy, "Why Don't We Use More Renewable Energy?", 2024. https://www.inspirecleanenergy.com/blog/clean-energy-101/why-dont-we-use-more-renewable-energy. Accessed: 2024-11-20.

[19] Intergovernmental Panel on Climate Change (IPCC), "Global Carbon and Other Biogeochemical Cycles and Feedbacks," in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, pp. 673–816, Cambridge, 2023.

[20] Myles ALLEN, David FRAME, Chris HUNTINGFORD, Chris JONES, Jason LOWE, Malte MEINSHAUSEN, Nicolai MEINSHAUSEN, "Warming Caused by Cumulative Carbon Emissions Towards the Trillionth Tonne," *Nature*, vol. 458, pp. 1163–1166, May 2009. https://doi.org/10.1038/nature08019.

[21] M. W. JONES, G. P. PETERS, T. GASSER, *et al.*, "National Contributions to Climate Change Due to Historical Emissions of Carbon Dioxide, Methane, and Nitrous Oxide Since 1850," *Scientific Data*, vol. 10, pp. 155, 2023. https://doi.org/10.1038/s41597-023-02041-1.

[22] Arman SHEHABI, Sarah SMITH, Dale SARTOR, Richard BROWN, Magnus HERRLIN, Jonathan KOOMEY, Eric MASANET, Nathaniel HORNER, Inês AZEVEDO, and William LINTNER, "United States Data Center Energy Usage Report," Lawrence Berkeley National Laboratory (LBNL), June 2016. https://www.osti.gov/biblio/1372902. Accessed: 2024-11-20.

[23] NOAA Climate.gov, "Climate Change: Global Temperature," 2024. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=Earth's%20temperature%20has%20risen%20by,2023%20global%20summary. Accessed: 2024-11-20.

[24] SRG Research, "Hyperscale Data Centers Hit the Thousand Mark: Total Capacity is Doubling Every Four Years," 2024. https://www.srgresearch.com/articles/hyperscale-data-centers-hit-the-thousand-mark-total-capacity-is-doubling-every-four-years. Accessed: 2024-11-20.

[25] Taylor G. SMITH *et al.*, "pmdarima: ARIMA Estimators for Python," 2017. http://www.alkaline-ml.com/pmdarima. Accessed: 2024-11-20.

[26] Robert NAU, "ARIMA Models for Time Series Forecasting," 2024. https://people.duke.edu/~rnau/411arim.htm#:~:text=A%20nonseasonal%20ARIMA%20model%20is,errors%20in%20the%20prediction%20equation.. Accessed: 2024-11-20.

[27] S. J. MILLER, "Autoregressive Models for Time Series Analysis," 2024. https://sjmiller8182.github.io/LearningTS/build/autoregressive_models.html. Accessed: 2024-11-20.

[28] SolarPower Europe, "Cumulative Installed Solar PV Capacity Worldwide from 2000 to 2023 (in Megawatts) [Graph]," June 2024. https://www.statista.com/statistics/280220/global-cumulative-installed-solar-pv-capacity/. In Statista. Retrieved November 19, 2024.

[29] GWEC, "Cumulative Installed Wind Power Capacity Worldwide from 2001 to 2023 (in Gigawatts) [Graph]," April 2024. https://www.statista.com/statistics/268363/installed-wind-power-capacity-worldwide/. In Statista. Retrieved November 19, 2024.

[30] Ember, "Electricity Production Volume from Natural Gas Worldwide from 2000 to 2023 (in Terawatt-Hours) [Graph]," July 2024. https://www.statista.com/statistics/1303724/global-gas-power-generation/. In Statista. Retrieved November 19, 2024.

[31] International Hydropower Association, "Cumulative Hydropower and Pumped Storage Installed Capacity Worldwide from 2014 to 2023 (in Gigawatts) [Graph]," August 2024. https://www.statista.com/statistics/1179170/global-hydropower-capacity/. In Statista. Retrieved November 19, 2024.

[32] International Atomic Energy Agency (IAEA), "Generation of Electricity from Nuclear Energy Worldwide from 2010 to 2023 (in Terawatt-Hours) [Graph]," October 2024. https://www.statista.com/statistics/1349363/global-nuclear-electricity-production/. In Statista. Retrieved November 19, 2024.

[33] Ember, "Electricity Production from Coal Worldwide from 2000 to 2023 (in Terawatt Hours) [Graph]," July 2024. https://www.statista.com/statistics/1082201/coal-fired-electricity-generation-globally/. In Statista. Retrieved November 19, 2024.

[34] Ember, "Electricity Production from Fossil Fuels Worldwide from 2000 to 2023 (in Terawatt Hours) [Graph]," July 2024. https://www.statista.com/statistics/1303807/global-fossil-fuel-power-generation/. In Statista. Retrieved November 19, 2024.

[35] Yingbo ZHANG, Hangxin LI, and Shengwei WANG, "The Global Energy Impact of Raising the Space Temperature for High-Temperature Data Centers," *Cell Reports Physical Science*, vol. 4, no. 10, pp. 101624, 2023. https://www.sciencedirect.com/science/article/pii/S2666386423004447. Keywords: data center, free cooling, energy saving, global impact, server development.

[36] ENERGY STAR, "How to Balance Ambient Data Center Setpoints with IT Equipment Energy Use," 2024. https://www.energystar.gov/products/ask-the-experts/how-balance-ambient-data-center-setpoints-it-equipment-energy-use. Accessed: 2024-11-20.

[37] ASHRAE, "ASHRAE TC 9.9 Power Trends and Cooling Applications White Paper," June 2016. https://www.ashrae.org/file%20library/technical%20resources/bookstore/ashrae_tc0909_power_white_paper_22_june_2016_revised.pdf. Accessed: 2024-11-20.