Problem Chosen	2024 HiMCM Summary Sheet	Team Control Number
B		15517

High-Energy, High-Impact: The Environmental Toll of HPC

Ever since stepping into the 21st century, High-Performance Computing (HPC) has become prevalent in response to technological advancements and society's increasing need for computational demands. On the whole, HPC is indispensable since it allows efficient processing of enormous, diversified datasets that would be impossible to be processed by standard computers.

However, the growing HPC usage results in higher energy consumption from the energy mix to power these systems, bringing up adverse environmental impacts such as carbon emissions contributing to global warming. We developed a concise model to quantify the actual energy consumption using important parameters like the system's full capacity, average utilization rates, and Power Usage Effectiveness (PUE). We estimated that the global energy consumption for HPC would skyrocket to 1,179TWh in 2030.

When pivoting to focus on the environmental consequences of energy consumption, we designed a time-dependent sinusoidal function to depict the cyclic daily power consumption pattern of every server, and we input it in a definite integral calculus formula to scale up and convert the power consumption to the annual global HPC carbon emissions. We specified how different energies–fossil fuels, renewable energy, and nuclear energy–are produced and contribute to carbon emissions within the HPC hardware's life cycles. Our model made reasonable and reliable predictions that, if HPC systems continue to follow the current energy mix proportions in 2030, the HPC carbon emissions will be 6.36 billion tons and cost a surprising 989 billion USD.

We expanded our model by speculating how increasing the proportion of renewable energy will lead to carbon emission reductions in the future. We mapped out a logistic growth model to show this by employing restriction factors of government policies and limited HPC resources. We used the least square method to demonstrate that, by switching to 100% renewable energy, the carbon emissions in 2030 will be much lower 2.06 billion tons. Additionally, we highlighted several challenges of using more renewable energy to reflect the duality of implementing it.

A further stretch is our air pollution model which converts each inorganic pollutant's concentration into the Air Quality Index (AQI) and connects to the logistic model by describing the severity of air pollution based on the volume of polluted air.

Eventually, we advised numerous policy-oriented and technological recommendations to mitigate the harmful environmental effects, such as having higher investments in HPC for technological innovations and applying energy-efficient HPC hardware. We also connected them back to the core concepts of utilization rates and peak and off-peak power consumption.

Keywords: Power usage effectiveness (PUE), actual energy consumption, energy mix proportions, logistic growth, least square method

Contents

Letter to the United Nations Advisory Board
1 Introduction 4
1.1 Background Knowledge 4
1.2 Scope of the Problem 4
2 Variables and Assumptions 5
2.1 Variable Declaration 5
2.2 Assumptions and Justifications 6
3 Annual Energy Consumption of HPC 7
3.1 IT Equipment Energy Consumption 7
3.2 Power Usage Effectiveness 8
3.3 Actual Energy Consumption 8
4 Modeling Carbon Emissions Caused by HPC 10
4.1 Dynamic HPC Power Consumption Over Time 10
4.2 Dynamic Energy Consumption Over Time 10
4.3 Total Worldwide HPC Carbon Emissions Every Year 11
4.4 Economic Loss Due to Carbon Emissions 12
4.5 Analyzing the Carbon Emissions of Fossils 13
4.6 Analyzing the Carbon Emissions of Renewable Resources 13
5 Model Application 14
5.1 How Our Model May Change with the Growth of HPC 14
5.1.1 Possible Variable Variations 14
5.1.2 Exploring Changes in the Model 15
5.2 Predictions Up to the Year 2030 15
6 Modifying Our Model 16
6.1 Shifting to Renewable Energy 16
6.2 HPC Air Pollution Model 19
7 Initiatives to Reduce Environmental Impacts 20
7.1 Technical Solutions 20
7.2 Policy-Oriented Solutions 20
7.2 Initiative Incorporated in Our Model 21
8 Strengths and Weaknesses 21
8.1 Annual Energy Consumption of HPC 21
8.2 Annual HPC Carbon Emissions 21
8.3 Logistic Model on Carbon Emission Reductions 22
8.4 HPC Air Pollution Model 22
9 Conclusion 23
References 24

Letter to the United Nations Advisory Board

Team 15517
November 19, 2024
To: The United Nations Advisory Board
Subject: Request to include HPC environmental impacts more in the development goals for 2030

Dear Members of the United Nations Advisory Board,

We, Team 15517, are writing a letter to express our concerns surrounding the inclusion of high-powered computing systems in your report published in September, *Governing AI for Humanity*.

We fully acknowledge that HPC has always played an irreplaceable role in artificial intelligence, where it enables the training of accurate and reliable AI models using diverse, large datasets. This is an essential capability pushing forward global AI governance that promotes our socioeconomic development. However, there is a duality in HPC power usage. While your report encompasses insightful and valuable explanations of the needs and gaps of global AI governance and how to enhance global cooperation, we believe the report underemphasizes the HPC environmental consequences regarding the energy mix. We feel that potential environmental consequences associated with the rapid growth of HPC systems are insufficiently addressed.

To better understand the adverse environmental impacts of HPC, our team has developed three comprehensive, credible mathematic models that estimate the annual worldwide HPC energy consumption, carbon emissions, and payment for emissions from the years 2023 to 2030. Our integral model predicts that, if HPC systems rely on the current energy mix of roughly 80% being fossil fuels, they will cause approximately 6.6 billion tons of carbon emissions in 2030. The financial costs of these emissions will be 989 billion dollars. The cost is not only substantial but also triggers severe climate changes such as global warming. Thus, we urge you to consider the environmental consequences in your future planning for AI's sustainable development.

Additionally, our logistic model estimates that increasing the renewable energy to power HPC will significantly reduce carbon emissions in 2030. We also proposed effective initiatives to mitigate the environmental impacts including the development of energy-efficient HPC hardware and optimizing computational processes to reduce unnecessary energy consumption.

Given that your scheduled goals for 2030 involve ensuring sustainable development in AI and across the environment, we respectfully request that you include a more detailed section in the report including the consequences and practical solutions to environmental impacts of HPC. Let us keep in mind these consequences alongside expanding AI technology worldwide to foster a sustainable, green future for all of us.

Thank you all for your consideration on this vital issue.

Sincerely,
Team 15517

1 Introduction

1.1 Background Knowledge

Throughout the past two decades, the significance of HPC, the acronym for high-performance computing, has rapidly increased and become an indispensable part of our lives. HPC is commonly used in technological areas like machine learning in AI, weather forecasting in climate modeling, health care, and genomics. In simple terms, it is a supercomputing system that solves problems that are too complicated or would take too long for standard computers. Overall, it enables the simulation or analysis of enormous data volumes with high diversity that would not otherwise be possible with standard computers [1].

An HPC system is like a group of computers being together as a cluster. Each computer in the cluster is a node with an operating system, processor, storage, and network capabilities that allow computers to communicate with each other. Parallel computing is used such that the problem is divided into many smaller tasks and given to the nodes, which work simultaneously on these tasks and calculate for specific results that will eventually be combined. Besides, a cooling system is needed to manage the heat produced by the hardware and minimize the chance of overheating during intensive computations, requiring abundant water usage [2].

Importantly, the HPC system's efficiency and powerful computing power inevitably lead to its immense energy consumption. The system will be supplied with energy by power plants and different sources in the energy mix, namely fossil fuels, renewable energy resources, and nuclear energy resources. The supply and demand for HPC systems and the availability and usage of different energy sources fluctuate over time, varying the energy consumption and thus the environmental impacts. If non-renewable resources like fossils are mainly used to power HPC systems, carbon emissions will be high due to combusting coal, natural gas, and oil to generate electric energy [3]. The emitted greenhouse gases trap the sun's heat and lead to climate issues—global warming and climate change [4]. Other environmental concerns include e-waste pollution, resource depletion, air pollution, noise pollution, land use effects on ecosystems, chemical contaminants of cooling systems, and substantial water usage that must be addressed to save planet Earth.

1.2 Scope of the Problem

The scope of analysis encompasses the annual global HPC energy consumption. We will consider the system's full capacity and average utilization rates when calculating the actual energy consumption of HPC systems.

An HPC system's full capacity refers to the maximum energy output the power plants or energy sources can produce under optimal conditions. This indicates the system's capability to generate energy for computing power when under full load. However, the system will not always be under full load since people's need for it fluctuates in different periods of the year. There are times like holidays, weekends, and nighttime when people relax, spend less time on screens, and require HPC less. On the other hand, there are times like work days, daytime, and days when large computing projects are launched, where corporations need substantial HPC to process their data. For example, developing large natural language processing systems like OpenAI's ChatGPT requires HPC systems with powerful GPUs to train the deep learning model on large, diversified language datasets across many computing nodes.

Hence, we must consider the system's average utilization rate to obtain a realistic estimation of the actual energy output. The rate gauges how effectively a power plant or energy source is being used over a specified

period to compare how consistently the system is operating compared to its full potential. We can simply use the average utilization rate across the year and the full capacity to calculate the actual energy consumption of the system in a year.

After calculating the actual energy consumption, we will combine it with environmental impact factors (like carbon emission factors) and even modify our model to assess the environmental impacts, namely carbon emissions, air pollution, noise pollution, water usage, and pollution caused by producing HPC hardware. We can also have a brief overview of the economic costs of releasing carbon dioxide gas.

2 Variables and Assumptions

2.1 Variable Declaration

Table 1: Symbols and Descriptions

Symbols	Descriptions	Units
$E_{servers}$	IT equipment energy consumption	TWh
T_{usage}	Hours of operation	Hour
P_{server}	Power consumption per server	TW
$N_{servers}$	Number of servers	
E_{max}	System's maximum energy consumption	TWh
PUE	Power usage effectiveness	
U_{avg}	Average actual utilization rate	%
E_{actual}	HPC system’s actual energy consumption	TWh
U_1	Peak utilization rate	%
U_2	Valley utilization rate	%
P_{base}	Baseline power consumption	TW
P_{peak}	Maximum additional power consumption during peak times	TW
T	Hour period of the fluctuation	Hour
E_{source_i}	Average server energy consumption from energy source i	TWh
P_{max_i}	Maximum power consumption of a server from source i	TW
C_{HPC}	Total HPC carbon emissions in a year worldwide	Billion Ton
$C_{emission\ factor_i}$	Carbon emission factor of a particular source i	Billion Ton/TWh
E_{loss}	Economic loss	Billion USD
C_{price}	Carbon price	USD
E_t	Total but limited amount of Earth’s resources available for HPC	
E_r	How much renewable energy takes up in the energy mix for HPC	
E_f	How much fossil fuels take up in the energy mix for HPC	
E_n	How much nuclear energy takes up in the energy mix for HPC	
V_f	Volume of emission of fossil fuels by producing one unit energy	kg CO_2
V_{PM}	Volume of emission of PM by producing one unit energy	mg PM
L_M	Sound pressure level of the machine	dB
$E_{machine}$	Energy produced by machine	TWh
λ	Factor related to how people restrict the growth of HPC by policies	

H	Scale of HPC	acre^{-1}
V_{ai}	Volume of air polluted by pollutant a_i	Billion Ton
$I(a_i)$	AQI of air pollutant a_i	
I_t	Sum of all AQI	
s_m	AQI of the AQI interval's upper bound	
s_{m-1}	AQI of the AQI interval's lower bound	
c_{actual}	Actual concentration of the pollutant	
c_m	Concentration of the concentration interval's upper bound	
c_{m-1}	Concentration of the concentration interval's lower bound	

2.2 Assumptions and Justifications

1. **Annual energy consumption of HPC**
   - Assumption: All the servers in a data center have the same power consumption in TW each day.
     Justification: In reality, different data centers serve different purposes and have different capacities. They contain diversified server configurations including hardware, software workloads, and usage patterns. By assuming a uniform power consumption amount for the servers, it simplifies the problem for us to devise a generalized and scalable formula that can be applied in different data centers without the need to account for every specific variation in server equipment or usage.
2. **Carbon emissions caused by HPC**
   - Assumption: An HPC system's power consumption strictly follows a repetitive cycle that fluctuates from peak to valley level consumption over every daily period.
     Justification: We assume that people's HPC power demand has a cyclic nature across time cycles to provide a macroscopic, generalized representation of the HPC system's periodic behavior. We want to capture the gest—the natural, recurring patterns of daily HPC usage, and present it without the need for complex data values. A cyclic representation also depicts the smooth transitions between high (daytime) and low (nighttime) power demand, which is more realistic than abrupt changes.
   - Assumption: We assume that each data center of the same energy source has the same number of servers.
     Justification: We do this to simplify the data collection processes and calculations. There are data centers with significantly more servers than others, but it is impossible to find all the data for the specific number of servers for each data center worldwide. Hence, we have to make a generalized assumption to balance out their discrepancies and represent an average number of servers per data center.
   - Assumption: The emission factors remain the same as time passes for a few years.
     Justification: Emission factors delineate the rate at which a given activity releases greenhouse gases into our atmosphere. It is extremely difficult to find adequate data for the discrepancies in the composition of a particular energy source and technological development for efficient energy in a few years. So, the change in emission factors could be negligible for our model that targets a general carbon emission.
   - Assumption: The hours of operation for servers remain constant as time changes.
     Justification: When time passes for a few years, the servers possess not too different computational capabilities and are made of similar materials. They are operated for around the same time each

day and roughly need the same time to cool down their internal temperatures.

3. **Logistic model on Carbon Emission Reductions**
   - Assumption: We classify the pollution by HPC into four types—air pollution, water pollution, noise pollution, and pollution caused by producing HPC hardware.
     Justification: Among all environmental impacts brought by HPC, these four are the most concerning factors people worry about. For simplicity and due to the inadequacy of viable resources, our model will mainly focus these types of pollution.
   - Assumption: The total energy demand for the HPC is constant in a short period of time.
     Justification: As energy is a necessity for every server and they need operate every day to maintain a sustainable infrastructure, the discrepancies between different days can be negligible.
   - Assumption: We assume that the energy consumption of HPC systems powered by renewable energy is 0 and we neglect the energy consumption caused by manufacturing their required hardware for renewable energy.
     Justification: Renewable energy does not deprive Earth's total resources in the long term because mother nature can continuously provide or quickly restore renewable resources. The small amounts of energy consumed to create the infrastructures are negligible and very difficult to find data for.
4. **HPC Air Pollution Model**
   - Assumption: We assume that air pollution only consists of inorganic compound pollutants.
     Justification: HPC systems produce organic compound pollutants that pollute the environment, which is much more complicated than modeling inorganic compound pollutants. Also, there are no decent criteria that can both measure the inorganic and organic compound pollutants' impact on air pollution. To simplify the model, we only consider the inorganic compound pollutants for air pollution.

3 Annual Energy Consumption of HPC

3.1 IT Equipment Energy Consumption

Following the scope of analysis, we will begin our model by quantifying the annual global HPC energy consumption. We will first calculate the energy consumed to operate the IT equipment, $E_{servers}$, of HPC. These types of equipment include major system components such as CPU or GPU processors, equipment for memory, hard drives, and cooling systems.

$$E_{servers} = P_{server} \times T_{usage} \times N_{servers}$$

The hours of operation, T_{usage}, will be a constant value since the daily operation duration of an HPC system during workdays will be the same. Also, we assume that the servers in data centers have the same power consumption, P_{server}, to obtain a general, consistent view of electrical power usage. The number of servers, $N_{servers}$, is likely to increase over time due to companies' increasing demands for HPC systems to process large amounts of data. The product of these three parameters provides the IT equipment's energy consumption in TWh and can be easily scaled to calculate the consumption on a worldwide scale.

3.2 Power Usage Effectiveness

The HPC system's full capacity is viewed as its maximum energy consumption in TWh in our model. We divide it by the calculated IT equipment's energy consumption to attain the power usage effectiveness, which determines the energy efficiency of a data center.

$$PUE = \frac{Total\ facility\ power}{IT\ equipment\ energy}$$

$$PUE = \frac{E_{max}}{E_{servers}}$$

The PUE value always ranges from 1.1 to 3, depending on the power usage efficiency of the data center. Now we may use the parameters above to calculate the actual HPC yearly energy consumption in the world. IT equipment's energy consumption and the max energy consumption will augment as more HPC components are manufactured and purchased to meet large corporations' increasing needs for data processing, so the PUE is approximately a constant value.

3.3 Actual Energy Consumption

We must first analyze the system's average utilization rate in a year. We considered two significantly different utilization rates in different periods of a year: U_1 and U_2. Specifically, U_1 represents the peak utilization rate when the system is under max load during the daytime of workdays from Monday to Friday, especially when large computing projects are launched. U_2 indicates the valley utilization rate when the system is almost idle, which might be during nighttime, maintenance downtime, weekends, and public holidays. We used weights to adjust the significance of these rates and summed them to obtain the average utilization rate within a year. The rate is a fraction between 0 and 1, inclusive.

$$U_{avg} = w_1U_1 + w_2U_2 \text{ where } 0 \leq U_{avg} \leq 1$$

The average utilization rate determines the actual energy consumption because it measures how effectively the system is being used now compared to its full potential.

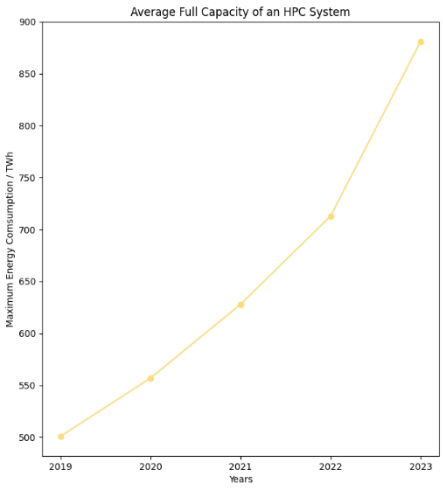
$$E_{actual} = E_{max} \times U_{avg}$$

Thus, we can figure out the actual energy consumption of a data center by multiplying the average utilization rate and energy consumption at full capacity.

Since $E_{max} = PUE \times E_{servers}$, E_{actual} can be rewritten as:

$$E_{actual} = P_{server} \times T_{usage} \times N_{servers} \times PUE \times U_{avg}$$

Using this formula for a global scale and researching several past years' data of these variables, we calculated and graphed the annual average full capacity, annual average utilization rate, and annual energy consumption of HPC systems worldwide from 2019 to 2023.



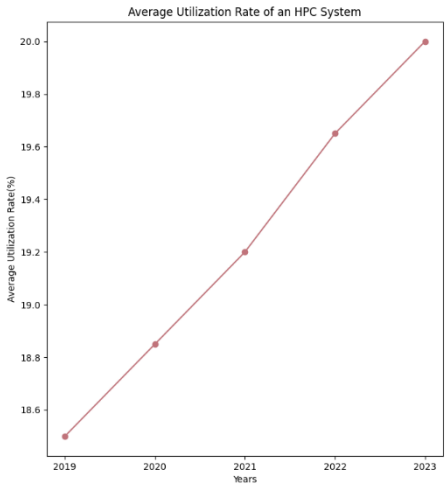


Figure 1: Average HPC Full Capacity **Figure 2: Average HPC Utilization Rate**

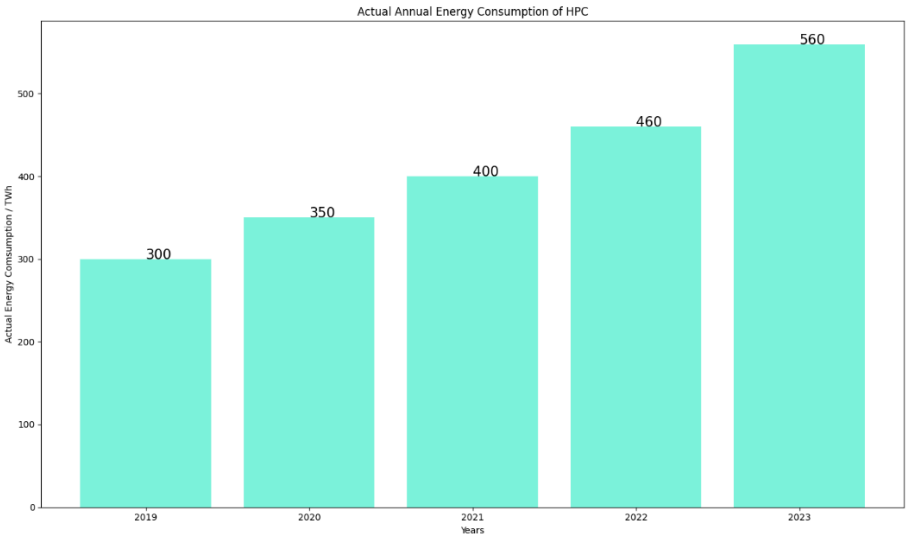


Figure 3: Annual Actual Energy Consumption

In Figure 1 and 2, both the average HPC full capacity and their utilization rates have increased in the past five years due to people's increasing demand for HPC. The full capacity increased faster because overusing the servers might cause them to overheat or result in diminished returns after pushing them to their extremes.

Figure 3 displays a positive, gradually increasing actual HPC energy consumption worldwide for each year from 2019 to 2023. To be specific, in 2019 there is only around 300TWh energy consumed to operate HPC systems, compared to a far higher level in 2023 with approximately 560TWh, increasing nearly by two folds. The increase from 2019 to 2020 is 50TWh with 2020 having 350TWh consumption, which is only half as much as the increase from 2022's 460TWh to 560TWh in 2023. This effectively supports the phenomenon that "global energy consumption growth accelerated in 2023 (+2.2%), much faster than its average 2010-2019 growth rate (+1.5%/year)" [5]. Our calculations are quite accurate compared to the real energy consumption in the past years. For example, the annual electricity report from the International Energy Agency (IEA) says data centers consumed 460TWh in 2022 [6], and our calculation for 2022 is 452TWh.

4 Modeling Carbon Emissions Caused by HPC

4.1 Dynamic HPC Power Consumption Over Time

We can incorporate definite integrals into the model for calculating the carbon emissions constantly produced in a year by HPC systems worldwide. Calculus can model the dynamic power consumption of HPC systems, considering the system's usage and the continuously varying energy sources over time. We can model temporal variations and continuous power consumption over time and account for how the energy mix changes in different periods.

We assume that an HPC system's power consumption follows a repetitive cycle that fluctuates between peak (full-load system) and valley (idle system) level consumptions over each day period. Hence, we will first develop a time-dependent function, $P(t)$, to describe the power consumption in TW at any time t. Later, the total energy consumption over a given time period can be calculated.

$$P(t) = P_{peak} \times \sin\left(\frac{2\pi t}{T}\right) + P_{base}$$

The inspiration for this time-dependent function is the sinusoidal function written in the form $y = \mathrm{a} \times \sin(bx) + k$. P_{peak} represents the amplitude to indicate the maximum additional power consumed during peak times when the HPC system is under the heaviest load while P_{base} is the shift that demonstrates the baseline power consumption when the system operates at an average, standard load. Meanwhile, $\sin\left(\frac{2\pi t}{T}\right)$ represents the period of the cyclic variations in HPC power consumption, with T being the time period over which the cycle repeats.

The maximum power demand occurs when $P(t)$ reaches $P_{base} + P_{peak}$, producing the highest possible daily power consumption. When the $P(t)$ reaches $P_{base} - P_{peak}$, the power consumption reaches its lowest point. Each period starts from the base level consumption, increases to the maximum, decreases to the minimum, and eventually returns to the baseline (just like data centers preparing for a new day of operation). We ensure that the power consumption will always be a positive value above the x-axis and fluctuates the same way for each period.

T will be 24 hours for daily fluctuations in our case when calculating the total HPC carbon emissions, then $P(t)$ can be used to model the fluctuations in power consumption within a day. In reality, this would demonstrate the differences in consumption between daytime when HPC companies all operate and nighttime when there is a less intensive human-driven workload. It is more likely for daytime to have higher demands for computing and cooling due to external factors, business operations, and human interactions with HPC systems. This cycle rises in the daytime and falls at night, repeating every day like a period in sine functions.

4.2 Dynamic Energy Consumption Over Time

Since the energy mix contains different types of energy resources, we can calculate the approximate average energy consumed by a server from a particular energy source i over a duration t:

$$E_{source_i} = P_{max_i} \times t \times U_{avg_i}$$

P_{max_i} gives us the peak power that the system could consume when it runs at full capacity. When multiplied by the time t and the average utilization rate of that server U_{avg_i} in the day, it gives us the general

energy consumption of a server over a time period. Since we want to reflect that the power consumption fluctuates sinusoidally within a day due to the varying power demand, we could modify E_{source_i}'s formula into:

$$E_{source_i} = \int_{t_1}^{t_2} (P_i(t) \times U_{avg_i}) dt$$

On a broad view, the energy mix might change over time depending on the availability of nonrenewable and renewable resources or the current power demand in the grid. Therefore, the total, continuous energy consumed by a server from source i over the period $[t_1, t_2]$ can be calculated more accurately and comprehensively using this formula. For us, $t_1 = 0$ and $t_2 = 24$.

4.3 Total Worldwide HPC Carbon Emissions Every Year

Now we must convert the average energy consumption of a server into its resulting carbon dioxide emissions, hence we will multiply E_{source_i} with its corresponding carbon emission factor, $C_{emission\ factor_i}$. Carbon emission factors are important because they are values between 0 and 1 representing the amount of carbon dioxide emitted per unit of energy consumption. There are three main energy sources—fossil fuels, renewable energy sources, and nuclear energy sources, so we must carefully consider each of their emission factors.

Factors of fossil fuels like coal, natural gas, and oil are high and close to 1. Burning them causes lots of chemical reactions between the carbon in fossils and oxygen in the air to bond covalently to form CO_2, thus relatively higher carbon emissions produced by fossil fuel energy resources than renewable or nuclear energy resources. We will elaborate more on fossil fuels' carbon emission calculations in section 4.5.

On the other hand, factors for renewable energy sources like wind power, hydropower, and solar power are much lower and close to 0 because they generate electricity from nature. The only energy consumption comes from the manufacturing, construction, operation, and decommissioning of infrastructures that obtain and utilize renewable energy sources. This will be further analyzed in 4.6.

Aside from that, nuclear energy's carbon emission factor is basically 0 because it is a zero-emission clean source. Nuclear energy power plants can generate electric power through fission by splitting uranium atoms to create substantial amounts of energy. The heat released by fission is used to create steam that spins a turbine to generate electricity without producing the harmful byproducts emitted by fossil fuels [7].

Since we only modeled the average energy consumption of a server from a particular energy source, we need to scale it up and pivot our attention to the total annual global HPC carbon emissions on Earth. We will convert a single server's energy consumption into its resulting carbon emissions by multiplying it with the carbon emission factor. n indicates the total number of servers in our formula below, so we will sum up all servers' carbon emissions in one day in a data center and multiply it by the total number of data centers, k. Now we know the total HPC carbon emissions in the world produced in a day. Then, we will multiply this product by 365 to calculate the total carbon emissions of HPC in a year, C_{HPC}. Putting all the information together, we came up with this formula:

$$C_{HPC} = 365 \cdot k \int_{t_1}^{t_2} \sum_{i=1}^{n} (P_i(t) \times U_{avg_i} \times C_{emission\ factor_i}) dt$$

To include the complexity of the energy mix, we can rewrite C_{HPC} in a detailed, refined manner for data collection and future calculations:

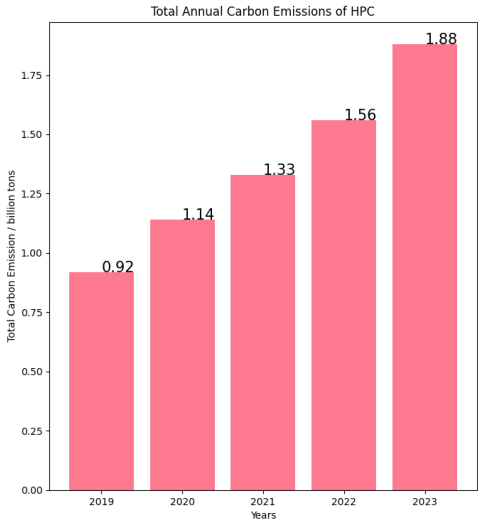
$$C_{HPC} = 365 \cdot (k_1 \int_{t_1}^{t_2} \sum_{i=1}^{n_1} (P_{fossil}(t) \times U_{avg_{fossil}} \times C_{emission\ factor_{fossil}}) dt$$

$$+ k_2 \int_{t_1}^{t_2} \sum_{i=1}^{n_2} (P_{renewable}(t) \times U_{avg_{renewable}} \times C_{emission\ factor_{renewable}}) dt$$

$$+ k_3 \int_{t_1}^{t_2} \sum_{i=1}^{n_3} (P_{nuclear}(t) \times U_{avg_{nuclear}} \times C_{emission\ factor_{nuclear}}) dt)$$

k_1, k_2, k_3 represent the different numbers of data centers for the three energies in the energy mix, and n_1, n_2, n_3 represent the different number of servers in a data center from the three energies. To simplify the data collection process and calculations, we assume that each data center of the same energy source has the same number of servers.

After collecting past years' useful data and running the codes, we graphed the yearly total HPC carbon emissions in the world from 2019 to 2023 and resulting annual economic loss in USD:



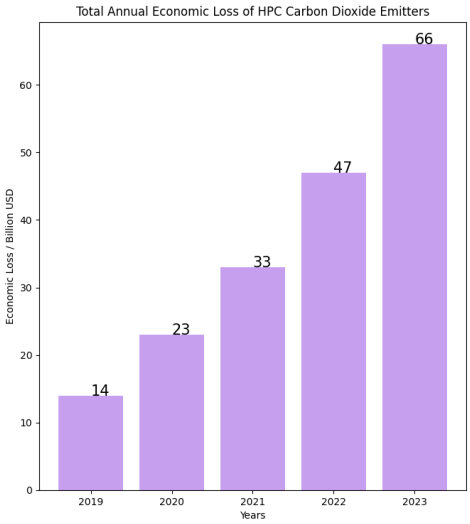


Figure 4: Annual Carbon Emissions **Figure 5: Annual Economic Loss**

Figure 4 depicts a similar trend to Figure 1 because they both increase more rapidly as the years passed, reflecting the consistency of our model. More energy consumed for HPC; more carbon dioxide emitted. Carbon emissions increased almost by two folds from 2019 to 2023, highly alike with the increment from 2019 to 2023 for energy consumptions. Our calculated worldwide carbon emissions for 2023 are around 1.88 billion tons, a lot higher than the 0.92 billion tons in 2019. Our calculations are quite accurate for this too. In 2023 the total carbon emissions worldwide were 37.55 billion tons and HPC took up around 5% of it [8], hence the actual HPC carbon emissions were around 1.8775 billion tons and resemble our 1.88 billion tons.

Figure 5 shows the annual economic loss that will be further discussed in the section below.

4.4 Economic Loss Due to Carbon Emissions

Companies are carbon dioxide gas emitters if they use nonrenewable energies or renewable energies to power their HPC, hence they have to pay the carbon price for their accumulated emissions. We modeled the

annual economic loss of all companies within a year, the price they have to pay in USD, by multiplying their total carbon emissions with 365 and the carbon price in USD:

$$E_{loss} = C_{HPC} \times C_{price} \times 365$$

We then graphed the worldwide economic loss in USD from 2019 to 2023 in Figure 5. The payment seems to increase much more drastically than energy consumption and carbon emissions. Not only do the carbon emissions increase, but the carbon prices worldwide also increase every year considering the current increment, accelerating the increase of economic loss. We estimated that in 2019 the payment would cost 14.0 billion USD and in 2023 it would be 66.0 billion USD, marking an increase of more than four folds.

4.5 Analyzing the Carbon Emissions of Fossils

Fossil fuels mainly include oil, coal, and natural gas, and we found data for them. Now we would like to share our detailing about the total carbon emissions caused by burning fossil fuels in a year. The structure of the formula is the same as the formulas for C_{HPC}:

$$C_{fossil} = 365 \cdot (k_4 \int_{t_1}^{t_2} \sum_{i=1}^{n_4} (P_{oil}(t) \times U_{avg_{oil}} \times C_{emission\ factor_{oil}})dt + k_5 \int_{t_1}^{t_2} \sum_{i=1}^{n_5} (P_{coal}(t) \times U_{avg_{coal}} \times C_{emission\ factor_{coal}})dt$$

$$+ k_6 \int_{t_1}^{t_2} \sum_{i=1}^{n_6} (P_{natural\ gas}(t) \times U_{avg_{natural\ gas}} \times C_{emission\ factor_{natural\ gas}})dt)$$

This formula is used to calculate the C_{fossil} that takes up most of the C_{HPC}. k_4, k_5, k_6 represent the different numbers of data centers for the three fossil fuels and n_1, n_2, n_3 represent the different number of servers in a data center from the three fuels. To simplify the data collection and detailed calculations, we also assume that each data center using the same type of fossil fuel has the same number of servers.

4.6 Analyzing the Carbon Emissions of Renewable Resources

As mentioned in 4.3, data centers using renewable energy sources also cause a bit of carbon emissions because there will be greenhouse gases emitted when manufacturing, constructing, operating, and decommissioning the facilities and infrastructure for these resources. Manufacturing involves building basic components to capture renewable energy, such as installing solar panels and wind turbines. Constructing includes transporting and installing these infrastructures. Operation refers to their daily usage. Lastly, decommissioning contains the recycling and disposal of these infrastructures at the end of their life cycles.

We will dissect how renewable energy is obtained and will showcase a way that calculates the energy consumed by renewable resources for a complete life cycle of renewable energy projects. The energy consumption of renewable resources, $E_{renewable}$, is equivalent to:

$$E_{renewable} = E_{manufacturing} + E_{construction} + E_{operation} + E_{decommissioning}$$

Since the infrastructures are manufactured, constructed, operated, and decommissioned at different times than t in our previous formulas, we will use τ when describing their life cycles. We will define $C(\tau)$ as the infrastructures' carbon emission rate.

$$E_{source_i} = \int_{\tau_a}^{\tau_b} C_{source_i}(\tau)d\tau$$

We will point out the carbon emission rate during manufacturing and construction periods:

$$C_{manufacturing}(\tau) = E_{materials}(\tau) \times C_{emission\ factor_{material\ unit}}$$

We can convert the amount of materials used over a time period, $E_{materials}$ into carbon emission rate by multiplying it with its carbon emission factor per material unit.

$$C_{construction}(\tau) = f_{machinery}(\tau) + f_{transportation}(\tau)$$

$f_{machinery}(\tau)$ depends on the machinery operating over the period τ and $f_{transportation}(\tau)$ depends on the transportation distances and the particular fuel type used by trucks and ships over the period τ.

Hence, $E_{renewable}$ can be rewritten as:

$$E_{renewable} = \int_{\tau_1}^{\tau_2} C_{manufacturing}(\tau)d\tau + \int_{\tau_2}^{\tau_3} C_{construction}(\tau)d\tau + \int_{\tau_3}^{\tau_4} C_{operation}(\tau)d\tau + \int_{\tau_4}^{\tau_5} C_{decommissioning}(\tau)d\tau$$

Since the life cycle happens step by step over a time period, we noted that $[\tau_1, \tau_2]$ indicates the period for the first step–manufacturing, $[\tau_2, \tau_3]$ represents the period for the second step–construction, and so on. The resulting carbon emissions can be simply calculated by multiplying the energy consumed by the general carbon emission factor of renewable energy:

$$C_{renewable} = E_{renewable} \times C_{emission\ factor_{renewable}}$$

We can define specific units for the formulas and scale them up to a full-year level and calculate for annual renewable energy's carbon emissions. However, it will be much more complicated to find data for specific steps in the life cycle.

5 Model Application

5.1 How Our Model May Change with the Growth of HPC

5.1.1 Possible Variable Variations

With the increasing demand for high-volume e-commerce transactions and data storage and management, the number of servers will soar subsequently to provide sufficient advanced computations to establish a sustainable HPC environment. Hence, P_{max_i} and $N_{servers}$ are expected to be much greater than the values illustrated in 2023. Noteworthily, researchers from the IDC expect that the global data center electricity consumption will double between 2023 and 2028 with a five-year CAGR of 19.5% and reach 857 Terawatt hours in 2028 [9].

Some alterations will also occur in renewable energy resources. As people are more concerned about the environmental issues, a considerable amount of payout will be invested in the renewable energy resources and result in a larger amount of renewable energy produced. Hence, $E_{renewable}$ is expected to be greater than the value calculated in 2023.

Another inevitable consideration is that nuclear energy will also represent an increasing trend on account of environmental issues. It is commonly referred to as a clean energy source that produces nearly zero carbon emissions or other greenhouse gas emissions. Hence, $E_{nuclear}$ is expected to be greater in 2030 too.

On the contrary, provided the fact that fossil fuel energy resources produce a significant amount of carbon emissions that deteriorate our fragile environment and that fossil fuels are non-renewable because we have to wait another million years for new batches, the average carbon price will continue to rise due to their limited amounts. On account of the money, the energy consumed by fossil fuels is expected to decrease. Hence, E_{fossil} is expected to be lower than the value illustrated in 2023.

5.1.2 Exploring Changes in the Model

The model will change when we modify the variables.

Overall, the total energy consumption by HPC will rise directly with the increasing number of servers. As fossil fuel energy source is the largest contributor to greenhouse gas emissions, the decreasing demand for fossil fuels will result in a smaller greenhouse impact. HPC systems will possess a cleaner carbon footprint, relying on clean energy resources more heavily in the future. For instance, Microsoft data centers illustrate methods to decarbonize its data center and construction operations. It made an ambitious commitment: By 2030, it would be "carbon negative"–meaning to take more carbon out of the atmosphere than emitting [10].

We might need to develop a logistic model to describe the total annual carbon emissions because fossil fuels are likely to be less used throughout the years to operate HPC and more renewable energy and nuclear energy sources will be used. Our estimations in 5.2 show that the carbon emissions will increase very rapidly in several future years to 2030, but we should demonstrate that the slope will decrease, and the line will approach a plateau once the energy mix transforms to focus mainly on using renewable energy and nuclear energy. A limited number of nonrenewable resources and government policies that favor cleaner energy sources would be the restrictions of the logistic growth. We will elaborate more in "6. Expanding the Model".

5.2 Predictions Up to the Year 2030

We will use the changed model to estimate the total HPC energy consumption, carbon emissions, and economic loss up to the year 2030. We estimated the data for the variables in our models and used the logistic regression of the least square method in Python to predict. We used the logistic line of best fit because the three estimations are not in linear growth, they are supposed to grow rapidly in the first few years up to 2030 but will eventually decrease in growth rate in the far future when the energy mix changes a lot. We chose the logistic line because it also mimics the trends depicted in precedent graphs for these three calculations.

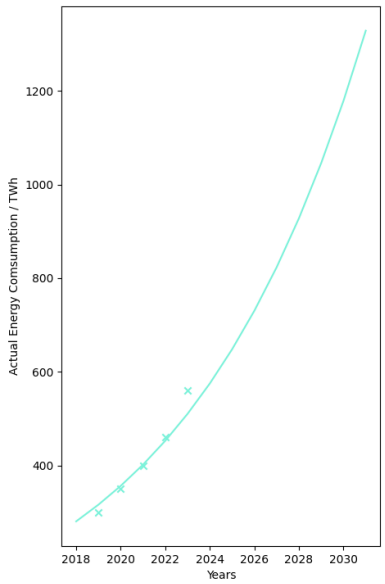


Figure 6: Energy Consumption

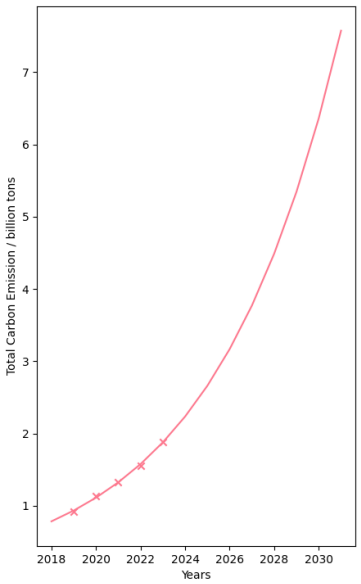


Figure 7: Carbon Emissions

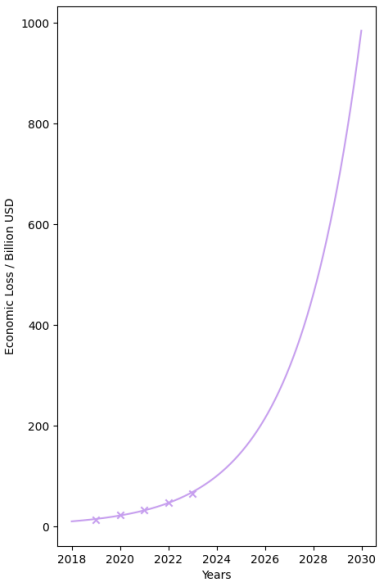


Figure 8: Economic Loss

In Figure 6, we estimated that, in 2030 the actual HPC energy consumption worldwide will increase to approximately 1179 TWh.

In Figure 7, the HPC carbon emissions worldwide will increase to approximately 6.36 billion tons in 2030.

In Figure 8, the economic loss due to carbon price payment worldwide will skyrocket to approximately 989 billion dollars in 2030 if the carbon price and carbon emissions increase at their current rates.

We made line charts to indicate the continuous trends of increase for the three estimations. Overall, the scope of the problem up to the year 2030 involves the magnitude, uncertainty, and environmental concerns of our calculated results. The three trends are in the early stages of a logistic model where the slope increases more rapidly year after year, obviously not a positive linear magnitude. Variables like HPC full capacity and average utilization rate would increase in future years due to people's increasing need for supercomputing to process data. Currently, our model estimates based on the current trends of data, uncertainties include not knowing whether major social events, natural disasters, or economic crisis factors like COVID-19 will happen from now to 2030 because they could drastically impede the growth in energy consumption, carbon emissions, and carbon prices and make our predictions inaccurate. Government policies and the availability of resources in the energy mix in the upcoming years might also make the actual trends vary from our estimations. However, based on our estimations, environmental problems such as global warming and climate change seem to be much more severe in the future due to the HPC carbon emissions increasing by several folds. Therefore, we will propose numerous recommendations to reduce the HPC environmental impacts in section 7.

6 Modifying Our Model

6.1 Shifting to Renewable Energy

We will include four types of pollution caused by HPC systems, namely water pollution, air pollution, noise pollution, and pollution caused by producing HPC hardware to discuss the effects of transferring the energy mix to mainly containing renewable energy sources. It is common sense that if severe pollution damages the local environment, the government authorities will implement policies to regulate it immediately. Thus, government policies on restricting pollutants will be a limiting factor in the growth of carbon emissions in our logistic model, which is mentioned in section 5.1.2.

We assume that the total pollution of HPC only includes the four we mentioned above:

$$P_t = P_a + P_w + P_n + P_p$$

We ought to mention that P_a, P_w, P_n, and P_p are all functions about E_{actual}. P_a, P_w, P_n, and P_p have a positive relationship with E_{actual} when the technological development stays the same within a short amount of time.

To modify our carbon emissions model into a logistic growth, we will start by defining f_x as the fraction of the actual pollution level, P_{actual}, over the maximum pollution level, P_{max}.This is because the units for different kinds of pollution are different, hence we will use fractions to indicate them in an uniform format. Therefore, for a particular pollution x:

$$f_x = \frac{P_{x,actual}}{P_{x,max}}$$

We then define f_t as the total pollution of f_a, f_w, f_n, and f_p:

$$f_t = f_a + f_w + f_n + f_p$$

We will use f_x to calcualte λ, the restriction factor related to how the government restricts the growth of the HPC by enacting energy-limiting policies based on the severeness of the pollution. λ is a function about f_a, f_w, f_n, and f_p.

$$\lambda = F(f_a, f_w, f_n, f_p)$$

To elucidate this, the partial derivatives of the pollution factors will be negative due to the logistic growth:

$$\frac{\partial \lambda}{\partial f_a} < 0, \quad \frac{\partial \lambda}{\partial f_w} < 0, \quad \frac{\partial \lambda}{\partial f_n} < 0, \quad \frac{\partial \lambda}{\partial f_p} < 0$$

We can simply take $\lambda = 1 - \frac{h_a f_a + h_w f_w + h_n f_n + h_p f_p}{f_t}$ where $h_a + h_w + h_n + h_p = 1$ and $0 < h_x < 1$ for since we consider only four kinds of pollution.

We will define E_t as a constant value indicating the total but limited amount of resources on planet Earth that can be used to power all HPC systems. Therefore, in terms of the actual energy consumption of HPC systems, we have the following relationship:

$$E_r + E_f + E_n = 1 \text{ and } 0 \le E_r, E_f, E_n \le 1$$

E_r, E_f, and E_n are percentages indicating how much each of the three types of energies (renewable, fossil, and nuclear) takes up in the total energy mix used to power HPC systems. They are all functions of t and change relatively slowly over time. If time passes for a few years, the constitutions of the energy proportion for HPC will unlikely experience drastic changes, so they can be represented by continuous functions where the independent variable is t.

Going back to the main topic, renewable energy's portion in the energy mix will increase in the future because of its sustainability and accessibility. Most importantly, for this model, we interpret the actual energy consumption as the sum of how much all three types of energies deprive of the total amount of resources available on Earth that can be used for powering HPC, E_t. We can infer that the consumption is contributed mostly by burning fossil fuels and powering nuclear power plants, which causes massive carbon emissions. Therefore, the government will restrict the usage of nonrenewable resources over time since we will have fewer amounts of them in the future and eventually run out of them.

There is a positive relationship between the HPC scale, H, and the actual energy consumption by HPC. Therefore, to obtain the scale, we have the following formula:

$$E_{actual} = q \cdot H(1 - E_r)$$

H depends on the growth of the HPC scales worldwide and q is a constant. There is a positive relationship between E_H and H will gradually weaken when E_r increases because HPC systems of the same scale use more renewable energy than in the past. There will be a decrease in total energy consumption at the same scale due to renewable energy's efficiency. If $E_r = 1$, all the energy used to power HPC is renewable energy and $E_{actual} = 0$ in this case. Because renewable energy does not further deplete Earth's total resources, and we neglect the low energy consumption caused by manufacturing infrastructures like hardware for renewable energy. In simple words, switching to 100% renewable energy will result massive reduction in carbon emissions and alleviate environmental issues significantly.

However, there are potential challenges to be considered if we switch to a 100% renewable energy source:

- Technical challenge: Renewable resources such as solar and wind energy provide unstable electricity that needs to be implemented with modern techniques such as battery storage and smart grids.

- Payout challenge: A considerable amount of infrastructure is needed to be established to provide renewable energy resources, resulting in massive payout that needs approval by policies and long-term investments.
- Peak load challenge: High power data centers require large-scale and stable power supply, while intermittent power supply from renewable energy sources can lead to insufficient power supply, requiring reliance on fossil fuel power generation for emergency power supply.

To simulate the HPC actual energy consumption, we borrowed and altered the logistic growth model created by Pierre Francois Verhulst in 1840:

$$\frac{dE_{actual}}{dt} = \lambda E_{actual} \times m\left(1 - \frac{E_{actual}}{E_t}\right) = \left[1 - \frac{1}{f_t}\sum h_x f_x\right] \times E_{actual} \times m(1 - \frac{E_{actual}}{E_t})$$

x represents a, w, n, and p. f_x is a function of P_x and P_x is a function of E_{actual}, so f_x is a composite function of E_{actual}. $1 - \frac{E_{actual}}{E_t}$ and $\frac{dE_{actual}}{dt}$ conform a positive relationship, and we introduce the constant m to represent the scale factor. We then used the formulas and carbon emission factors to calculate the resulting carbon emissions in our logistic model. We estimated the logistic growth based on the calculated carbon emissions of year 2023 in Figure 7.

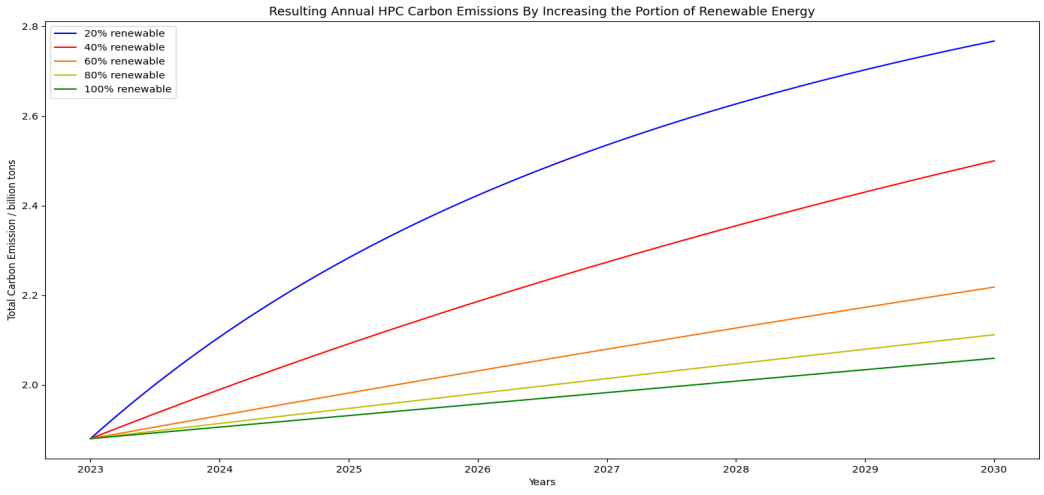


Figure 9: Carbon Emission Reductions

In Figure 9, we used five different colors to indicate the carbon emissions from 2023 to 2030 following a logistic growth if we transfer to 20%, 40%, 60%, 80%, and 100% renewable energy. It is apparent that the carbon emission growth is restricted the most and reaches the logistic plateau the earliest when we have 100% renewable energy in the energy mix, while the growth is restricted the least and reaches the logistic plateau the slowest when we have 20% renewable energy. 20% is approximately the current proportion of renewable energy in the world's energy mix, so the blue curve suggests a reasonable prediction of the resulting carbon emissions if we maintain this proportion. The higher the proportion of renewable energy used in the future, the more possible we can witness carbon emission reductions than the prediction for a 20% proportion.

Switching to use more renewable energy sources to power HPC systems would cause other effects, too. Carbon prices could be more stable because renewable energy resources are not subject to the same dynamic

market fluctuations as fossil fuels. From an HPC company's point of view, they may improve their appeal and reputation to customers, investors, and users who are very environmentally conscious. It also encourages technological advancement in HPC systems because the higher integration of renewable energy stimulates people to create more resilient and distributed energy grids.

6.2 HPC Air Pollution Model

To model another environmental impact caused by HPC systems, we choose to model air pollution because, from our carbon emission models, we realize how severe the environmental crisis is that HPC could cause. We believe that air pollution may provide a more comprehensive, broader view on the air quality because it will involve numerous pollutants rather than solely calculating for carbon dioxide. Hopefully we can properly assess the air quality under HPC's effects and raise awareness for public health purposes and regulatory compliance and encourage authorities to design pollution control strategies.

Firstly, we need to identify several types of pollutants involved in HPC to estimate the air pollution produced by HPC. We will choose sulfur dioxide, nitrogen oxide, PM, and carbon monoxide because they are the major particulate compounds or gases damaging the air quality. We will denote sulfur dioxide as a_1, nitrogen oxides as a_2, PM as a_3, and carbon monoxide as a_4. After that, we will introduce the Air Quality Index (AQI) to provide an indicator that gauges the air pollution severity based on concentration levels of different pollutants.

We will define the Air Quality Index as function I, and the following formula can be used to calculate the index for specific pollutants:

$$I(x) = \frac{c_{x,actual} - c_{x,m-1}}{c_{x,m} - c_{x,m-1}} \times (S_m - S_{m-1}) + S_{m-1}, \qquad c_{x,actual} \in [c_{x,m-1}, c_{x,m}]$$

c_{actual} is the actual concentration of a pollutant in a concentration interval between c_{m-1} (lower bound) and c_m (upper bound), inclusive. To convert c_{actual} into its corresponding AQI, we will first find out how much the part that c_{actual} is higher than c_{m-1} takes up in the interval $[c_{x,m-1}, c_{x,m}]$ by calculating $\frac{c_{x,actual}-c_{x,m-1}}{c_{x,m}-c_{x,m-1}}$. Then, we convert the part that c_{actual} is higher than c_{m-1} into its AQI by multiplying $\frac{c_{x,actual}-c_{x,m-1}}{c_{x,m}-c_{x,m-1}}$ with $S_m - S_{m-1}$, the difference between the lower bound S_{m-1} and upper bound S_m in the AQI interval corresponding to the concentration interval. Finally, we add the product to the S_{m-1} to obtain the AQI of a pollutant with c_{actual} concentration.

We included a table to demonstrate the pollutant intervals and AQI intervals.

Table 2: Concentrations and AQI of Inorganic Air Pollutants

Air Quality Index (AQI)	The regulated concentrations of pollutants (μg/m^3)			
	Sulfur dioxide (SO_2) /24h	Nitrogen dioxide (NO_2) /24h	Carbon monoxide (CO) /24h	PM2.5/24h
0	0	0	0	0
50	150	40	2	35
100	250	80	4	75
150	475	180	14	115
200	800	280	24	150
300	1600	565	36	250

To determine the severity of pollution caused by each pollutant, we will introduce I_t, the sum of all AQI, and define the fraction $f_a = \frac{P_{a,actual}}{P_{a,max}}$ as the severity for each pollutant. We devised a formula for P_a:

$$P_a = k \sum_{i=1}^{4} (\frac{I(a_i)}{I_t} \times V_{a_i})$$

Air pollution P_a conforms a positive relationship with the sum of all $\frac{I(a_i)}{I_t} \times V_{a_i}$, where V_{a_i} indicates the volume of air polluted by pollutant a_i. In addition, we use a constant k in the formula as a scale factor to adjust the impact of each pollutant on the overall pollution level. The actual P_a and maximum P_a can be calculated to obtain the severity fraction f_a for the previous logistic model.

Our model relates to other key areas too, especially energy consumption. Increasing demands in HPC data processing will cause higher energy consumption to power the servers. Air pollution will be significantly more severe if most energy consumed is from burning fossil fuels since countless pollutants will be released into the atmosphere. By analyzing the HPC system's energy consumption, we can potentially reckon its contributions toward air pollution. The air pollution model also connects to HPC's average utilization rates because if we invent energy-efficient hardware, chips, cooling systems, and computing algorithms, energy consumption will be relatively lower in the same data center scale and result in less imminent air pollution conditions. We can investigate the effect of the increasing U_{avg} on the P_a.

7 Initiatives to Reduce Environmental Impacts

7.1 Technical Solutions

Based on particular aspects, one of the most direct means is that data centers can be equipped with energy-efficient hardware including servers, and hard drives to optimize energy efficiency to a maximum extent. To be more specific, equipping with Solid-State Drives (SSD) [11] possess several advantages such as performance expedition, and power savings that can optimize overall energy consumption. Compared to traditional Hard Disk Drives (HDDs), we found that SSDs possess more advanced capabilities such as compactivity that guarantees sustainability. Such replacements can result in a lower overall energy consumption and thus lower carbon emissions, alleviating the environmental conditions.

Another inevitable consideration is to identify software that requires lots of energy to run and applications equipped with inefficient algorithms. Once identified, data centers can apply optimization methods to reduce overall energy consumption. Specifically, we suggest that data centers can optimize database queries to eliminate redundancies and improve data retrieval efficiency. Once applied, it can effectively boost energy efficiency and produce less harmful emissions.

7.2 Policy-Oriented Solutions

The government can also make great contributions to a green planet Earth. Specifically, the government can propose and enact policies and laws to regulate the pollution produced by HPC systems. Not only could carbon emissions be paid, for example, but the government can also raise taxes for producing one unit amount of water pollution, prompting data centers to reduce effluent immediately.

Another viable solution is that the government can allocate more investment to HPC systems to expedite innovative technologies and techniques. This would bring significant benefits because a rise in wages can attract more specialists to work on technical conundrums, working toward a thriving HPC workplace that takes account of environmental protection.

7.2 Initiative Incorporated in Our Model

According to the technical solutions, data centers can be equipped with more energy-efficient servers and hard drives in the future. Such changes may result in a higher average utilization rate and an overall better energy usage efficiency. For example, U_{avg} might increase from around 20% now to 25% in 2030. With the operation of efficient CPUs or GPUs, relatively less energy is required for servers to perform advanced computations, resulting in a considerable drop in total energy consumption. The surge in utilization rate will directly cause the computer throughput to increase. Subsequently, energy restraints will also be reduced. In our model, several circumstances such as P_{server}, PUE, U_1, and U_2 will be reconsidered and recalculated. PUE and P_{server} will be relatively lower and U_1, and U_2 will be higher. Such alterations might cause the dependent variable C_{HPC} to drop significantly.

Moreover, technological improvements in hardware efficiency can also result in a reduction in power consumption spikes. Lots of energy is saved by energy-efficient hardware, making a great contribution to environmental protection. In our model, such modifications can lead to evident reductions in the discrepancies between independent variables like P_{peak} and P_{base}. The dynamic power consumption $P(t)$ will maintain a more stable level while the HPC systems will have near-constant, high utilization rates, U_{avg} over long periods without encountering thermal or power issues. There is a lower need to balance peak and off-peak loads to avoid power spikes as power is produced effectively and unnecessary energy consumption is avoided.

8 Strengths and Weaknesses

8.1 Annual Energy Consumption of HPC

- Strengths: Strengths: By assuming that all servers have the same power consumption each day, we simplified the problem of quantifying the general total energy consumption and made the model scalable to data centers of different sizes. We can easily adapt it to those data centers by changing the variable $N_{servers}$ for different data center sizes. The model involves a straightforward and concise method to calculate actual energy consumption and is easily scaled to a full-year scale, combining key factors like full capacity, average utilization rate, and PUE for comprehensive calculations. The calculated data values are highly similar to the actual ones from 2019 to 2023, proving its accuracy.
- Weaknesses: The assumption ignores individual differences among servers and overlooks their possible variations in usage patterns, energy efficiency, and hardware configurations. The model neglects periodic fluctuations in utilization rates and shifts in energy demand since it takes a weighted average from U_1, and U_2, which may lead to inaccuracies in future estimations.

8.2 Annual HPC Carbon Emissions

- Strengths: We modeled the dynamic power consumption and energy mix using definite integrals, capturing the realistic feature of time-varying HPC energy consumption and energy resources

proportions to provide reasonable and relatively accurate calculations for the past five years and futuristic predictions up to 2030. We account for the emission factors of the energy mix, including all three types of energy—fossil fuels, renewable energy, and nuclear energy, we investigated the major types of fossil fuels' emission factors and corresponding power consumptions, and we even discussed how carbon emissions are unavoidably produced for obtaining renewable energy. We are very attentive to detail and clearly showed how we convert a server's energy consumption into carbon emissions and scale it up to the worldwide level within a year. We thoroughly depicted the HPC system's periodic fluctuation in power demand and thus its cyclic power consumption using a sinusoidal formula.

- Weaknesses: We assumed that carbon emissions do not vary within several years and every server's operating time and sometimes their power consumption are the same. This deviates from reality because technological innovation in the upcoming years may let emission factors evolve and shit the energy mix. In addition, the servers' operation time might vary due to companies' power demand, which might reflect our model's inflexibility and potential inaccuracies in this area.

8.3 Logistic Model on Carbon Emission Reductions

- Strengths: It fully considers the growth of E_{actual} being restricted by the limiting factors of Earth's total resources and government policies on energy usage in the face of pollution caused by HPC growth. It considers that each factor has different weights to fit the actual situation. At the same time, when dealing with the weights of various pollutants, the fraction form of $\frac{P_{actual}}{P_{max}}$ is used to avoid the computability problem caused by the different units of representative parameters selected for each pollution. When dealing with complex functions such as λ, we fully consider the properties that the function should have from a practical perspective and use these properties to determine the simplified form of the function. Our predictions for higher carbon emission reduction due to more renewable energy being used is fairly reasonable.
- Weaknesses: Various variables are interrelated and some implicit functional relationships between variables are omitted for the sake of simplicity, which may result in some deviation from the actual situation. We assumed the energy consumption (resource depletion) of renewable resources to be 0 and ignored the small amounts of energy consumed when manufacturing, constructing, operating, and decommissioning infrastructures for renewable energy. We did not include the renewable energy sources' hardware energy consumption during their life cycles to simplify the problem, hence our predictions on carbon emissions reductions might differ from the reality.

8.4 HPC Air Pollution Model

- Strengths: The air pollution model fully considers the various contributions of air pollution by different pollutants like particulate matters and different oxide gases. It also quantifies the weights of different air pollutants using scientific standards that vividly reflect real-world scenarios.
- Weaknesses: The actual polluted scenario is far more complicated than our model. For instance, we did not consider organic compounds as pollutants because we needed to simplify the model. Therefore, there will be discrepancies between the actual P_a and I_t and our calculations.

9 Conclusion

Our paper thoroughly investigates and communicates mathematically about typical environmental impacts caused by high-powered computing. Our findings inspire and help governmental policymakers mitigate the impacts and technological industries seek sustainable energy methods to power their HPC systems.

First and foremost, we initiated by zooming into and quantifying the annual energy consumption of HPC systems worldwide, inputting vital variables like HPC systems' full capacity, average utilization rate, and power usage effectiveness to calculate highly accurate energy consumptions between 2019 and 2023, such as 460TWh in 2022, which is equal to what happened in reality.

Secondly, we magnified our view on the dynamic patterns of HPC energy consumption by devising a sinusoidal time-dependent function to delineate the cyclic pattern of daily power consumption. We applied it in a definite integral calculus formula that converts continuous energy consumption into carbon emissions. It appropriately scaled from a server's daily power usage to the worldwide HPC data center's carbon emissions in a year. The formulas split the energy mix into three categories—fossil fuels, renewable energy, and nuclear energy, and calculate each of their carbon emissions. Not only HPC processing data will result in increased carbon emissions, but we also pay high attention to the construction process and life cycles of HPC facilities that also contribute to emissions. We predicted viable future values of HPC annual energy consumption, carbon emissions, and payment due to emitting carbon dioxide gas up to the year 2030 based on the current energy mix proportions, carbon emission factors, and utilization rates.

Thirdly, we deeply understand that the energy mix is dynamic with renewable energy becoming dominant, hence we explored the possibility of a logistic model that restricts carbon emissions through government policies and limited resources to power HPC. We compared the carbon emissions up to 2030 with different proportion uptakes of renewable energy in the energy mix. Our model suggests that switching to higher amounts of renewable energy will result in relatively lower carbon emissions, diminishing global warming. The graph displays a logistic trend where carbon emission rates become much more slowly over the years. Switching to 100% renewable energy is estimated to have a low 2.1 billion ton emissions, but we also proposed the challenges of the transition to reflect its complex duality.

Finally, we briefly developed an air pollution model to calculate the AQI based on inorganic compound pollutants and map out air pollution's severity. This connects to the energy consumption model and is a stretch of the carbon emission models, enhancing people's awareness of their health and solving air quality issues through effective regulations.

Understanding the significant environmental impacts of HPC through extensive research, accurate calculations, reliable estimations using multi-faceted approaches, we listed proper means to reduce the impacts, which includes policy-oriented and technical solutions such as having energy-efficient HPC hardware. Once applied, considerable amounts of effluent and greenhouse emissions would be reduced.

Overall, our model provides a robust, distinct framework to evaluate HPC's sustainability, it involves lots of preceding assumptions such as having the same power consumption for every server of the same energy source and not accounting for organic pollutants. Our future work should investigate and incorporate more factors for a comprehensive, detailed analysis that guarantees more accurate estimations.

In conclusion, our paper scrutinizes typical circumstances of environmental impacts for HPC. Particularly, we investigated the conundrum via carbon footprint and air pollution, establishing refined models to obtain credible results to convince institutions and governments to make actionable changes toward a sustainable

future. The authorities can enact precise laws to reduce a significant amount of carbon footprint for HPC, making considerable contributions to the protection of our precious homeland—Earth.

References

[1] “What Is High Performance Computing?” *YouTube*, Google Cloud Tech, 14 Nov. 2020.
[2] “Understanding Direct-to-Chip Cooling in HPC Infrastructure: A Deep Dive into Liquid Cooling.” *Vertiv*, Accessed 16 Nov. 2024.
[3] “Burning of Fossil Fuels.” *Understanding Global Change*, 8 Mar. 2022.
[4] “Causes and Effects of Climate Change.” *United Nations*, United Nations, Accessed 16 Nov. 2024.
[5] “World Energy Consumption Statistics.” *Enerdata*, Accessed 18 Nov. 2024.
[6] Iea. “Executive Summary – Electricity 2024 – Analysis.” *IEA*, Accessed 18 Nov. 2024.
[7] “3 Reasons Why Nuclear Is Clean and Sustainable | Department of Energy.” *Energy.Gov*, Office of Nuclear Energy, 31 Mar. 2021.
[8] “Global CO2 Emissions by Year 1940-2023.” *Statista*, 13 June 2024.
[9] Micheal Shirer, 2024, IDC Report Reveals AI-Driven Growth in Datacenter Energy Consumption, Predicts Surge in Datacenter Facility Spending Amid Rising Electricity Costs
[10] Sally Beatty, 2024, Microsoft builds first datacenters with wood to slash carbon emissions
[11] “Sustainable Data Center Infrastructure: Reducing the Carbon Footprint of Hardware“， *DCS Content Team*, 11 Jan, 2024.