



HAL
open science

Power limits in data centers: what can we expect from improving energy efficiency and refreshing servers?

Pablo Leboulanger, Anne-Cécile Orgerie

► **To cite this version:**

Pablo Leboulanger, Anne-Cécile Orgerie. Power limits in data centers: what can we expect from improving energy efficiency and refreshing servers?. IC2E 2025 - 13th IEEE International Conference on Cloud Engineering, Sep 2025, Rennes, France. pp.1-12. <hal-05215097>

HAL Id: hal-05215097

<https://hal.science/hal-05215097v1>

Submitted on 19 Aug 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Power limits in data centers: what can we expect from improving energy efficiency and refreshing servers?

Pablo Leboulanger, Anne-Cécile Orgerie
Univ. Rennes, Inria, CNRS, IRISA, Rennes, France
Emails: pablo.leboulanger@irisa.fr, anne-cecile.orgerie@irisa.fr

Abstract—Ten years after the drafting of the Paris Agreement, the objective to reduce greenhouse gas emissions by half by 2030 seems difficult to reach. The impact of ICT technologies is growing year after year. Data centers have a large impact in this domain. As a lot of research is focused on improving energy efficiency of data centers, one could ask if this improvement would suffice to reduce greenhouse gas emissions. To study the impact of the energy efficiency on the long term, we propose to model a data center that has a limited power capacity and is regularly refreshed with new servers, more energy-efficient than the previous ones, to cope with its workload. Our results explore various growth rates for the load, and the energy efficiency and show that without reduction of usage of data center resources, the improvements in energy efficiency will not be enough to reach the Paris Agreement’s objective.

Index Terms—Energy efficiency, Cloud computing, Data centers, Sustainability.

I. INTRODUCTION

It has been nearly 10 years since the drafting of the Paris Agreement. A first observation is that the objective to cut by 50% the global greenhouse gas (GHG) emissions by 2030 seems difficult to reach. Indeed, since 2015, the GHG emissions have been constantly rising according to IPCC [1]. In 2019, the global anthropogenic GHG emissions rose at 59 $GtCO_2eq$ [1]. The share of Information and Communications Technology’s (ICT) field produces between 1.8% and 3.9% of global GHG emissions [2]. Data centers are responsible for 18% to 41% of ICT’s GHG emissions according to various dedicated studies [2]. These emissions, as all the others, must be reduced to reach Paris Agreement’s objective.

Hyperscale data centers are growing bigger and bigger and more and more energy hungry [3], [4]. In the first years on exploitation, Cloud providers tend to extend their infrastructures by adding new devices alongside with the existing. The refresh (or renewal) of servers appears in the following years of exploitation, usually by batch, to cope with the growing load and to benefit from more energy efficient servers. Yet, refreshing servers is usually insufficient for cloud providers to cope with the growing users’ demand, and the number of data centers is also rapidly increasing [3].

Cloud computing, that is driving the development of data centers, is a growing market. As an example, according to a consulting company, the market volume of data centers in

France is supposed to grow from 127,000 MW in 2025 to 168,000 MW in 2030 at a +5.70% rate per year [5].

Figure 1 presents the overall electricity consumption of 4 major Cloud companies (Google [6], Microsoft [7], Meta [8], Apple [9]). For instance, Google’s annual electricity consumption has jumped from 12.237 TWh to 25.307 TWh between 2019 and 2023, which means more than doubling.

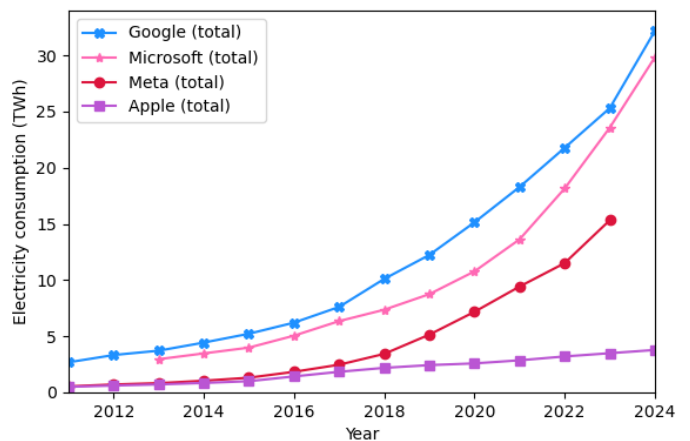


Fig. 1: Yearly electricity consumption of 4 major Cloud companies according to their own annual sustainability reports (Google [6], Microsoft [7], Meta [8], Apple [9])

Facing this growth of the data center demand [5], most of the proposed solutions to stabilize or decrease their GHG emissions are focused on renewable energy sources and energy efficiency. One can be concerned by the capacity of such solutions to achieve this aim for different reasons. Indeed, improving energy efficiency can be subject to rebound effects, and consequently it can lead to an increase rather than a decrease of overall GHG emissions [10]. Moreover, since the 70s, empirical laws (e.g. Moore’s law, Koomey’s law) depict the energy efficiency trend that is always improving at high rates. Yet, these laws show signs of slowing down and without major hardware technology innovations in the forthcoming years, energy efficiency will also slow down [11]. What would be the impacts in terms of GHG emissions with a slowing down energy efficiency and a still increasing demand for data centers?

The environmental impacts are obviously not limited to GHG emissions, but also affect water and material resources. Indeed, on the 43 chemical elements that appear in the production of decarbonized electricity, energy storage and ICT, only two elements do not have criticality issue [12]. This over-exploitation of materials and water is environmentally and societally concerning.

Taking this into consideration, the research community has recently started to question ICT's ability to reduce overall GHG emissions in other sectors [13] and to explore ways to establish limits for ICT with regard to planet boundaries [14]. Following this path, in this paper, we consider the case of a single data center with limited power capabilities, and we study how this data center can evolve through the years. It means that the instantaneous power demand of the facility cannot increase, yet, the load can and servers can be refreshed.

In this work, we propose to look at which kind of load such a data center, limited in power, would be able to satisfy and what would be the impact on its GHG emissions and water consumption. We propose a model to study its evolution over the years and determine under which conditions (e.g. energy efficiency trend) it can satisfy the users' demand.

The rest of this paper is structured as follows. In Section II, we present the state of the art. Section III, we propose a model for a power-limited data center. Then, Section IV presents the results in terms of load and GHG emissions of different evolution scenarios. Section V presents the water consumption of this data center under various scenarios. In Section VI, we highlight different limitations of our approach and Section VII concludes and provides future work.

II. RELATED WORK

A. Life Cycle Assessment

The lifelong impact of ICT's devices is usually represented in four phases: the manufacturing, the transport, the use and the end of life. For example, from [15], the carbon footprint of a PowerEdge R940xa server is around 13,200 kgCO₂e, of which 93.6% is for the use phase, and 5.9% for the manufacturing one. Other parts are actually small (inferior at 0.5%). The importance of the manufacturing part has long been overlooked because of the difficulty of calculating it and the lack of data from industry. The end of life phase represents the impact of waste or recycling of the device. As detailed in [15], the share of the end-of-life phase is usually low in the current life-cycle analyses. The manufacturing, transport and end of life phases are usually called the *embodied* part.

The use phase represents the energy consumption of the device during its life. This share is inherently linked to the electrical mix of the geographical location of the device. For example, a server powered by a heavy carbon electrical mix will have higher GHG emissions than the same server powered with highly nuclear electricity mix. In the last years, Cloud providers have been claiming an increasingly high percentage of carbon-free energy mix. For example, Google claims a 64% carbon-free energy over their data centers and aims for a 24/7 carbon-free in all their infrastructure by 2030 [6]. This

means that the primary source of GHG footprint may become hardware manufacturing [16]. Yet, carbon footprint of the embodied and use phases are different by nature. The GHG footprint of the use phase is mainly due to the emissions of generating electricity. The embodied phase, and in particular the manufacturing, is related to extracting and transporting materials (mostly consuming oil).

B. Energy consumption reduction

In order to respond to the increasing data center demand in the last decade, one major approach consists in focusing on data center efficiency. The energy efficiency is defined as a quantity of work per unit of energy. In the following, energy efficiency is defined as Flops/Watt with Flops (also denoted Flop/s) being the number of floating point operations per second.

The Power Usage Effectiveness (PUE) is defined as the energy consumption of the IT devices over the energy consumed by the entire data center. The global PUE in data centers has been decreasing from 2.5 in 2007 to 1.65 in 2013 [17]. This decrease is mostly due to improvements in the cooling system. However, the global PUE has been quite stable since 2013 until 2024 from 1.65 to 1.56 [17]. In the same time, in large facilities, the PUE has been decreasing to 1.1 [18], almost reaching the ideal value of 1. But, the PUE does not display the growing energy consumption of data centers since it only reflects an energy ratio and not the absolute energy consumption values, which are growing rapidly has shown on Figure 1.

C. Aging Moore's and Koomey's laws

The energy efficiency of processors has been observed since the 70s by Moore. In [19], he observed that the number of component per integrated circuit doubles every two years. This observation remained quite accurate over the years. Focusing on data centers, Koomey, in [20], observes that the energy efficiency doubles every 1.57 years. In [21], the authors show that energy efficiency is still growing exponentially but at lower rate. On the 2008-2023 period, the energy efficiency has doubled every 2.29 years in high-performance computers. They also show that the performance is growing faster than the energy efficiency (doubling every 1.85 years). This induces a power growth in the infrastructures. In the light of this, the authors deduce that it is necessary not only to improve the energy efficiency, but also to reduce the number of instructions executed by a program.

D. Kaya Identity

While a growing power demand is observed on the data centers' side, it is difficult to relate this growth to planetary boundaries and to the Paris Agreement. In [22], Kaya proposes a model, known as the Kaya Identity, to estimate the global GHG emissions with regard to economics tendencies. This model has been criticized since it only sketches coarse grain tendencies and does not fully capture rebound and indirect effects. Yet, it has been adapted to different domains of ICT,

like semiconductor manufacturing, mobile Internet access, data center usage and video streaming [23], [24] to identify the factors that influence GHG emissions' evolution. In [23], the authors adapt the Kaya identity to multiple ICT domains. The adaptation to global data center usage is based on Koomey's *law* and some revisions of it in order to provide an evolution of the energy efficiency. Practically, in this model, carbon emissions are estimated as the product of the number Internet users, the workload per user, the IT energy efficiency, the power usage effectiveness and the carbon intensity of the electricity mix. This work considers the global use of data centers with a baseline scenario (i.e. current trends) without analyzing other possible scenarios and practical implications of the trends (i.e. the refreshment of servers to benefit from energy efficiency improvements). In [24], the author explores the energy efficiency in video streaming, and models the evolution over time of its global GHG emissions. Both studies conclude that energy efficiency improvements will not be sufficient to meet the Paris Agreement. Facing this statement, both work pledge for a reduction in the usage either by reconsidering innovation in companies [23] or through legislative means [24].

E. Data center models

In [25], the authors propose a model to compute the best refreshing rate between two servers. An old server will be refreshed by a newer one more efficient in energy. The cumulative energy is the sum of the embodied energy and the energy consumption of the server over the years. The proposed model computes the payback time when the cumulative energy of the new server will be lower than the old one. The authors demonstrate that the embodied part is not significant in the case of GHG emissions [25]. This is due to the fact that servers are continuously powered on and that the authors do not consider the use of low carbon energy as Cloud providers do.

In this paper, we explore the impact of the energy efficiency in the case of a data center limited in power. First, we propose an analytical data center model over years. Based on [25], we compute when to refresh its servers in order to satisfy a growing load in time. Moreover, based on [23] and [24] adapted to our model, we estimate the tendencies of GHG emissions over time for our power-limited data center.

F. Water Consumption

Alongside of the GHG emissions, another concerning environmental domain is the water footprint of data centers. While the GHG emissions tend to impact climate change at a global scale, the water consumption have more local impacts. Indeed, the impact can be really different regarding the water scarcity in the location region of the data center. For instance, the Water Scarcity Factor (WSF) in France was only of 21.6% in 2021 when it was of 66.5% in India [26] (a higher value meaning less water availability).

The water footprint is usually divided in 3 parts: onsite footprint, offsite footprint and embodied footprint [27]. The

onsite water footprint is mostly the water consumed by the cooling system of the data center. This is usually represented by the Water Usage Effectiveness (WUE). In [28], the authors define WUE as the ratio of the annual water usage over the energy consumption of the IT equipment. WUE depends on the climate of the geographical region where is located the data center. A data center in a hot region may have a bigger WUE than one in a cold region. For instance, OVHCloud claims a mean WUE of 0.37 but with a per data center WUE which can vary from 0.1 to 1.20 [29]. The offsite footprint is the amount of water used in the generation of electricity employed to power the data center. It depends on the local electricity mix. Finally, the embodied water footprint is the amount of water used in the manufacturing and transport of IT equipment, notably servers.

III. METHODOLOGY

In order to model a limited data center, several elements are required, like an estimation of the energy consumption of equipment in the data centers (servers, switches, cooling system), the evolution over time of the energy efficiency of servers and a way to compute of the number of servers in a refresh. In this section, we present step by step the models employed hereafter. Firstly, in Section III-A, we present the linear power consumption model for a server that is inspired by [25] and how to model for the carbon footprint of the server. Then, in Section III-B, we detail the global trends that are considered in the studied scenarios. Section III-C provides the analytical model of the power-limited data center. Lastly, Section III-D presents our adaptation of Kaya identity based on [23] in order to estimate the GHG emissions. Table I summarizes the variables that are introduced in the following.

A. Energy consumption and carbon footprint of a server

First, we introduce the energy model used to compute the energy of servers. This linear model is taken from [25]. Let s be a server, either s_o old or s_n new. Such a server is characterized by the set P_i^s, P_f^s, C^s where,

- P_i^s is the power of s at idle state,
- P_f^s is the power of s at full load state,
- C^s is the capacity of s .

The annual energy consumption of a server s running at the average utilization level can be estimated in kWh:

$$\Psi_u^s = (P_i^s \alpha + P_f^s \beta) \times 8.76 \quad (1)$$

With:

- β is the annual average utilization rate of server,
- α is the annual average idle rate of server ($\alpha = 1 - \beta$),
- 8.76 is the conversion factor from Joules to kWh. (8760 hours a year divided by 1000).

The annual use phase consumption of s , E_u^s , is estimated by the following:

$$E_u^s = \Psi_u^s \times PUE \quad (2)$$

TABLE I: Variables employed in the model of a power-limited data center

| Variable | Definition | Unit |
|--------------------|---|-----------------------|
| x | Server, switch or data center | |
| s | Server | |
| s_o | Old server | |
| s_n | New server | |
| s_m | TOP500 mean server (2014-2024) | |
| S | Set of servers | |
| r | Switch | |
| R | Set of switches | |
| μ | Number of ports per switch | |
| P_f^x | Power at full state of x | W |
| P_i^x | Power at idle state of x | W |
| C^x | Capacity of x | GFlops |
| β | Annual average utilization rate of x | |
| α | Annual average idle rate of x | |
| Ψ_u^x | Annual energy consumption of x | kWh |
| E_e^x | Embodied energy of x | kWh |
| E_u^x | Use phase energy of x | kWh |
| E_c^x | Cumulative energy of x | kWh |
| E | Energy | kWh |
| ω | Load | GFlops |
| λ | Power | W |
| λ^ρ | Scaled power | W |
| A | Power demand | W |
| ε | Power efficiency | GFlops/W |
| ρ_y | Ratio of y | |
| Φ | Amortization function | kgCO ₂ e |
| T^Φ | Amortization duration | |
| S_Φ | Set of servers to amortize | |
| τ | Payback time | |
| γ | GHG factor | kgCO ₂ e/W |
| η_1 | GHG emissions | |
| η_2 | GHG emissions for Kaya adapted model | |
| n | Year of refresh | |
| t | Year of simulation | |
| T | Maximum simulation duration | |
| D | Simulation duration | |
| B | Power budget | W |
| PUE | Power Usage Effectiveness | |
| δ^{s_o} | Number of old servers s_o for a refresh | |
| δ^r | Number of new switches for a refresh | |
| WUE | Water Usage Effectiveness | L/kWh |
| EWIF | Energy Water Intensity Factor | L/kWh |
| WSF | Water Scarcity Factor | |
| Ω | Water footprint | L |
| Ω^{onsite} | Water onsite footprint | L |
| $\Omega^{offsite}$ | Water offsite footprint | L |
| Ω^{emb} | Water embodied footprint | L |

The embodied energy of a server s is the following:

$$E_e^s = \sum_{t=0}^{T^\Phi} \Phi(t, E_e^s) = 1000 \text{ kWh} \quad (3)$$

With :

- Φ is the amortization function, discussed later in Section III-B,
- T^Φ the amortization time,
- The value 1000 kWh is taken from [25]

To estimate the efficiency of a refresh of an old server s_o by a new one s_n , one can look at the payback time of the refresh. In order to compute it, one need the cumulative energy of the servers. The cumulative energy consumption of s_o and s_n , over the years t , is computed with the following:

$$E_c^{s_o}(t) = E_u^{s_o} \times t + \sum_{i=0}^t \Phi(i, E_e^{s_o}) \quad (4)$$

$$E_c^{s_n}(t) = E_u^{s_n} \times t + n \times (E_u^{s_n} - E_u^{s_o}) + E_e^{s_o} + \sum_{i=n}^t \Phi(i-n, E_e^{s_n}) \quad (5)$$

With:

- $t \in \mathbb{N}$,
- $n < t$, the refresh year of the server,

The payback time τ of refreshing s_o with s_n is the time taken by the cumulative energy of s_n to be less than the cumulative energy of s_o :

$$\tau = \min(t \mid E_c^{s_o}(t) > E_c^{s_n}(t)) - n \quad (6)$$

This payback time is sensitive to the difference of power efficiency between the new and the old servers. The larger is the difference, one can expect a smaller payback time (in the case that the new server is more efficient than the old one). Another factor is the embodied cost of the new server, a small embodied footprint would lead to a smaller payback time. However, here, for simplicity's sake, we consider that all servers have the same embodied cost, even for newer and more energy efficient servers. This reflects the efforts made by server manufacturers to decrease their carbon footprint while increasing performance.

To model the evolution of servers' capacity over time, we use the TOP500 data from 2014 to 2024 to determine a mean server s_m . The TOP500 is used as a good estimate of the trends for top-level servers, both in terms of efficiency (i.e. performance) and energy efficiency. Moreover, both of these metrics are regularly measured (every 6 months) on the latest high-performance infrastructures and are publicly available. $P_f^{s_m}$ and C^{s_m} are calculated using the average values of power and respectively capacity of all TOP500 supercomputers [4] for each year divided by the mean number of servers. $P_i^{s_m}$ is set as a proportion of $P_f^{s_m}$, here 30%. This is an optimistic value since the idle power consumption of a server typically ranges between 30% to 60% of its full power consumption [30], [31]. With this, we find $s_m = (61.19 \text{ W}, 203.97 \text{ W}, 530.93 \text{ GFlop/s})$. The capacity of s_m is also optimistic. This is due to the massive use of GPUs in the TOP500 data set. GPUs are more efficient than CPUs in computing matrix operations employed by the benchmark used to rank the TOP500 data set. Consequently, our mean server is optimistic in terms of capabilities and thus represents an upper bound in terms of energy efficiency evolution in the future.

B. Global trends in load, power consumption and energy efficiency

To model the evolution of different characteristics of data center, as load, power or power efficiency, we use various data sources. We use different functions ω to simulate the load evolution of our data center. The first one is using the

compound annual growth rate (CAGR) of data center market of different countries. For example, the data center market in France is predicted to grow from 127 GW in 2025 to 168 GW over 5 years at a 5.70% rate each year according to a consulting firm [5]. The first load function is the function using CAGR.

$$\omega_1(t) = b \times (1 + a)^t \quad (7)$$

In order to study a more theoretical profile, the second function follows a linear evolution of load:

$$\omega_2(t) = a \times t + b \quad (8)$$

Where,

- a is the evolution per year either CAGR or the gradient for respectively ω_1, ω_2
- b is the initial capacity of the data center, C^{DC} , at the starting state.

The evolution of the power of server at full capacity, λ , is modeled using the TOP500 data set [4]. Using the data from 2014 until 2024, shown in Figure 2, we calculate the mean power of HPC's data centers during the last decade and λ represents the linear regression for these values during the considered period.

$$\lambda(t) = 0.91 \times t + 0.12 \quad (9)$$

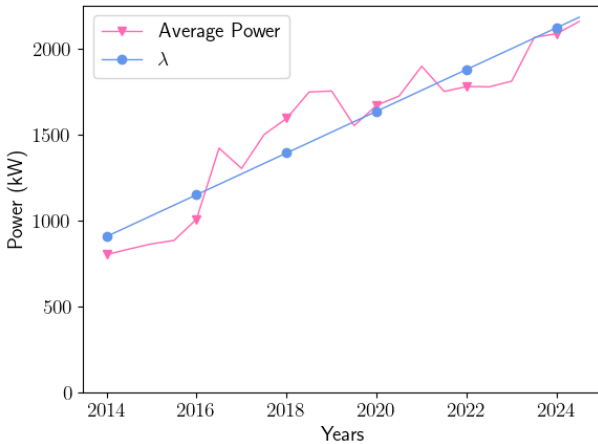


Fig. 2: Average power of the TOP500 computers

Several functions, ε , are used to simulate the different evolutions of energy efficiency. To estimate the evolution of the energy efficiency, shown in Figure 3, we use the same data set as for power (TOP500).

$$\varepsilon = 0.85 \times e^{0.28 \times t} \quad (10)$$

The previous functions are the power and power efficiency of data centers. To suit our model, we need to adapt them to

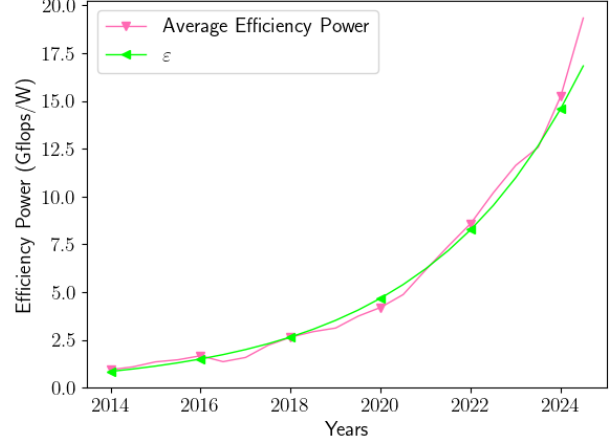


Fig. 3: Average Efficiency Power on the TOP500 computers

a single server. We choose to scale the functions λ and ε with respectively ρ_λ and ρ_ε .

$$\rho_\lambda = \frac{P_f^{s_m}}{\lambda(0)} \quad (11)$$

$$\rho_\varepsilon = \frac{C_f^{s_m} / P_f^{s_m}}{\varepsilon(0)} = \frac{\varepsilon^{s_m}}{\varepsilon(0)} \quad (12)$$

If we apply both ratios to their corresponding function, we get the following scaled function:

$$\lambda^\rho(t) = \rho_\lambda \times (0.91 \times t + 0.12) \quad (13)$$

$$\varepsilon^\rho(t) = \rho_\varepsilon \times 0.85 \times e^{0.28 \times \rho_\varepsilon \times t} \quad (14)$$

Embodied cost of a server can also be seen over time through amortization functions. We use a function Φ of amortization of the embodied cost E_e^s of a server s : Φ pay all the embodied cost at the first year of exploitation of s .

$$\Phi(t, E_e^s) = \begin{cases} E_e^s, & t = 0 \\ 0, & t > 0 \end{cases} \quad (15)$$

C. Power-limited data center

Now that we have detailed how to compute the energy consumption of a server and described the functions related to global trends, we can present our model of limited data center shown on Figure 4. The idea behind such a data center is that, in order to reduce its impact, one can put limits. Here, we explore a power limit that currently holds for actual data centers. Indeed, when building a new data center, Cloud operators contract with an electricity provider and the agreement stipulate the maximum instantaneous power that can be drawn by the data center. Several countries where data centers are abundant (like Ireland) suffer from electrical grid constraints and starts to enforce strict limitation in power supply [32].

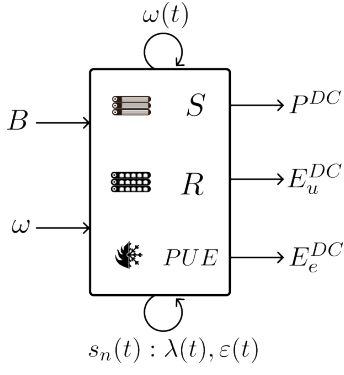


Fig. 4: Limited data center model

A power-limited data center is represented by the following tuple:

$$DC = (B, S, R, PUE, C, \omega) \quad (16)$$

With,

- B , the budget of the data center in Watts,
- S , the set of servers in the data center,
- R , the set of switch in the data center,
- C , the capacity of the data center *i.e.* the total capacity of all its servers (determined by the initial set of servers),
- PUE , the power usage effectiveness of the data center,
- ω , the load of the data center.

The growth of demand on the data center's market gives an impression of the growth of the needed computation power. Indeed, to ensure to satisfy the load demand in every situation, the data center's infrastructure is scaled at the load peak (worst case). From this, we can deduce that the power demand represents the power peak demand.

By definition of PUE, in the worst case (when servers loaded at full capacity), the data center power demand A can be expressed as:

$$PUE = \frac{A(t)}{P_f^R + P_f^S} \quad (17)$$

From our experiments, we have seen that the power of switches (P_f^S) is negligible, under 2% of the total power of the data center. So, we neglect it in the previous equation to have:

$$A'(t) = \frac{A(t)}{PUE} \approx P_f^R \quad (18)$$

Consequently, the load of the data center can be deduced from the projected evolution of power demand A and the efficiency power ε :

$$\omega(t) = A'(t) \times \varepsilon(t) = \frac{A(t)}{PUE} \times \varepsilon \quad (19)$$

In order to determine when to refresh servers, we propose the following algorithm in 4 phases:

- initialization of the data center,
- computing the time of refresh,
- refreshing servers,
- checking payback time.

Algorithm 1 MODEL

```

1: Procedure MODEL( $T, B, s, r, \omega, \lambda^p, \varepsilon^p, \Phi$ )
2: initialize( $B, s, r$ )
3: for  $t \in \{0 \dots T\}$  do
4:    $\omega_n \leftarrow \omega(t)$ 
5:   if  $\omega_n > C^{DC}$  then
6:      $s_n \leftarrow server(t, s, \lambda^p, \varepsilon^p)$ 
7:      $refreshed \leftarrow refresh(s_n)$ 
8:     if  $refreshed$  then
9:        $t' \leftarrow yearNextRefresh()$ 
10:      for  $s_o \in S$  do
11:         $\tau \leftarrow paybackTime(s_o, s_n)$ 
12:        if  $t + \tau \leq t'$  then
13:           $replaceMaximumServer(s_o, s_n)$ 
14:        end if
15:      end for
16:    end if
17:  end if
18: end for
19: End Procedure

```

The parameters are the following :

- T , the maximum model duration
- s , the initial server of the data center
- r , the initial switch of the data center, in the following, we assume there is no evolution of energy efficiency of switches since as explained before, the switches' power consumption is negligible compared to the servers' one
- ω , the load function of the data center (evolving over time)
- λ^p , the function of the full-capacity power consumption of servers
- ε^p is the function of the energy efficiency of servers
- Φ is the amortization function for embodied footprint of servers.

To initialize the data center, we solve an integer programming problem. The number of initial servers in the data center is calculated with the following:

$$\begin{aligned} \#s = \max(x \mid \\ (P_f^S \times x + P_f^R \times |R|) \times PUE \leq B, \\ \wedge x \leq \mu \times |R|) \end{aligned}$$

With

- μ , the number of ports per switch.

In our model, we initialize the load of the data center at the initial capacity to simulate a nearly perfectly dimensioned data center at start. It is important to note that, due to this choice, the model has to make a refresh at the first year of simulation.

Every year of simulation, we need to compute a potential new server following the growing energy efficiency trend. The

new server s_n available at year t is computed using λ^ρ and ε^ρ .

$$\begin{aligned} s_n &= (P_f = \lambda^\rho(t)), \\ P_i &= \rho_i \times \lambda^\rho(t), \\ C &= \varepsilon^\rho(t) \times \lambda^\rho(t) \end{aligned}$$

With ρ_i is the ratio P_i/P_f that is set to 30% as explained in Section III-A.

We describe how to compute the number of new servers needed during a refresh as another linear integer programming problem. The first condition of this problem is that the power at full capacity of the new servers and new switches must be at most equivalent to the power of those replaced (to cope with the power limit). The second condition is that the difference of capacity between the new servers and the replaced ones must be superior to the difference between the new load and the capacity of the data center before the refresh. Otherwise, even with new servers, the data center cannot absorb the increasing load. So, the number of new servers in a refresh is the following:

$$\begin{aligned} \#s_n &= \min(x \mid \\ (P_f^{DC} + P_f^r \times \delta^r + P_f^{s_n} \times x - \sum_{s_o \in S} P_f^{s_o} \times \delta^{s_o}) \times PUE &\leq B, \\ x \times C^{s_n} - \sum_{s_o \in S} \delta^{s_o} \times C^{s_o} &\geq \omega_n - C^{DC}, \\ |S| + x - \sum_{s_o \in S} \delta^{s_o} &\leq (|R| + \delta^r) \times \mu \end{aligned}$$

With

- C^{DC} , the data center capacity before the refresh
- δ^{s_o} , the number of server s_o replaced during the refresh
- δ^r , the number of new switches of the refresh

If the payback time of refreshing a server s_o by s_n is less than the time to the next refresh, it is preferable to replace a maximum number of server s_o .

$$\delta^{s_o} = \max(x \mid (P_f^{DC} + P_f^{s_n} \times x - P_f^{s_o} \times x) \times PUE \leq B) \quad (20)$$

If a refresh is required to cope with the increasing load but is not possible due to power limit, we opt for a strict enforcement of the power limit, meaning that the refreshment is not performed and consequently, the data center cannot cope with the load. Each year, the situation is re-examined and the load follows its initial evolution even if, for one or more years, the data center cannot cope with it.

D. Evaluation of GHG emissions

The GHG emissions of our data center model η are estimated using the electricity's emissions factor.

$$\eta_1 = E \times \gamma \quad (21)$$

Where,

- E is an energy in kWh either E_e , E_u , or E_c
- γ is the electricity's emissions factor in gCO_2e/kWh

Emissions factors are different between countries. This is due to the energy mix that greatly varies around the globe. A country using a lot of high carbon emitting energy sources has a high emissions factor. For example, France has a low carbon energy with $58\text{gCO}_2e/\text{kWh}$ [33]. On the other hand, Germany with more heavy carbonate energy has a higher emissions factor of $381\text{gCO}_2e/\text{kWh}$ [34].

One way to estimate the evolution of GHG emissions over time, is to use the Kaya identity as detailed in Section II-D. We use an adaptation of this model proposed by [23], which is the following:

$$\eta_2(t) = p(t) \times \frac{\omega(t)}{p(t)} \times \frac{E^{IT}(t)}{\omega(t)} \times \frac{E^{DC}(t)}{E^{IT}(t)} \times \frac{\eta(t)}{E^{DC}(t)} \quad (22)$$

With:

- p , the population of Internet users,
- ω/p , the computation affluence,
- E^{DC}/E^{IT} , the IT energy intensity,
- η/E^{DC} , the carbon intensity factor,

To adapt this equation to our model, we set:

- $p(t) \times \omega(t)/p(t)$ to the load of our data center $\omega(t)$
- $E^{IT}(t)/\omega(t)$ to the inverse of energy efficiency of servers converted in GFlops/kWh
- $E^{DC}(t)/E^{IT}(t)$ to the inverse of the PUE
- $\eta(t)/E^{DC}(t)$ to the emission factor γ

$$\eta'_2 = \omega \cdot (\varepsilon^{-1} \cdot 8.76) \cdot (PUE^{-1}) \cdot \gamma \quad (23)$$

In this model, as in [24], we set the improvement of γ at a rate of 1% per year. Furthermore, for each year, we take into account the embodied cost of servers. This cost is amortized linearly over the seven years of servers' warranty. So, the total amount of GHG emissions is the following:

$$\eta_2(t) = \eta'_2(t) + \frac{1}{7} \times \eta_{emb} \times S_\Phi(t) \quad (24)$$

With:

- η_{emb} , the embodied cost of a server [15]
- S_Φ , the number of servers to amortize.

IV. EVALUATION OF THE LOAD'S AND GHG EMISSIONS' EVOLUTION

In this section, we look at multiple scenarios in order to understand the implications of the various studied factors onto the evolution of the GHG emissions.

The baseline scenario, S_0 , is based on the French market demand and emissions factors. Furthermore, the evolution of the energy efficiency follows the trends of the TOP500 data set over the last decade, ε_1 . We also explore the case of Germany S_1 and India S_2 which have different market demand and emissions factors. In S_3 and S_4 , we propose to fix the refresh rates of servers to, respectively, every year and every seven years, that can be considered as the warranty duration of servers. In the last scenario S_5 , we reduce the load, we propose to increase the data center load by its initial capacity every year.

Table II summarizes all the scenarios. GHG emissions factors, γ , are expressed in $\text{gCO}_2\text{e}/\text{kWh}$. The line *Fixed* indicates when the scenario uses a fixed interval of refresh. The line *MW* shows when the load is computed from the power demand of data center market and not from the workload the data center is able to handle.

TABLE II: Scenarios

| | S_0 France | S_1 Germany | S_2 India | S_3 Yearly refresh | S_4 Seven years refresh | S_5 Linear Load |
|--------------|-----------------|------------------|----------------|----------------------------|------------------------------------|-------------------------|
| T | 10 | 10 | 10 | 50 | 50 | 10 |
| ω | 1 | 1 | 1 | 1 | 1 | 2 |
| slope | 5.70% [5] | 6.64% [35] | 15.11% [36] | 5.70% | 5.70% | $\omega(0)$ |
| T^Φ | 7 | 7 | 7 | 1 | 7 | 7 |
| γ | 58 [33] | 381 [34] | 713 [34] | 58 | 58 | 58 |
| <i>MW</i> | ✓ | ✓ | ✓ | ✓ | ✓ | |
| <i>Fixed</i> | | | | ✓ | ✓ | |

A. Different countries: different growing rates for the load and different emissions factors

The French scenario concerns a low growth rate of data center demand (*i.e.* 5.70%) compared to other countries. In Figure 5, the first observation with regard to the capacity of the data center is that, every year, the model is not able to satisfy the load of the data center because of the power limit. This behavior is due to the fact that the load is growing faster than the energy efficiency. The GHG emissions are, excepted for the first year, stable because there is no new servers added to the data center (since refreshing servers cannot cope with the load increase) and the embodied cost is amortized at the first year of exploitation. This is why the electricity consumption is stable along the years. One can see that this constraint of a limited power budget is too strong in order to satisfy the current growth of load.

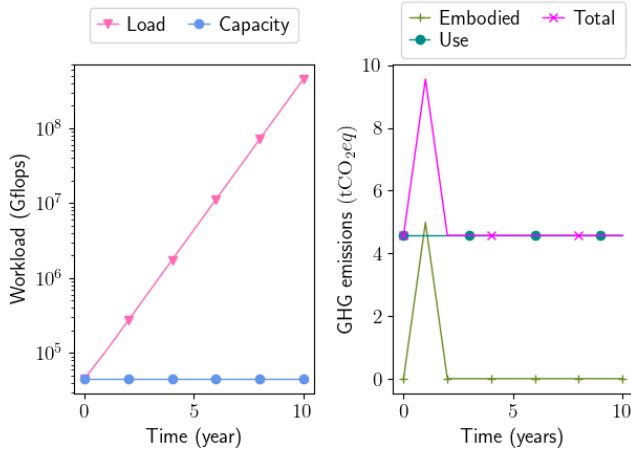


Fig. 5: Workload and GHG emissions in France (S_0)

The same behavior is observed in the case of Germany (S_1) and India (S_2), as shown on Figure 6. The amount of GHG emissions differs by the size of their emissions factor γ . As India’s GHG emissions factor is 10 times bigger than the one of France, the total of GHG emissions is 10 times bigger.

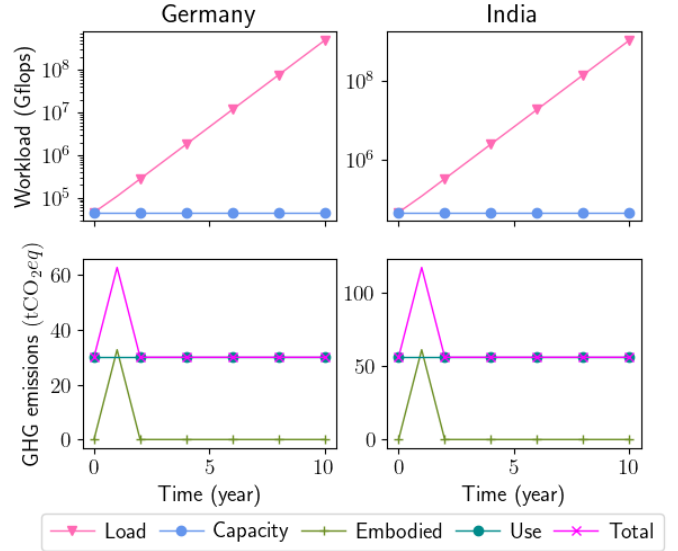


Fig. 6: Workload and GHG emissions in India (S_1) and Germany (S_2)

B. Fixed refresh times

We now look at scenarios where all servers are refreshed after a given number of years. In S_3 , servers are replaced every year in order to fully exploit the energy efficiency improvements (Figure 7). On the contrary, in S_4 servers are replaced every 7 years, which is assumed to be the warranty time of servers (Figure 8).

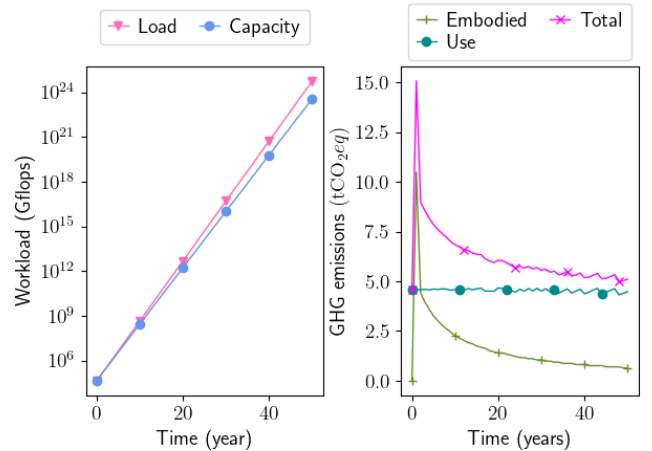


Fig. 7: Workload and GHG emissions with yearly refresh of servers (S_3)

When refreshing servers every year on Figure 7, one can see that even this intensive (and unrealistic) refreshment rate is not enough to satisfy load. Yet, we observe that the total GHG emissions decrease over time because the embodied footprint has a smaller part. Indeed, since the power of new server increases with time, the maximum number of servers that can be installed within the power limit decreases.

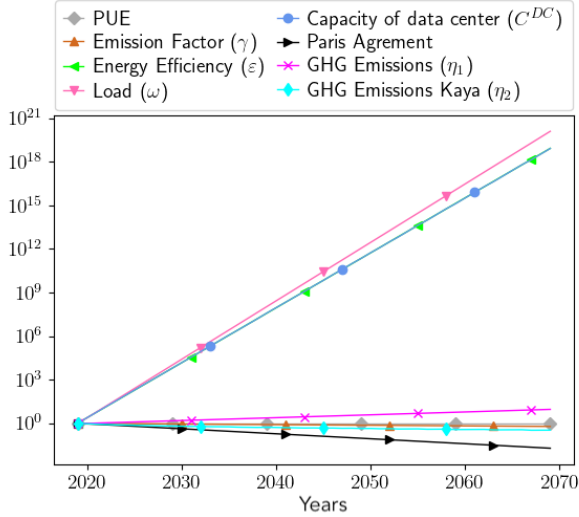


Fig. 8: Evolution of GHG emissions when yearly refreshing servers (S_3)

Figure 8 shows the tendencies of GHG emissions over time with regard to the expectations from the Paris Agreement. In S_3 , as servers are changed every year, one can see that the capacity grows with the same tendency as the energy efficiency. This induces that in Figure 8 both curves are superposed (ϵ and C^{DC}). One can see that the GHG emissions (*i.e.*, η_1) do not decrease. If we limit the load at the capacity, one can see that GHG emissions (η_2) decrease but not fast enough to respect the Paris Agreement.

In order to reduce these emissions, in the scenario S_4 , servers' refresh periods are spaced out every seven years. One can see in Figure 9 that the capacity of the data center grows like a staircase (with a step size of seven years). This staircase curve appears in reflection in the GHG emissions. Even if the overall trend of the GHG emissions does not achieve to reach the Paris Agreement in the long term, the decrease provided by the spacing of refresh periods allows to punctually reach the expected reduction in 2030.

C. Scenarios linear Load

Load is the only factor that remains unmodified in the previous scenarios. So, we now look at a scenario S_5 with a constrained load. Here, the load of the data center arbitrarily rises by the initial capacity of the data center every year. So the load doubles on the first year compared to the initial state, increases by 50% on the second year compared to the first year, and so on. In Figure 10, one can see that until the fifth year of

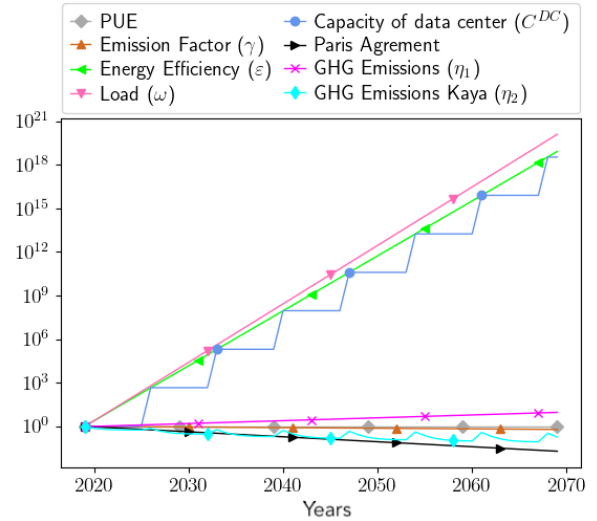


Fig. 9: Evolution of GHG emission when refreshing servers every 7 years (S_4)

modeling, the energy efficiency is not sufficient to satisfy the load demand. However, from the fifth year of modeling the energy efficiency has grown enough to handle the load and only needs some refresh over time in order for the capacity to stay above the load.

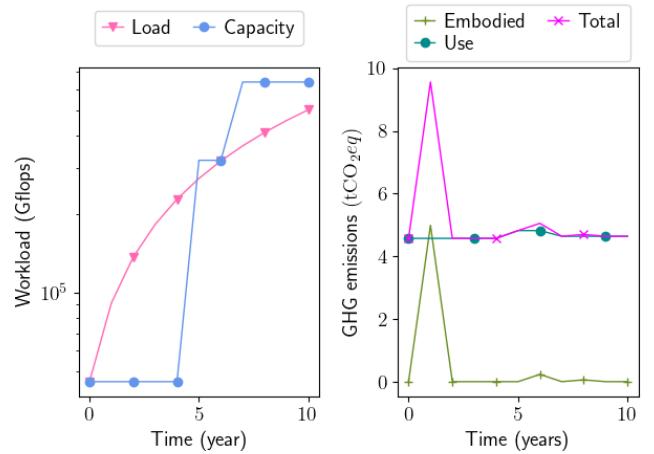


Fig. 10: Workload & GHG emissions with a linear growing load (S_5)

S_5 also successfully achieves, in Figure 11 the reduction of GHG emissions expected by the Paris Agreement. This shows that it may be possible to respond to a growing demand in the order of magnitude of the previous years while fulfilling GHG emissions reduction engagements. However, one may take into considerations that in a small timescale (few years time) the GHG emissions n_2 may still increase due to the growing load and the multiple server refreshments whose embodied cost has to be amortized.

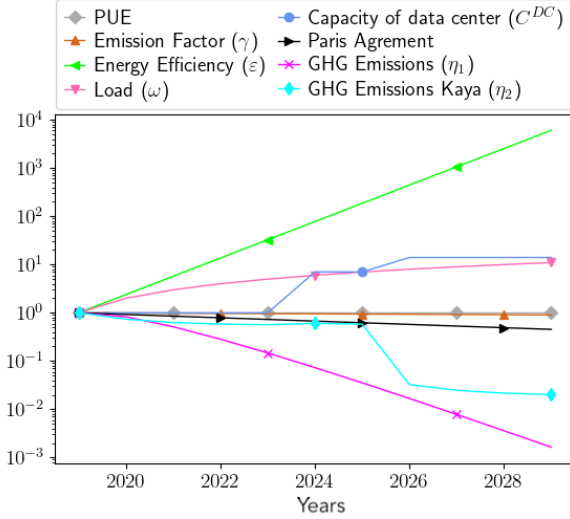


Fig. 11: Evolution of GHG emissions with a linear growing load (S_5)

V. EVALUATION OF THE WATER CONSUMPTION'S EVOLUTION

GHG emissions are currently among the most studied environmental impacts. Yet, it is not the only one to evolve in a worrying way. Water consumption is also concerned by environmental impacts and can be subject to high local constraints depending on data center's location. The water consumption of a data center is usually split in 3 main parts. The first part is the onsite water consumption. This is mainly represented by the consumption of the cooling system. From [37], we can estimate the annual water consumption of the cooling system with the following:

$$\Omega^{onsite}(t) = WUE(t) \times E_u^{DC}(t) \times (1 + WSF^{DC}(t)) \quad (25)$$

With :

- WUE , Water Usage Effectiveness of the data center,
- E_u^{DC} , annual electrical energy consumption the data center,
- WSF^{DC} , Water Scarcity Factor of the geographical region of the data center,

The second part of water consumption is the amount of water used in the electricity generation named offsite water footprint. The offsite water footprint is heavily dependent on the electrical mix of the region of the data center. To estimate the evolution of the electrical mix, we use prospective net-zero carbon mixes in 2050 proposed by the French transmission system operator RTE [38].

Table III shows the different electrical mixes for two possible trajectories in 2050: $m0$ with a fully renewable electricity mix, and $n3$ with a heavily nuclear mix. We use a linear regression to draw the evolution over the years and to reach these mixes in 2050 from the current situation. The Energy Water Intensity Factor (EWIF) represents the water footprint when generating electrical energy and differs from energy

TABLE III: RTE-Futures 2050 Electrical Mix [38] and the average Energy Water Intensity Factor (EWIF) per energy source [37]

| | 2019 Actual | m0 Full Renewable | n3 New Nuclear | EWIF (L/kWh) |
|-------------------|----------------|-------------------------|----------------------|-----------------|
| Hydraulic | 10 | 9 | 10 | 12.64 |
| Nuclear | 72 | 0 | 51 | 2.42 |
| Photovoltaic | 2 | 36 | 13 | 0 |
| Onshore wind | 6 | 21 | 13 | 0 |
| Offshore wind | 0 | 31 | 12 | 0 |
| Bio-energies | 2 | 2 | 2 | 2.10 |
| Ocean energies | 0 | 2 | 0 | - |
| Thermal | 8 | 0 | 0 | 1.24 |

sources [39]. We define EWIF of a certain electrical mix by the weighted mean over the EWIF of its energy sources. EWIF of ocean energies is not set due to lack of data, but ocean energies represent only a small part of the $m0$ electrical mix (*i.e.*, 2%). From [37], we can estimate the annual water consumption from the electricity generation.

$$\Omega^{offsite}(t) = PUE(t) \times E_u^{DC}(t) \times EWIF(t) \times (1 + WSF^{DC}(t)) \quad (26)$$

With :

- PUE , Power usage Effectiveness of the data center,
- E_u^{DC} , annual electrical energy consumption the data center,
- $EWIF$, Energy Water Intensity Factor of certain electrical mix, only the use phase and not manufacturing of the source,
- WSF^{DC} , Water Scarcity Factor of the geographical region of the data center,

The last part of the water consumption of the data center is represented by the embodied cost of the equipment. Here, we consider only the servers' cost. Since the data center is power-limited, the data center building does not evolve over time. Consistent with the previous assumptions, the switches' cost is neglected due to its small overall proportion. The embodied cost of the cooling system is also not considered since the power limit ensures it does not have to evolve over time, and thus represents a constant parameter over time (only increasing when the system is replaced due to aging). We use a fixed water footprint for a server, $\Omega_s^{emb} = 1,797,590.2$ L which represents the mean over 4 different Dell servers [40]. From this, we can estimate the annual water embodied footprint with the following:

$$\Omega^{emb}(t) = \Omega_s^{emb}(t) / T^{\Phi}(t) \times S_{\Phi}(t) \quad (27)$$

With :

- Ω_s^{emb} , the embodied water footprint of a server
- T^{Φ} , the amortization time set to 7 years
- S_{Φ} , the number of servers to amortize

The total water footprint is the sum of onsite, offsite and embodied water parts.

$$\Omega(t) = \Omega^{onsite}(t) + \Omega^{offsite}(t) + \Omega^{emb}(t) \quad (28)$$

Similar to Section IV, we study scenarios with yearly and seven years refresh (S_3 , S_4) and linear load (S_5) with different evolutions of the electrical mix, respectively m0 and n3.

In Figure 12, one can see that the embodied water footprint is largely predominant in the total water footprint. Over all scenarios, the part of the embodied footprint is 99% of the total water footprint. One can see this in result of the linear load scenario where the total water footprint drastically drops when the first batch of servers is amortized (at year 2026).

From the offsite water footprint perspective, one can see that with the m0 mix, the offsite footprint decreases quicker than in with the n3 mix. This difference is due to the nuclear part in the n3 mix. Indeed, the nuclear electricity generation is more water consuming than renewable generation (*e.g.*, wind or photovoltaic).

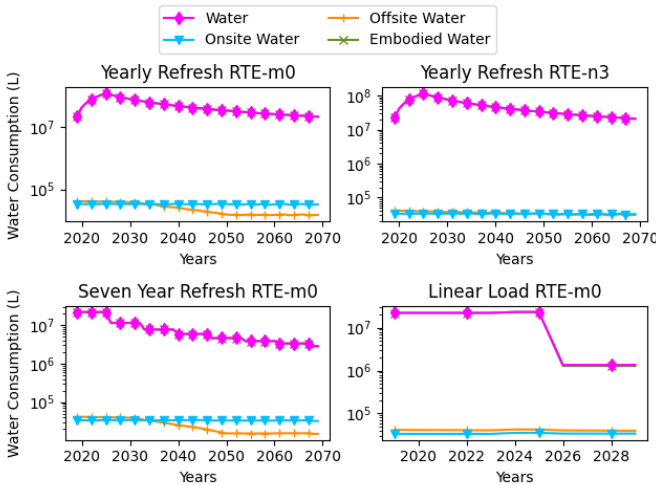


Fig. 12: Water Footprint over different scenarios. Embodied cost (green and cross) is hiding behind the pink line.

The onsite and water consumption are both negligible in front of the embodied cost. This shows that in order to reduce the water footprint, one must reduce the number of manufactured server and thus increase the lifetime of servers.

VI. DISCUSSION

In this section, we discuss some limitations of our model and how they may affect our results.

First, the TOP500 data set is not about Cloud data centers but high performance computers (HPC). The types of workloads are different between Cloud and HPC data centers. Yet, real data describing data centers' evolution are rare and the TOP500 is frequently used in other work for the same purpose (for instance in [21]), to estimate the energy efficiency evolution.

Another limitation with regard to the TOP500 data set is that to compute our base server, we use the capacity of each super computer which is expressed in Flops. Flops represents the typical operations which GPUs are designed for. But our model is simplified to propose servers that are yearly adapted to the energy efficiency improvements without

considering practical hardware constraints (CPU and GPU sizes, heterogeneous servers and workload, etc.). This leads to a clear overestimation of the energy efficiency of our base server (see Section III-A), and consequently of its capacity. Yet, the results have shown that even with this overestimation, no scenario based on current tendencies (S_0 , S_1 and S_2) can achieve to meet the current trend of the growing users' demand within the power limit.

We assume a fixed water and carbon embodied footprint for the servers over the years despite an increasing performance (both in Flops and Flops/W). This representation omits the future scarcity of abiotic materials which is difficult to predict. These depletions may lead to an increase of embodied cost with harder or more energy demanding access to these resources. One could see in this an underestimation of future GHG emissions and water consumption. In particular, this could affect the linear load scenario S_5 . One could ask if this increase of embodied cost would, in the end, prevents this model to achieve the Paris Agreement. This note is utterly valid. However, S_5 seems like a very unrealistic scenario considering the actual trends. On the other hand, this comment may point to the fact that other scenarios, especially scenarios with a fixed refresh time of servers S_3 and S_4 , would even less succeed in the diminution of their GHG emissions.

To conclude, we want to insist that these results, and especially our adaptation of Kaya identity, are to take as trends over time and not as exact values. We do not claim to talk about magnitudes only to prospective future trends of GHG emissions with regard to certain constraints factors that one could apply on data centers.

VII. FUTURE WORK AND CONCLUSION

In this work, we study the impacts of limiting the power of a data center with respect to global trends, such as increasing workload and increasing energy efficiency of servers. Firstly, we have shown that, with such a power constraint, a data center is able to handle the current growing rate of load, but only if servers are refreshed very frequently. Secondly, our work shown that the data center power limitation alone is not sufficient to reverse the actual direction taken by the GHG emissions. This is even more concerning because of the optimistic assumptions on energy efficiency that are taken in this work (conservative Moore's law for instance). As a matter of fact, one could think that the energy efficiency will continue to grow slower compared of the Koomey's original observations as it has been the case in previous years. Considering this, one could see that the most impactful solution would be to reduce the usage of such technologies in order to stay in the safe zone of the planet boundaries.

We also show that the water footprint of highly dependent of the embodied footprint. This strengthens the idea of a reduction of usage in order to limit the manufacturing of new servers.

In the future, we would like to adapt our model in order to consider other possible limits for data center, in particular GHG emissions limits, and to understand what would be the

impacts of such limits over the acceptable workload in these conditions. This work would have to take into account different ways to amortize in time the embodied cost and to study how the amortization options impact the refresh periods. Another direction would be to look at other environmental indicators like materials in order to explore limits on these resources.

ACKNOWLEDGMENT

This work received funding from the France 2030 program, managed by the French National Research Agency under grant agreement No. ANR-23-PECL-0003.

REFERENCES

- [1] IPCC, “AR6 Synthesis Report: Climate Change 2023,” <https://www.ipcc.ch/report/sixth-assessment-report-cycle/>, 2023.
- [2] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. S. Blair, and A. Friday, “The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations,” *Patterns*, vol. 2, no. 9, 2021.
- [3] IEA, “Energy and AI – Analysis,” <https://www.iea.org/reports/energy-and-ai>, 2025.
- [4] TOP500, “November 2024 ranking,” <https://www.top500.org/>, 2024.
- [5] Mordor Intelligence, “France Data Center Market Size,” <https://www.mordorintelligence.com/industry-reports/france-data-center-market>, 2025.
- [6] Google, “2024 Environmental Report,” <https://sustainability.google/reports/google-2024-environmental-report/>, 2024.
- [7] Microsoft, “2024 Environmental Sustainability Report Data Fact Sheet,” <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/2024-Environmental-Sustainability-Report-Data-Fact.pdf>, 2024.
- [8] Meta, “2024 Sustainability Report,” <https://sustainability.atmeta.com/2024-sustainability-report/>, 2024.
- [9] Apple, “Environmental Progress Report,” https://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2025.pdf, 2025.
- [10] C. Gossart, “Rebound Effects and ICT: A Review of the Literature,” in *ICT Innovations for Sustainability*, L. M. Hilty and B. Aebischer, Eds. Springer, 2015, pp. 435–448.
- [11] J.-R. L  quepeys, M. Duranton, S. Bonnetier, S. Catrou, R. Fournel, T. Ernst, L. H  rault, D. Louis, A. Jerraya, A. Valentian, F. Perruchot, T. Signamarcheix, E. Vianello, and C. Reita, “Overcoming the Data Deluge Challenges with Greener Electronics,” in *IEEE European Solid-State Device Research Conference (ESSDERC)*, 2021, pp. 7–14.
- [12] S. Cerf, A. Luxey-Bitri, C. Quinton, R. Rouvoy, T. Simon, and C. Truffert, “Untangling the Critical Minerals Knot: When ICT hits the Energy Transitions.”
- [13] T. Pirson, L. Golard, and D. Bol, “Evaluating the (ir)relevance of IoT solutions with respect to environmental limits based on LCA and backcasting studies,” in *Computing within Limits*, 2023.
- [14] A. Mansoux, B. Howell, D. Barok, and V.-M. Heikkil  , “Permacomputing Aesthetics: Potential and Limits of Constraints in Computational Art, Design and Culture,” in *Computing within Limits*, 2023.
- [15] Dell, “PowerEdge R940xa CarbonFootprint,” <https://corporate.delltechnologies.com/content/dam/digitalassets/active/en/unauth/data-sheets/products/servers/poweredge-r940xa.pdf>, 2021.
- [16] U. Gupta, Y. G. Kim, S. Lee, J. Tse, H.-H. S. Lee, G.-Y. Wei, D. Brooks, and C.-J. Wu, “Chasing Carbon: The Elusive Environmental Footprint of Computing,” in *IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 2021, pp. 854–867.
- [17] A. Isazadeh, D. Ziviani, and D. E. Claridge, “Global Trends, Performance Metrics, and Energy Reduction Measures in Datacom Facilities,” *Renewable and Sustainable Energy Reviews*, vol. 174, p. 113149, 2023.
- [18] Google, “Power Usage Effectiveness - Google Data Centers,” <https://datacenters.google/efficiency>, 2025.
- [19] G. E. Moore, “Progress in digital integrated electronics,” in *Technical Digest. International Electron Devices Meeting, IEEE*, 1975.
- [20] J. Koomey, S. Berard, M. Sanchez, and H. Wong, “Implications of Historical Trends in the Electrical Efficiency of Computing,” *IEEE Annals of the History of Computing*, vol. 33, no. 3, pp. 46–54, 2011.
- [21] A. Prieto, B. Prieto, J. J. Escobar, and T. Lampert, “Evolution of computing energy efficiency: Koomey’s law revisited,” *Cluster Computing*, vol. 28, no. 1, p. 42, 2024.
- [22] Y. Kaya and K. Yokobori, “Environment, energy and economy: Strategies for sustainability,” *United Nations University Press*, 1997.
- [23] D. Bol, T. Pirson, and R. Dekimpe, “Moore’s Law and ICT Innovation in the Anthropocene,” in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021, pp. 19–24.
- [24] T. Maugey, “Towards digital sobriety: Why improving the energy efficiency of video streaming is not enough,” in *IEEE International Workshop on MultiMedia Signal Processing*, 2023, pp. 1–4.
- [25] R. Bashroush, “A Comprehensive Reasoning Framework for Hardware Refresh in Data Centers,” *IEEE Transactions on Sustainable Computing*, vol. 3, no. 4, pp. 209–220, 2018.
- [26] H. Ritchie and M. Roser, “Water Use and Stress,” Our World in Data, <https://ourworldindata.org/water-use-stress>, 2015.
- [27] P. Li, J. Yang, M. A. Islam, and S. Ren, “Making AI Less ‘Thirsty,’” *Communications of the ACM*, vol. 68, no. 7, pp. 54–61, Jun. 2025.
- [28] D. Azevedo, S. C. Belady, and J. Pouchet, “Water usage effectiveness (WUE): A green grid datacenter sustainability metric,” *The Green Grid*, vol. 32, 2011.
- [29] OVHcloud, “Sustainability - Environment,” <https://corporate.ovhcloud.com/fr/sustainability/environment/>, 2024.
- [30] C. Jin, X. Bai, X. Zhang, X. Xu, Y. Tang, and C. Zeng, “A measurement-based power consumption model of a server by considering inlet air temperature,” *Energy*, vol. 261, 2022.
- [31] D. Meisner, B. T. Gold, and T. F. Wenisch, “PowerNap: eliminating server idle power,” *SIGARCH Comput. Archit. News*, vol. 37, no. 1, p. 205–216, 2009.
- [32] I. V. Soares, M. Yarime, and M. Klemun, “Balancing the trade-off between data center development and its environmental impacts: A comparative analysis of Data Center Policymaking in Singapore, Netherlands, Ireland, Germany, USA, and the UK,” *Environmental Science & Policy*, vol. 157, p. 103769, 2024.
- [33] ADEME, “Base empreinte,” <https://base-empreinte.ademe.fr/donnees/jeu-donnees>, accessed July 2025.
- [34] “Carbon Intensity of Electricity Generation,” Our World in Data, <https://ourworldindata.org/grapher/carbon-intensity-electricity>, 2025.
- [35] Mordor Intelligence, “Germany Data Center Market Size,” <https://www.mordorintelligence.com/industry-reports/germany-data-center-market>, 2025.
- [36] —, “India Data Center Market Size,” <https://www.mordorintelligence.com/industry-reports/india-data-center-market>, 2025.
- [37] Y. Jiang, R. B. Roy, R. Kanakagiri, and D. Tiwari, “WaterWise: Co-optimizing Carbon- and Water-Footprint Toward Environmentally Sustainable Cloud Computing,” in *ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP)*, 2025, p. 297–311.
- [38] RTE, “Futurs   nerg  tiques 2050,” <https://rte-futursenergetiques2050.com/>, 2022.
- [39] P. Reig, T. Luo, E. Christensen, and J. Sinistore, “Guidance for Calculating Water Use Embedded in Purchased Electricity,” World Resources Institute, Feb. 2020.
- [40] Sphera, “LCA of Dell Servers R6515, R7515, R6525, R7525,” <https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/full-lca-of-dell-severs-r6515-r7515-r6525-r7525.pdf>, 2021.