Optimal sizing of a globally distributed low carbon cloud federation

Miguel Vasconcelos, Daniel Cordeiro, Fanny Dufossé, Jean-Marc Nicod, Veronika Sonigo

Optimal sizing of a globally distributed low carbon cloud federation

Miguel Felipe Silva Vasconcelos, Daniel Cordeiro, Georges da Costa, Fanny Dufossé, Jean-Marc Nicod, Veronika Rehn-Sonigo

Optimal sizing of a globally distributed low carbon cloud federation

Miguel Vasconcelos[✦★], Daniel Cordeiro[✦], Georges Da Costa[✧], Fanny Dufossé[★], Jean-Marc Nicod[☆] and Veronika Rehn-Sonigo[☆]

[✦]*School of Arts, Sciences and Humanities, University of São Paulo, Brazil*,
[✧]*IRIT, Université de Toulouse, CNRS UMR 5505, Toulouse, France*,
[★]*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, Grenoble, France*,
[☆]*FEMTO-ST Institute, Univ. Bourgogne Franche-Comté, CNRS, ENSMM, Besançon, France*
miguel.silva-vasconcelos@inria.fr, daniel.cordeiro@usp.br, georges.da-costa@irit.fr,
fanny.dufosse@inria.fr, jean-marc.nicod@ens2m.fr, veronika.sonigo@univ-fcomte.fr

***Abstract*—The carbon footprint of IT technologies has been a significant concern in recent years. This concern mainly focuses on the electricity consumption of data centers; many cloud suppliers commit to using 100% of renewable energy sources. However, this approach neglects the impact of device manufacturing. We consider in this paper the question of dimensioning the renewable energy sources of a geographically distributed cloud with considering the carbon impact of both the grid electricity consumption in the considered locations and the manufacturing of solar panels and batteries. We design a linear program to optimize cloud dimensioning over one year, considering worldwide locations for data centers, real-life workload traces, and solar irradiation values. Our results show a carbon footprint reduction of about 30% compared to a cloud fully supplied by solar energy and of 85% compared to the 100% grid electricity model.**

***Index Terms*—cloud computing, renewable energy, energy storage, linear program, job scheduling, follow-the-sun, green computing**

I. Introduction

Data centers (or DCs), the infrastructure that hosts the cloud, consume 1% of the total power generated in the world [1]. Research and industry sectors are unitedly making efforts to reduce this impact: during the period from 2010 and 2018, there was a 10-fold increase in IP traffic, a 25-fold increase in storage capacity, and a 6-fold increase in DCs workload. However, energy consumption only increased by 6% thanks to improvements in efficiency [2]. On the other hand, some studies predict that the energy demand will keep growing. For instance, [3] consider scenarios for the period 2016–2030, with predictions ranging between a wavering balance and a significant increase in electricity needs.

In order to reduce the environmental impact of data centers' operation, major cloud players committed to include renewable energy in their operation, such as Google, Apple, Microsoft, Amazon AWS, and Facebook [4]. Furthermore, solar energy costs have fallen 85% (and are expected to keep decreasing) and its deployment increased over than 10 times from 2010 to 2019 [5]. Moreover, the solar irradiation has a lower variation than wind speed [6], which can increase the accuracy of predictions for the workload scheduling decision. However, integrating renewable energy is not trivial given its intermittent nature: its production is variable and affected by the weather, time of the day, geographic location, and seasons. One possible strategy to reduce the impact of intermittency is the adoption of Energy Storage Devices (ESD), such as lithium-ion batteries, to store the renewable energy to be used later.

Defining the optimal area of the photovoltaic panels (PV) and the capacity of the ESDs (referred in the literature as sizing, dimensioning, or capacity planning) should not neglect the following facts: the electricity from the regular electrical grid in some locations of the world already incorporates a share of renewable energy; manufacturing photovoltaic panels and batteries does generate carbon emissions, and each region in the world has a different potential to produce green power.

The notation of Carbon-Responsive Computing [7] classifies strategies that are aware of the carbon intensity and use this information to make decisions. Follow-the-renewables approaches [8] are one example of carbon-responsive strategies for reducing the environmental impact of cloud operations. It incorporates information on the availability of renewable energy in the scheduling decision. This way, the workload can be allocated or migrated to locations with more green energy.

In this work, we explore the adoption of both strategies, sizing the PVs and batteries and scheduling with follow-the-renewables to make the operations of existing cloud platforms greener. More specifically, this paper presents the following contributions: (i) we model the two sub-problems—PVs and batteries sizing, and workload scheduling—as a single problem, which allows evaluating scenarios such as: should the battery capacity or the PV area be increased, or should the workload be scheduled in a data center located in another part of the world? (ii) we propose a model that uses a linear programming approach (LP) with real variables, allowing us to optimally solve the problem we address in polynomial time using classical LP solvers. This allows a large number of scenarios to be evaluated over broad time horizons (i.e., one year) to take the seasonal behavior of renewable energy production into account. This model can be extended to multiple scenarios, and it may help decision-makers evaluate

The list of authors is sorted in alphabetical order, except for the first one who led all the experiments.

which regions need more investment to reduce the cloud operation's environmental impact.

The remainder of the paper is organized as follows: Section II gives an overview of the state of the art about sizing IT platforms dedicated to the cloud in general or in the green IT context in particular. Section III defines the problem addressed, the assumptions, the models, and the objective function. Details about the problem constraints and how to optimally solve the problem are given in Section IV. Comprehensive experiments, presented in Section V, are discussed in Section VI before we conclude in Section VII.

II. Related work

Most sizing research focuses on a single DC. There are two approaches, either to consider that the DC can use the electrical grid as a fallback or to consider how to size a DC only with on-site renewable sources.

Most sizing approaches consider the capability to use the electrical grid. Padma Priya and Rekha [9] use a Particle Swarm Optimization approach for sizing a smart microgrid to supply fog DCs located in a rural area in India. The objective of the optimization is to reduce the capital cost of buying solar panels, wind turbines, diesel generators, and batteries. Power from the regular electrical grid can be used when there is no green power production. The authors also propose a scheduling algorithm to maximize green energy usage. Niaz et al. [10] evaluates using curtailed renewable energy to power DCs and provide hydrogen to hydrogen refueling stations. The authors model their problem as a MILP (Mixed Integer Linear Programming) with the objective of minimizing the costs. System components included natural-gas–powered combined cooling, heating, and power systems, electrolyzers, hydrogen fuel cells, heat pumps, hydrogen tanks, and battery energy storage systems. The results were that using only power from the electrical grid was the worse in both economic and environmental terms. Using a mix of curtailed renewable energy and electricity from the grid was the most economical. Using only renewable energy was the best for the environment; however, it had the highest costs.

In some cases, the approach considers how also to size on-site energy production, removing the need to access the electrical grid. Richter et al. [11] proposes a planning methodology for net-zero energy systems, and performed a qualitative study to evaluate a net-zero energy DC located in Germany. The conclusion is that by selecting appropriate technologies for energy generation, increasing energy efficiency, and optimal sizing Energy Storage Systems, the DC showed large potential to operate as a net-zero energy system. A DC as a net-zero energy system can increase the marketing image and add economic value to the related company. Haddad et al. [12] proposes to size a DC using only on-site renewable energy and energy storage systems (batteries and hydrogen). This work focuses on a single DC and discusses the impact of its location, its workload, and its context on the resulting sizing (number of servers, renewable sources, and storage). Benaissa et al. [13] proposes to reduce the usual oversizing of renewable-powered DCs. Classical sizing approaches based on traces are defined by a few days with unusually high workloads and/or low renewable availability. In this work, the authors propose to reduce such sizing and evaluate the impact on the Quality of Service and on the sizing itself. Contrary to the previous studies, they use a binary search approach to find the best relevant sizing instead of MILP formulation.

Some research focuses on the sizing of particular elements, such as the electrical infrastructure. Sheme et al. [14] studies the impact of the battery size to reach a specific green coverage of 50% (half of the energy consumption of the DC needs to be green). They develop a simulation tool that uses as input the area of PVs and capacity of the batteries. Experiments comparing countries (Finland, Crete, and Nigeria) show that the number of solar panels needed in Crete and Finland is slightly higher than in Nigeria, 17% and 45%, respectively. However, although Finland provides only 15% less annual solar energy than Nigeria, it requires a battery size of 39 times bigger to achieve wasted energy at level 0. While in Crete, a battery capacity of only 27% greater than in Nigeria is needed.

Overall, most studies focus on sizing individual DCs. This is similar in the context of scheduling renewable-powered DCs: Song et al. [15] reviews recent publications on the field of DCs powered by Renewable Energy mix. It shows that among more than 100 publications, only a quarter focuses on geographically distributed data centers partially powered by renewable energy mix. It also shows that most research in this field focuses on workload scheduling, while few articles focus on the adaptation or sizing of the infrastructure.

The Carbon Explorer framework is an example of a study that explores sizing multiple DCs [16]. The framework explores three solutions to achieve 100% renewable operation of DCs distributed over the United States of America: i) only use renewable energy; ii) use renewable energy and energy storage; and iii) use renewable energy and schedule the workload. These DCs already have access to local solar power, wind power, or both. The carbon emissions from manufacturing PVs, wind turbines, batteries, and servers are considered. An exhaustive search is used to find the solutions. The work concludes that 100% renewable operation may not be the optimal solution when considering the geographic location of the DC, and the carbon emissions from the manufacturing phase. Furthermore, the authors say that choosing the optimal solution is still an open research question for future work.

Our approach focuses on optimally sizing of geo-distributed DCs across the globe, which has not yet been studied to the best of our knowledge. Furthermore, in contrast to the Carbon Explorer framework, our solution allows using the regular electrical grid when opportune, given it may be supplied by a low-carbon intensive source. Finally, contrary to most studies using a MILP, our model uses a linear program formulation.

III. Problem statement

This section is dedicated to the description of the addressed problem and hypothesis. The next two sections give details

about the chosen model, notations, and the optimal approach to solve the decision problem that we tackle within the paper.

A. Addressed problem

Our goal is to reduce the carbon footprint of existing cloud platforms by increasing renewable energy usage and reducing their sizing. The considered platform consists of several data centers spread worldwide, on all continents, and in both hemispheres. The chosen approach aims to design an additional solar-based power supply infrastructure to the classical power grid connection and to define an optimal way to operate the global IT cloud platform considering a given workload to complete. To green the cloud infrastructure and reduce its carbon footprint, we have to limit the usage of energy from fossil fuels to operate the data centers. Using renewable energy is a promising option. However, the production of solar panels also has a carbon footprint, as does the construction of batteries, even if we pay this carbon footprint only once. We have to take this aspect into account.

Another problem is that the locality of the DC determines how sustainable it can be. The carbon footprint of the electricity offered by the local power grid depends on how it is produced: natural gas, coal combustion, hydraulic or nuclear energy. Also, producing electricity from solar energy is less efficient depending on whether the solar panels are installed near or far from the tropics. Powering the cloud federation is a balance or mix between using low-carbon electricity from the grid and using solar panels, with the understanding that the solar panels have to be installed and are necessarily associated with batteries to mitigate intrinsic solar power intermittency.

This decision problem aims at defining the additional renewable power supply architecture from solar energy to reduce the carbon footprint of the global cloud infrastructure. Results depend on the inputs and hypothesis of the problem. We consider as input: (i) carbon footprint of manufacturing the targeted electrical components (PV and batteries), as well as the emissions from using grid electricity depending on the country of the production; (ii) the computing demand from clients (jobs with a given amount of computation); and (iii) weather conditions (solar irradiation) in areas where each data center operates for the federation. It is assumed that job submission is centralized, 100% of the jobs must be completed in time, and the cloud platform is homogeneous.

We now introduce the models and notations before the objective function to optimize.

B. Models and notations

To propose a solution in terms of job operations, we define a decision horizon $\mathcal{H}$ in which job scheduling decisions can be taken. To do so, we propose to discretize $\mathcal{H}$ into K indivisible time slots whose duration is Δt such that $\mathcal{H} = K \times \Delta t$. To simplify the notations, we consider $\Delta t = 1u.t.$ (unit of time), knowing that, in practice, we assume that $\Delta t = 1h$ such that $K = 8760\,h$ with $\mathcal{H} = 1$ year. Let k be the index of the time slot that addresses any time instant t such that $k\Delta t \leq t < (k+1)\Delta t$ with $0 \leq k < K$.

TABLE I
MAIN NOTATIONS FOR THE IT MODEL FOR EACH DC^d $(1 \leq d \leq D)$ DURING TIME SLOT k $(0 \leq k < K)$

Δt	time duration of each time slot in unit of time $[u.t]$
$\mathcal{H}$	decision horizon $\mathcal{H} = K\Delta t$
K	number of time slots $\Delta t = 1\,\text{h} = 1\,u.t.$
k	time slot between dates $k\Delta t$ and $(k+1)\Delta t$ excluded
DC^d	data center number d of the cloud federation
$\mathcal{DC}$	the set of all data centers $\{DC^d \mid d = 1, \ldots, D\}$
C^d	number of cores within DC^d
$Pcore$	dynamic power consumption of one core
$Pidle^d$	static power consumption of DC^d
$Pintranet^d$	consumption of the DC^d interconnection network
P_k^d	the power demand to perform tasks on DC^d during time slot k
PUE	Power Usage Effectiveness (constant value)
$\mathcal{T}$	the workload to perform $(= \{T_i \mid 1 \leq i \leq N\})$
T_i	task i of the workload $\mathcal{T}$ $(1 \leq i \leq N)$
r_i	release date of tasks T_i
p_i	processing time of tasks T_i
c_i	number of cores needed to execute task T_i
w_k	number of cores needed during the kth time slot
w_k^d	number of cores needed during the kth time slot on DC^d

1) IT part model: As we plan to green an existing cloud federation, it is assumed that the data centers have already been designed in terms of the IT and electrical infrastructure, and they are homogeneous regarding the number of CPU cores. Let $\mathcal{DC} = \{DC^d \mid d = 1, \ldots, D\}$ be the set of data centers in the federation. Considering a given data center DC^d, let C^d be its number of cores and $Pcore$ the energy used to power one core. In order to execute all tasks assigned to DC^d at time slot k, the DC will need an amount of energy of P_k^d.

The power consumption of a cloud data center can be classified as static or dynamic [17]. For the static part, the current model considers the idle power consumption $Pidle^d$ of the servers, the Power Usage Effectiveness (PUE) to represent the power consumption used to cool the DC infrastructure, and the power consumption of the network switches $Pintranet^d$ that interconnect the servers in each data center DC^d. Regarding the latter, cloud data centers usually adopt the fat-tree topology to interconnect servers in the DC [17]. In this topology, one can compute the number of network switches needed to match the number of servers. The power consumption of the network switches is considered to be static based on actual measurements, which have shown that the consumption does not change significantly with the device usage [18]. Moreover, each geographic location has different cooling needs, therefore each DC has a specific PUE^d value.

Finally, the dynamic part is represented by the additional power consumption generated by using the CPU cores in each data center. Equation (1) represents the power consumption of each DC for each step k $(0 \leq k < K)$:

$$P_k^d = PUE^d \times \left(Pidle^d + Pintranet^d + Pcore \times w_k^d\right) \quad (1)$$

TABLE II
MAIN NOTATIONS FOR THE ELECTRICAL PART MODEL OF EACH DC^d ($1 \leq d \leq D$) DURING EACH TIME STEP k ($0 \leq k < K$)

I_k^d	solar irradiation at time slot k [W/m^2]
Apv^d	surface of PVs of DC DC^d [m^2]
η_{pv}	PV efficiency
$Pgrid_k^d$	power from the grid at time slot k [W]
Pre_k^d	power from PVs at time slot k [W]
BAT^d	battery capacity installed in DC^d [Wh]
B_k^k	battery level of energy at the time $k \times \Delta t$ [Wh]
Pch_k^d	power charge during time slot k [W]
$Pdch_k^d$	power discharge during time slot k [W]
η_{ch}	charging process efficiency
η_{dch}	discharging process efficiency

2) Workload model: Considering the workload to complete, let $\mathcal{T} = \{T_i \mid i = 1, \ldots, N\}$ be the set of N tasks that have to be executed in time upon the whole cloud federation during the time horizon $\mathcal{H}$. Each task T_i has a release date r_i, a processing time p_i, and must be executed on c_i cores. Let w_k be the total number of cores needed to compute tasks during the time slot k in order to complete the workload in time. At each time step, w_k is the sum over all cores required by the tasks executed in time slot k. This is also the sum of the number of cores w_k^d of each data center DC^d in $\mathcal{DC}$ (with $1 \leq d \leq D$) and $0 \leq k < K$) as shown by Equation (2):

$$w_k = \sum_{T_i | r_i \leq k\Delta t < r_i + p_i} c_i = \sum_d w_k^d \tag{2}$$

3) Electrical part model: The power supply of the whole cloud platform $\mathcal{DC}$ is coming both from the classical power grid $Pgrid_k^d$ of each country on which $DC^d \in \mathcal{D}$ is hosted, and from renewable energy (the sun) thanks to solar panels (PVs) installed on each DC^d site. As renewable energies are inherently intermittent and cannot be controlled, storage devices are mandatory either to store the overproduction when the sun shines or to provide the missing energy during the night. Batteries have been chosen to play this role because of their good efficiency in terms of costs, power and energy density, charge and discharge rates, and self-discharge [19]. So, the whole cloud platform power demand can be supplied by the regular grid, power generated from the PV panels, and power discharged from the batteries. As a constraint, the power from the grid ($Pgrid_k^d$) is always positive because the platform is not supposed to sell energy. DCs may use the regular electrical grid as backup when the PV power production is not enough, or if the stored energy in the batteries is not sufficient.

The amount of renewable power that can be produced depends on the solar irradiation I_k^d received at the location of DC^d during the time slot k, on the surface of the solar panels Apv^d and on the efficiency η_{pv} of the PVs. Equation (3) models the on-site renewable power production.

$$Pre_k^d = I_k^d \times Apv^d \times \eta_{pv} \tag{3}$$

Batteries are systematically installed next to the PVs for the reason mentioned above. Let B_k^d be the amount of energy (in Wh) at time kt stored in batteries of capacity BAT^d installed in DC^d ($B_0^d = Binit^d$ being the amount of energy at the beginning the time horizon $\mathcal{H}$). These batteries are charged during the time slot k thanks to the power charge Pch_k^d. Conversely, $Pdch_k^d$ is the power discharge of the batteries. If Pch_k^d is greater than zero, $Pdch_k^d$ equals zero and vice versa. Regarding the batteries modeling, the charging and discharging process have an efficiency η_{ch} and η_{dch} less than 1. The self-discharge property has not been modeled because for lithium-ion batteries the value is very low (0.5 % per day) [19]. Finally, to increase the lifetime, the batteries cannot be discharged more than its Maximum Depth of Discharge. We added restrictions to the B_k^d variable to model this property. Equation (4) models the battery in terms of the level of energy:

$$B_k^d = B_{k-1}^d + Pch_{k-1}^d \times \eta_{ch} \times \Delta t - \frac{Pdch_{k-1}^d}{\eta_{dch}} \times \Delta t \tag{4}$$

with $0.2 \times BAT^d \leq B_k^d \leq 0.8 \times BAT^d$ for any time slot k and DC^d ($0 \leq k < K$ and $1 \leq d \leq D$). The modeling of the batteries and PVs are based on [20].

4) Footprint model: In the current model, carbon emissions originate from three sources: (i) the regular power grid and the manufacturing of both (ii) the photovoltaic panels, and (iii) the batteries. The carbon footprint of the regular power grid given by Equation (5) is calculated based on the carbon intensity of the power grid $gridCO2^d$ in the region of data center DC^d times the amount of energy used during the time slot k.

$$FPgrid_k^d = Pgrid_k^d \times \Delta t \times gridCO2^d \tag{5}$$

The $gridCO2^d$ input reflects the share of carbon-intensive power sources in the regular electricity grid: if it is supplied by solar, wind power, hydroelectric or nuclear power, the value will be low. However, the value will be higher if it is supplied by coal, oil, biomass, or natural gas. It is considered that the electrical grid may be supplied by multiple power sources.

For the PVs, in order to account for the fact that the amount of solar irradiation received is not homogeneous for different geographic regions, one must also consider the expected power output that PVs can produce over their lifetime relative to the cost of manufacturing. Therefore, the carbon footprint of PVs is also related to the location of each data center. The carbon footprint thus defined is modeled by Equation (6):

$$pvCO2^d = \frac{FPpv_{1m2}}{expectedEpv^d} \tag{6}$$

where $FPpv_{1m2}$ is the cost of manufacturing $1\,m^2$ PV in $g\,\mathrm{CO}_2 - eq$, and $expectedEpv^d$ is the expected energy production in Wh that $1\,m^2$ of PV can produce during its lifetime at the location of DC^d. As a result, the unit of this metric is expressed in $g\,\mathrm{CO}_2 - eq.Wh^{-1}$, and so, the total emissions from the PVs are related to its power production, as shown in Equation (7).

$$FPpv_k^d = pvCO2^d \times Pre_k^d \times \Delta t \quad (7)$$

Finally, the batteries' carbon footprint $FPbat^d$ of the DC DC^d is related to their capacity BAT^d in kWh and carbon emissions of the manufacturing process $batCOS$ in $g\,\mathrm{CO}_2 - eq.kWh^{-1}$, as seen in Equation (8). To be consistent with the calculation of $FPpv_k^d$, $batCO2$ is the share of the carbon footprint of the battery type chosen for a capacity of $1\,kWh$ over the time horizon of $\mathcal{H}$, assuming a battery has a lifetime of 10 years. Thus $batCO2$ is the tenth of the total carbon footprint of the considered battery, given that we are considering 1 year of cloud operation.

$$FPbat^d = BAT^d \times batCO2 \quad (8)$$

These modifications regarding the lifetime of PVs and batteries were necessary because we are considering only one year of cloud operation. If we use the total carbon emissions for manufacturing the PVs and batteries, the solver will find a solution where there is few to no PV or batteries, because using the regular electrical grid would be less carbon-intensive.

C. Objective function

Now that the models have been introduced, the objective function can be defined (see Equation (9)). It consists of minimizing the carbon footprint of the globally distributed cloud federation in order to reduce as much as possible carbon emissions, which come from both the consumption of electricity from the power grid, as well as from the manufacturing of photovoltaic panels and batteries with k and d defined as follows: $0 \leq k < K$ and $1 \leq d \leq D$.

$$\text{minimize} \sum_{k=0}^{K-1} \sum_{d=1}^{D} (FPgrid_k^d + FPpv_k^d) + \sum_{d=1}^{D} FPbat^d \quad (9)$$

IV. Optimal resolution

The models presented in the previous section consist of several linear equations. We show in this section that constraints governing the use of the globally distributed cloud platform can be expressed as linear expressions. New real variables are introduced to finalize the linear program that needs to be solved to achieve the targeted objective. The solution obtained after solving the linear program is optimal in nature and computed in polynomial time, as long as the variables are not integers. Polynomial time is mandatory if we consider the number of variables needed for a time horizon $\mathcal{H}$ as long as one year. We assume to choose real positive values for all variables even if variables denote discrete objects like cores. Indeed, w_k^d is the number of cores needed to run tasks on DC^d during time slot k. Considering the size of the cloud with its thousands of cores, the decimal part of each w_k^d can be neglected. Having the solution with more or less than a core on a given DC does not change the order of magnitude for the PV and battery sizing process.

The globally distributed cloud platform that we plan to optimally size, as mentioned in the problem statement (Section III), has only one goal: completing a given amount of work during a given year and knowing the weather conditions during the same year. We present a set of constraints that must be respected to make this mission possible. Some constraints are explicit, and some are implicit to avoid the addition of integer variables which would transform this LP into a MILP whose solving process would not scale at all.

A. Constraints to address the workload

Since the distributed cloud federation configuration is defined a priori by a set of existing cloud DCs at each chosen location, the amount of work to be performed must respect each data center's computational capabilities DC^d. Equation (10) expresses that the number of cores that are switched on does not exceed the existing number of cores of DC^d:

$$w_k^d \leq C^d \quad (10)$$

B. Constraints to reach the power demand

The electric part of each DC has to supply the DC power demand using renewable energy (Pre_k^d), from batteries ($Pdch_k^d$ and Pch_k^d) and/or from the classical grid ($Pgrid_k^d$). Equation (11) presents the restriction for the power consumption.

$$P_k^d \leq Pre_k^d + Pgrid_k^d + Pdch_k^d - Pch_k^d \quad (11)$$

C. Constraints on batteries

The batteries are defined by their capacity, which is different for each DC and depends on how the intermittency of the renewable energies is managed on each site. The other quantities concerning the batteries depend on the total capacity of the batteries, DC by DC. This allows realistic behaviors for the batteries to be assumed. These limitations concern the practical level of use of the energy stored in the batteries, which cannot be completely emptied, for example. In addition, the power to charge or discharge a battery is also limited by the level of energy remaining in the associated battery, so that it is not possible to reach a forbidden energy level. Equations (12), (13) and (14) express these constraints:

$$0.2 \times BAT^d \leq B_k^d \leq 0.8 \times BAT^d \quad (12)$$

$$Pch_k^d \times \Delta t \times \eta_{ch} \leq 0.8 \times BAT^d - B_{k-1}^d \quad (13)$$

$$Pdch_k^d \times \Delta t \;/\; \eta_{dch} \leq B_{k-1}^d - 0.2 \times BAT^d \quad (14)$$

One may notice that we are not modeling any restrictions for charging and discharging simultaneously. Such restrictions would require the usage of binary variables that would significantly increase the required computational time to find the optimal solution to the problem. We performed experiments with a shorter duration (around 1 month), and the sizing results were the same between both versions: using and not using binary variables. Furthermore, it is possible to calculate an alternative solution for the linear program where there would be no charge and discharge at the same time slot by increasing or decreasing the value of the variables Pch_k^d and $Pdch_k^d$.

D. Linear program

This following linear program (LP) summarises what has been described before concerning the model that has to be respected to solve the tackled problem. All variables given by the solution obtained after the solving process are used to completely define both the renewable power supply part and the core operating process of the distributed low carbon cloud federation and the way each DC is used time slot by time slot for one year on the considered weather conditions. Comprehensive experiments have been led to highlight the pertinence of the approach. These experiments are shown in the next section, and a discussion is proposed in Section VI.

$$\text{(LP)} \begin{cases} \text{minimize} \sum_{k=0}^{K-1} \sum_{d=1}^{D} (FPgrid_k^d + FPpv_k^d) + \sum_{d=1}^{D} FPbat^d \\ \\ \text{s.t.} \quad (1)\,(3)\,(4)\,(5)\,(7)\,(8)\,(10)\,(11)\,(12)\,(13)\,(14) \end{cases}$$

where all variables are positive real variables.

V. Experiments

In this section, we present the settings and the results of our experiments. More details for reproducing the experiments can be found in Appendix A.

A. Settings

1) Cloud infrastructure: The servers are homogeneous and based on equipment of real cloud infrastructure: the Taurus server of the Grid'5000 testbed[1]. The servers are equipped with two Intel Xeon E2630 CPUs, with a total of 12 cores. For modeling the power consumption of the servers, real measurements conducted by Ahvar et al. [17] were considered: in the idle state, each server consumes 97 W, and their maximum power consumption (when using 100% of the 12 cores) is 220 W. The value of Pcore is 10.25 W, and it was obtained by linear interpolation between the power consumption of the idle and the fully used state.

Each data center is equipped with 23,200 servers (and a total of 278,400 cores). This number matches what can be seen in production data centers of major cloud players: Microsoft operates over 4 million servers distributed over 200 DCs [21].

We considered a network with a 48-ary fat-tree topology linked by 2,880 switches with 48 ports each. The power consumption of the switches was based on real measurements by Hlavacs et al. [18]: the HP ProCurve 2810-48G was selected, with 48 ports and approximately 52W per device.

For the location of the data centers, it was based on the real cloud infrastructure of Microsoft Azure[2], and different regions in different continents, hemispheres, and time zones were selected. Figure 1 presents the details of the locations.

We used values for the PUE inspired by real data from Microsoft Azure for each region: for the Americas, the PUE is 1.17 (DCs São Paulo and Virginia), Asia Pacific has a PUE of 1.405 (DCs Pune, Canberra, Singapore, and Seoul), and for the Europe region, Middle East, Africa the PUE is 1.185 (DCs Johannesburg, Dubai, and Paris) [22].

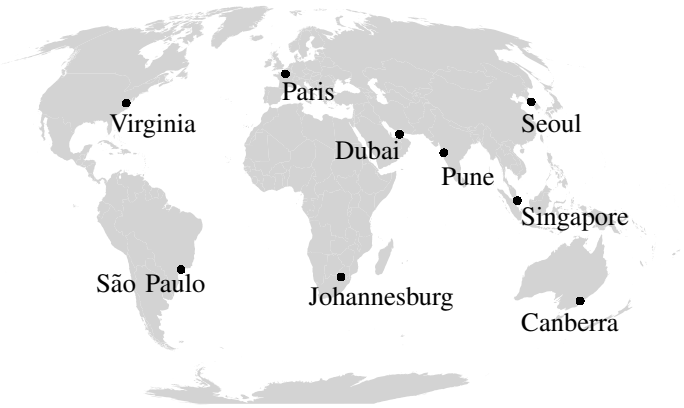

Fig. 1. Selected locations for the data centers.

2) Workload: The workload used was generated using the Grog generator[3], a workload generator based on analysis of properties of the execution trace made available by Google in 2011 [23]. For reproducibility purposes, the parameters regarding the number of tasks were set to 350,000, the duration was 30 days, and it was executed 12 times (1 per month). The tasks have a duration of one hour.

3) Photovoltaic power production: Global Solar Horizontal Irradiation (in Wh/m^2) data was collected from the MERRA-2 project [24], since it provides information for anywhere on earth. Figure 2 illustrates different solar irradiation of each location throughout the year 2021.

4) Carbon footprint: For solar panels, it is considered a lifetime of 30 years, and manufacturing $1m^2$ emits 250 $kg\,CO_2 - eq$, inspired from real measurements [25]. To compute the emissions in the form of $g\,CO_2 - eq.kWh^{-1}$ as stated in Section III-B4, we considered the total solar irradiation that was produced during the year 2021 multiplied by 30 (to account for the PV module lifetime of 30 years). For the electrical grid, we also considered the real-world data of the carbon footprint ($g\,CO_2 - eq.kWh^{-1}$). Table III lists the carbon emission values for each region.

Regarding the batteries, the emissions are only considered for the manufacturing step—59 $kg\,CO_2 - eq$ per kWh. In our experiments, the considered lifetime of the batteries is ten years. Therefore, the input used is equal to 5.9 $kg\,CO_2 - eq$ per kWh, given that we simulated one year.

5) Execution environment: We ran the experiments on a machine with an Intel i9-11950H CPU, and 32 GB of RAM. The solver used was the Gurobi Optimizer (version 9.5.2). The execution time for solving the LP with the inputs listed in the previous sections — which resulted in a total of 394,263 variables — was in the order of 30 seconds.

[1]https://www.grid5000.fr/w/Lyon:Hardware#taurus

[2]Azure global infrastructure: https://infrastructuremap.microsoft.com/.

[3]https://pypi.org/project/grog/

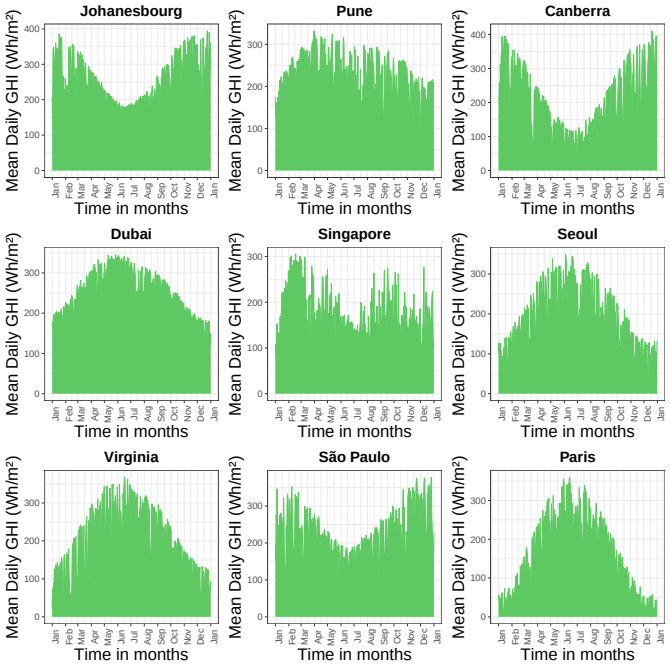

Fig. 2. Average daily solar irradiation per location throughout the year 2021.

TABLE III
EMISSIONS (IN $g\,CO_2 - eq.kWh^{-1}$) FOR BOTH PV USAGE AND USING THE REGULAR GRID. SOURCE FOR GRID EMISSIONS: ELECTRICITYMAP, CLIMATE-TRANSPARENCY.ORG.

Location	Grid	PV
Johannesburg	900.6	24.90
Pune	702.8	27.96
Canberra	667.0	29.71
Dubai	530.0	24.84
Singapore	495.0	36.19
Seoul	415.6	34.00
Virginia	342.8	31.71
São Paulo	61.7	27.99
Paris	52.6	39.93

B. Results

In this section, we present the results in terms of the computed optimal area of the PVs and capacity of the batteries, the source of energy that was consumed by the DCs operation (grid, batteries, or PV panels), and the total emissions of the cloud operation, generated from both manufacturing PVs and batteries, and power consumption of the regular electrical grid. Furthermore, to assess the solution computed by the LP, we compare it with two other scenarios: i) only power from the regular electrical grid is used to supply the DCs (represent current DCs), and ii) only power generated from the PV panels, and stored and discharged from the batteries are used to supply the DCs. Finally, we present an evaluation using metrics to assess the environmental impact of the results.

Figure 3 illustrates the area of the photovoltaic panels and the capacity of the batteries computed from the LP using the inputs described in Section V.

To analyze the sources of energy that supplied the DCs

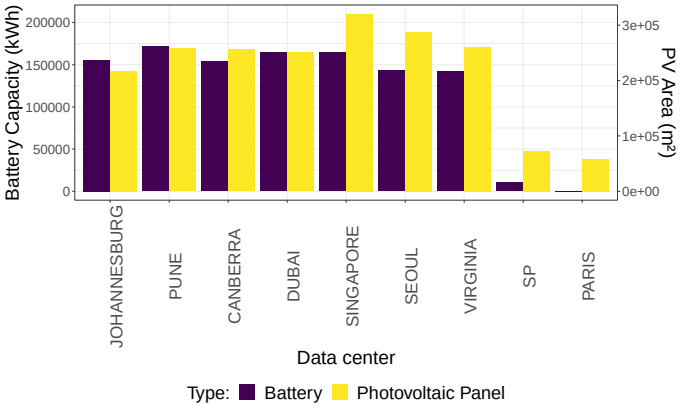

Fig. 3. Optimal result for the area of PV panels and capacity of the batteries.

operation, we present in Figure 4 the percentage that each source (grid, renewable, and batteries) was used to daily supply the DCs throughout the year. Figure 5 is a fine-grain visualization of the DC operation regarding the power consumed or produced: it illustrates hour-by-hour the DC total power demand, how much power was consumed from the grid, discharged from the batteries, and produced by the PV panels.

In order to assess the optimal solution of the LP, we compared it with two other scenarios in terms of total carbon emissions ($t\,CO_2 - eq$): i) the DCs are only supplied by power from the regular electrical grid, and ii) the DCs are only supplied by power from the photovoltaic panels and batteries. Table IV presents the results. In comparison with the first scenario (only grid power), the reduction in the CO_2 emissions was approximately 85%, and it was approximately 30% for the second scenario (only renewable power).

TABLE IV
TOTAL EMISSIONS FOR THE DIFFERENT SCENARIOS.

Scenarios	Emissions ($t, CO_2 - eq$)
Electrical grid	201211.3
PV and batteries	42370.6
PV, batteries, and grid	29600.6

To further evaluate these scenarios, we present in Table V results in terms of the average load each DC executed throughout the year. Equation (15) represents how the metric was computed for each DC d.

$$\frac{\sum_k w_k^d}{C^d \times K} \tag{15}$$

To evaluate the environmental impact of the solution, we used metrics extracted from [26]. The first metric, the Green Energy Coefficient (or GEC), is the ratio between the total green power generated and the DC total energy consumption, and it can illustrate the oversizing of the green power supply infrastructure. The second metric is the CO_2 savings, which represents the emissions reduction after DC equipment upgrade or flexibility mechanisms. CO_2 savings is computed as seen in Equation 16, where: $CO2_{current}$ represents the

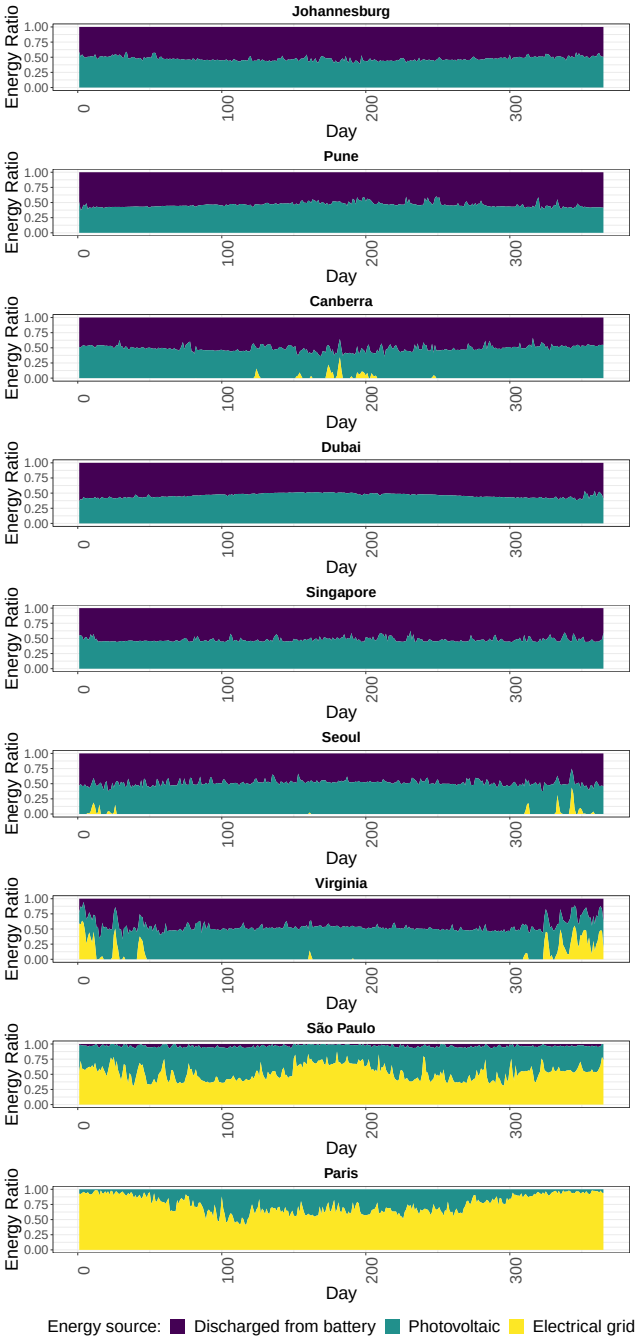

Fig. 4. Composition of the DCs' daily energy consumption throughout the year considering the different sources of energy, where 1.0 is the DC's total energy consumption.

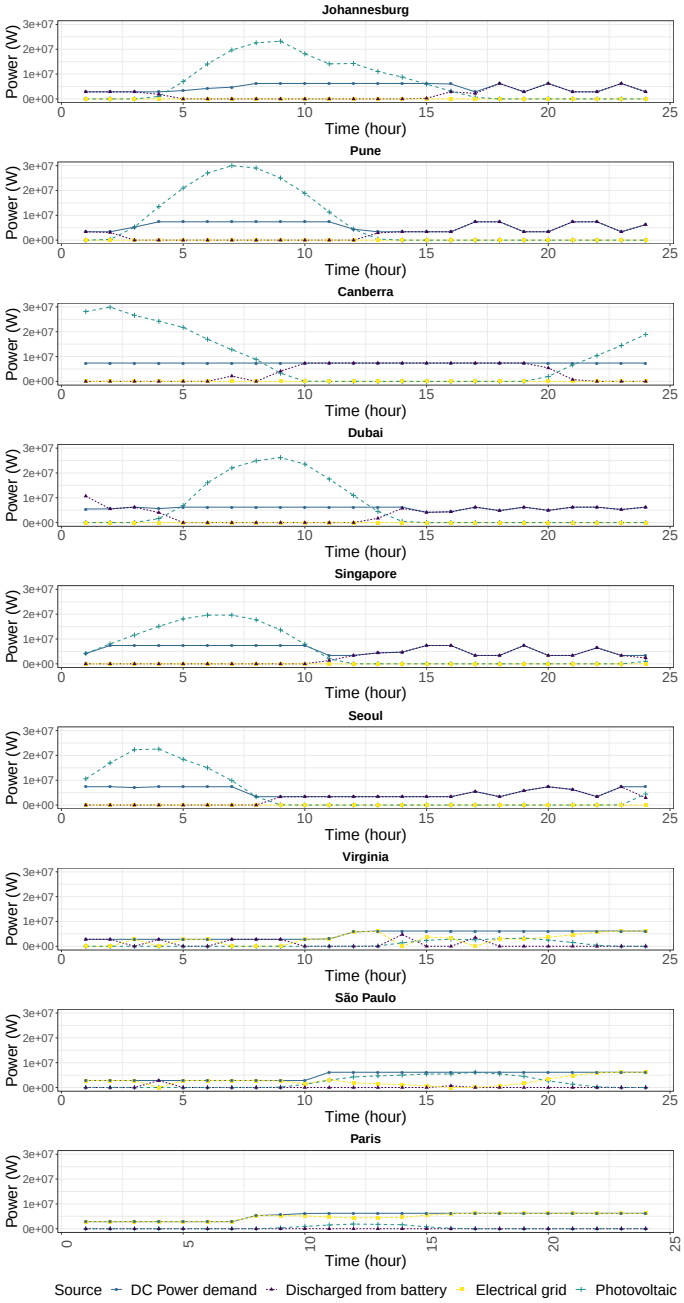

Fig. 5. Composition of the DCs' hourly power consumption throughout the first day of the year. Time follows the Universal Time (UT) standard.

system studied after the modifications (the result of the linear program for the sizing of PVs and batteries) and $CO2_{baseline}$ the system in its original state. Here, it was considered that $CO2_{baseline}$ has the same workload allocation of $CO2_{current}$; the difference between the two is that $CO2_{baseline}$ does not have PVs and batteries, and thus only consumes power from the grid. Table VI shows the computed values for both metrics.

$$CO2_{savings} = \left(1 - \frac{CO2_{current}}{CO2_{baseline}}\right) \times 100 \qquad (16)$$

In order to assess the robustness of the sizing process for the area of PV panels and the capacity of the batteries, it is necessary to take into account other meteorological conditions, given that the DCs will operate for decades and not only for one year. The metric selected is the Mean Absolute Percentage Error (MAPE) defined by: $\frac{1}{n}\sum_{i=1}^{n}\frac{|R_i - F_i|}{R_i}$, where n represents the number of values being considered, i the index

TABLE V
AVERAGE DC LOAD THROUGHOUT THE YEAR

Location	Grid	PV + Bat	PV + Bat + Grid
Johannesburg	0	79.31	86.20
Pune	10.25	82.07	89.34
Canberra	99.72	66.62	67.95
Dubai	99.97	93.93	95.11
Singapore	99.93	72.6	85.18
Seoul	99.99	81.87	65.39
Virginia	100.0	88.54	75.51
São Paulo	100.0	63.67	59.06
Paris	100.0	81.24	86.11

TABLE VI
RESULTS OF THE SUSTAINABILITY METRICS FOR THE EXPERIMENTS

Location	GEC	CO_2 savings (%)
Johannesburg	1.47	93.93
Pune	1.45	91.5
Canberra	1.57	89.59
Dubai	1.59	89.1
Singapore	1.42	85.75
Seoul	1.53	82.51
Virginia	1.46	75.99
São Paulo	0.5	20.05
Paris	0.24	5.25

of the value being considered, R_i the real value for the year, and F_i the estimated value (in this case, the computed sizing for the year 2021 that was used in the experiments). Table VII presents the results of the MAPE for both the area of PV and capacity of the batteries when we solve the LP using as input the solar irradiation for the years 2018, 2019, and 2020. Results indicate a variation of less than 10% in the different DCs over the years.

TABLE VII
EVALUATING SIZING FOR DIFFERENT YEARS USING THE MAPE METRIC (VALUES ARE IN %)

Location	PV Area	Battery Capacity
Johannesburg	1.72	1.64
Pune	3.72	0.76
Canberra	8.62	4.25
Dubai	2.31	2.88
Singapore	7.22	0.34
Seoul	3.15	1.11
Virginia	2.2	0.87
São Paulo	5.81	8.05
Paris	2.76	0

VI. ANALYSIS AND DISCUSSION

These results permit the evaluation of the carbon footprint impact of different electricity supply policies for Clouds. On the one hand, as shown in Table IV, there is a significant reduction to obtain by including renewable energy in the electricity sources of DCs. We observe a 5-fold decrease in the footprint in our experiments. Many Cloud providers have committed to using 100% renewable energy supplies for their DCs in the following years. On the other hand, this objective of 100% renewable is, in our opinion, more ideological than pragmatic, and there is more benefit to obtain by combining grid and renewable electricity. We observe in our experiments a further reduction of a fourth in the optimal solution compared to the 100% renewable scenario. This study thus gives further insight into the debate of energy sources in Clouds.

The locations used in this paper for the different DCs allow us to benefit from the diversity of latitudes, hemispheres, and climates, as shown in Figure 1. This variety of longitudes and hemispheres permits mitigation of the impact of seasonal and daily variations of solar irradiation on electricity production and always has at least some DCs with good PV production, as shown in Figure 2. The diversity of climates is highlighted by the case of Singapore's solar production, which is the second lowest with Paris, while its location close to the equator could permit better irradiation.

As indicated in Table III, we observe significant heterogeneity in the carbon footprint of grid electricity of the different DCs, which results in two categories for the optimal solution: i) Paris and São Paulo, DCs with a reduced number of PVs and batteries (no battery in Paris), and ii) the other locations have quite similar sizes of PV and batteries. In the second category, the larger PV area is mainly associated with low solar irradiation. It might appear counterintuitive to allocate more PVs to locations with lower solar production, but this is more comprehensive considering the static part of the power consumption of DCs. When the workload is mainly sent to locations with solar production, the electricity consumption of DC also includes a static part for the idle consumption of servers and the interconnection network, as referred to in Equation (1). This static electricity consumption implies either using the carbon-intensive grid or a sizing of PV and batteries that matches the demand, even during winter days of low PV production. This results in a large PV and battery sizing — the PVs are producing up to 1.6 times the DC energy consumption as seen in Table VI — and, as shown in Figure 4, the grid energy consumption of these DCs is very low.

The results for Paris and São Paulo show the carbon footprint of ESDs compared to grid electricity. There is a small benefit in São Paulo for intensive usage, so with reduced sizing, there is no benefit in Paris to using batteries. The PV sizing on these DCs is reduced, probably due to the fact that little energy can be stored in case of overproduction.

The detail of hourly electricity consumption is highlighted in Figure 5. The workload is allocated in DCs with PV production. If all this production is used, or the corresponding DCs are full, then the allocation is driven by the battery state of charge, and when none of these possibilities are available, the allocation is for the DC with the lowest grid electricity footprint. For example, in the last hours, the electricity consumption of the different DCs is furnished by battery discharge, in the limit of a state of charge, and the remaining is allocated in the DCs of Paris, São Paulo, and Virginia. Thus, the DC of Virginia consumes grid electricity in two cases: either when Paris and São Paulo DC are full (from hours 10 to 24), or when the DC is empty and only local electricity can be used (hours 3, 5, and 6 in Figure 5). The follow-the-sun approach can be partially observed between hour 7

and 8, when Seoul PV production fall and the workload is transferred to Paris, with grid consumption. Then, at hour 10, the same append between PV production in Singapore and grid electricity in São Paulo and at hour 11 between Pune and Virginia. The figure also shows the impact of location, season, and PV sizing on the solar production between Pune and Canberra, large PV production in the best hours, and the tiny production in Paris.

Table V presents the impact of the different scenarios for energy sources on the load of the different DCs. In the first scenario, the workload is only allocated based on the grid electricity footprint. Thus, we could expect the workload order to be the same as the footprint order per kWh. However, the consumption does not only depend on the workload but also the PUE of the different DCs. We can thus observe a higher workload in Dubai compared to Singapore, considering that Dubai has electricity with a slightly higher footprint but the lowest PUE. Globally, the range of values highlights the workload variations, requesting at least 4 DCs, at most 8, and most of the time 7. The second scenario considers a model without grid electricity. The allocation is surprisingly distinct from the solar irradiation of the different DCs. For example, the DC of Paris has the lowest yearly irradiation but the median workload in this scenario. Its workload is higher than the one in Johannesburg, which has the second-highest yearly irradiation and is on a similar longitude. The workload is thus not only driven by yearly irradiation. The extremely low PV production in Paris during winter, associated with the static part of the electricity consumption in each data center, implies a high sizing of PV and battery, which lead to a high production during the other seasons that permit a large workload. On Johannesburg, the seasonal variation is lower, so static constraints do not drive PV sizing. Another surprising result is that the DCs with the lowest workload in this scenario are the 4 in the southern hemisphere (including Singapore). This contradicts the intuition of "follow-the-summer" allocation. The case of São Paulo and Canberra could be similar to the one of Johannesburg with the value of the minimal daily production in Virginia and Seoul. The largest workload concerns DCs with the more stable production (Dubai and Pune) and the lowest minimum daily production (Virginia, Paris, and Seoul). Finally, for the last complete scenario, the DCs with the largest workload are the 3 with the largest irradiation (Dubai, Pune, and Johannesburg), followed by Paris with the lowest grid electricity footprint. The only surprise is the workload of São Paulo, which is low considering its low grid electricity footprint and high solar irradiation, and the workload of Singapore, which is high considering its low PV production. Concerning São Paulo, this is probably because it has the second-lowest grid footprint. This implies a low battery sizing, thus a low PV sizing, and finally, it mainly receives workload only when no more DC can provide electricity from PV or battery discharge, and when the DC of Paris is full, that makes many constraints. Considering Singapore, it is probably due to its position close to the equator, which implies no "winter" season, and its large PV sizing. Finally, the reduction of carbon footprint of each DC between the complete scenario (PV + bat + grid) and the scenario with only grid electricity is showed in Table VI. It shows a small decrease in Paris and São Paulo, and a large decrease in the other locations, correlated to the electricity footprint.

VII. Conclusion

In this paper, we tackled the problem of greening a distributed cloud data center (DC) federation to lower its carbon footprint. The IT part of the cloud platform already exists, and the idea is to add the equipment on site to introduce renewable energy into the brown energy from the classical grid into the power supply of the DCs. Since the sun is shining everywhere on earth, we have proposed photovoltaic panels (PVs) to produce renewable energy and batteries as storage devices to mitigate the intrinsic intermittency of this energy during the day. The question is how to size the PV array and associated battery size, given an existing federation of DCs distributed around the earth. We have provided a formulation of the problem as a linear program. The particularity of our formulation is that we do not need integer variables; a solution is possible using only real variables given our objective and the context of the problem. As a result, the linear program allows to optimally solve large problem sizes, e.g., minimize the carbon footprint of a nine-site federation, each with its own weather conditions, upon a one-year horizon, hour by hour. We have demonstrated that our program is able to calculate the optimal sizing for PVs and batteries in just a few minutes. Numerous experiments have brought forward results that we have analyzed and discussed to explain what these results express. As an example, an interesting result, depending on the DC locations considered, the optimal solution to reduce the carbon footprint is a hybrid configuration between using PVs and the regular electrical grid. Moreover, batteries are not always mandatory in each location. Finally, our model has the flexibility to be extended to assess other scenarios (more DCs, other locations, values for carbon emissions, or workloads) and it may help decision-makers build their strategy to reduce the environmental impact of the cloud operation.

In future work, we plan to propose a sizing process that also includes the IT part. Since this investment has been made for years, another perspective is to introduce uncertainty into this sizing process to obtain a more robust distributed DC platform that can provide satisfying service to clients even if the weather conditions change and the submitted workload evolves. The goal always being to remain as virtuous as possible.

Acknowledgement

This work has been partially supported by the LabEx PERSYVAL-Lab ("ANR-11-LABX-0025-01") funded by the French program *Investissement d'avenir*, by grant #2021/06867-2, São Paulo Research Foundation (FAPESP), by the EIPHI Graduate school (contract "ANR-17-EURE-0002"), by the EuroHPC EU Regale project (g.a. 956560), and by the ANR DATAZERO2 (contract "ANR-19-CE25-0016") project.

REFERENCES

[1] IEA, "Data centres and data transmission networks," IEA, Paris, Tech. Rep., 2022. [Online]. Available: https://www.iea.org/reports/data-centres-and-data-transmission-networks

[2] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020.

[3] M. Koot and F. Wijnhoven, "Usage impact on data center electricity needs: A system dynamic forecasting model," *Applied Energy*, vol. 291, p. 116798, 2021.

[4] Greenpeace, "Clicking Clean," Greenpeace International, 2017. [Online]. Available: www.greenpeace.org/international/publication/6826/clicking-clean-2017/

[5] IPCC, *Summary for Policymakers*. Cambridge, UK: Cambridge University Press, 2022, p. In Press, in Press.

[6] N. Y. Krakauer and D. S. Cohan, "Interannual variability and seasonal predictability of wind and solar resources," *Resources*, vol. 6, no. 3, 2017. [Online]. Available: https://www.mdpi.com/2079-9276/6/3/29

[7] D. Nafus, E. M. Schooler, and K. A. Burch, "Carbon-responsive computing: Changing the nexus between energy and computing," *Energies*, vol. 14, no. 21, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/21/6917

[8] J. Shuja, A. Gani, S. Shamshirband, R. W. Ahmad, and K. Bilal, "Sustainable cloud data centers: a survey of enabling techniques and technologies," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 195–214, 2016.

[9] R. Padma Priya and D. Rekha, "Sustainability modelling and green energy optimisation in microgrid powered distributed FogMicroData-Centers in rural area," *Wireless Networks*, vol. 27, no. 8, pp. 5519–5532, 2021.

[10] H. Niaz, M. H. Shams, M. Zarei, and J. J. Liu, "Leveraging renewable oversupply using a chance-constrained optimization approach for a sustainable datacenter and hydrogen refueling station: Case study of california," *Journal of Power Sources*, vol. 540, 2022.

[11] M. Richter, P. Lombardi, B. Arendarski, A. Naumann, A. Hoepfner, P. Komarnicki, and A. Pantaleo, "A vision for energy decarbonization: Planning sustainable tertiary sites as net-zero energy systems," *Energies*, vol. 14, no. 17, 2021.

[12] M. Haddad, G. Da Costa, J.-M. Nicod, M.-C. Péra, J.-M. Pierson, V. Rehn-Sonigo, P. Stolf, and C. Varnier, "Combined it and power supply infrastructure sizing for standalone green data centers," *Sustainable Computing: Informatics and Systems*, vol. 30, p. 100505, 2021.

[13] M. Benaissa, J.-M. Nicod, and G. Da Costa, "Standalone data-center sizing combating the over-provisioning of the it and electrical parts," in *Proceedings of the Workshop on Cloud Computing (WCC)*, 2022.

[14] E. Sheme, S. Lafond, D. Minarolli, E. K. Meçe, and S. Holmbacka, "Battery size impact in green coverage of datacenters powered by renewable energy: A latitude comparison," in *Advances in Internet, Data & Web Technologies*, L. Barolli, F. Xhafa, N. Javaid, E. Spaho, and V. Kolici, Eds. Cham: Springer International Publishing, 2018, pp. 548–559.

[15] J. Song, P. Zhu, Y. Zhang, and G. Yu, "Versatility or validity: A comprehensive review on simulation of datacenters powered by renewable energy mix," *Future Generation Computer Systems*, vol. 136, pp. 326–341, 2022.

[16] B. Acun, B. Lee, F. Kazhamiaka, K. Maeng, M. Chakkaravarthy, U. Gupta, D. Brooks, and C.-J. Wu, "Carbon explorer: A holistic approach for designing carbon aware datacenters," *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, 2023.

[17] E. Ahvar, A.-C. Orgerie, and A. Lebre, "Estimating energy consumption of cloud, fog, and edge computing infrastructures," *IEEE Transactions on Sustainable Computing*, vol. 7, no. 2, pp. 277–288, 2022.

[18] H. Hlavacs, G. Da Costa, and J.-M. Pierson, "Energy Consumption of Residential and Professional Switches," in *2009 International Conference on Computational Science and Engineering*, vol. 1, 2009, pp. 240–246.

[19] D. Wang, C. Ren, A. Sivasubramaniam, B. Urgaonkar, and H. Fathy, "Energy Storage in Datacenters: What, Where, and How Much?" in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '12, NY, USA, 2012, p. 187–198.

[20] M. Haddad, J.-M. Nicod, M.-C. Péra, and C. Varnier, "Stand-alone renewable power system scheduling for a green data center using integer linear programming," *Journal of Scheduling*, vol. 24, pp. 523 – 541, 2021.

[21] J. Roach, "Microsoft's virtual datacenter grounds "the cloud" in reality," *Microsoft Innovation Stories*, 2021. [Online]. Available: https://news.microsoft.com/innovation-stories/microsofts-virtual-datacenter-grounds-the-cloud-in-reality/

[22] N. Walsh, "How microsoft measures datacenter water and energy use to improve azure cloud sustainability," *Microsoft Azure Blog*, 2022. [Online]. Available: https://azure.microsoft.com/en-us/blog/how-microsoft-measures-datacenter-water-and-energy-use-to-improve-azure-cloud-sustainability/

[23] G. Da Costa, L. Grange, and I. de Courchelle, "Modeling, classifying and generating large-scale google-like workload," *Sustainable Computing: Informatics and Systems*, vol. 19, pp. 305–314, 2018.

[24] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A. M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J. E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S. D. Schubert, M. Sienkiewicz, and B. Zhao, "The modern-era retrospective analysis for research and applications, version 2 (merra-2)," *Journal of Climate*, vol. 30, no. 14, pp. 5419 – 5454, 2017.

[25] D. Yue, F. You, and S. B. Darling, "Domestic and overseas manufacturing scenarios of silicon-based photovoltaics: Life cycle energy and environmental comparative analysis," *Solar Energy*, vol. 105, pp. 669–678, 2014.

[26] V. D. Reddy, B. Setz, G. S. V. R. K. Rao, G. R. Gangadharan, and M. Aiello, "Metrics for sustainable data centers," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 3, pp. 290–303, 2017.

APPENDIX A
ARTIFACT DESCRIPTION: OPTIMAL SIZING OF A GLOBALLY DISTRIBUTED LOW CARBON CLOUD FEDERATION

A. Abstract

This document details the necessary steps and instructions for the reader to be able to: i) install the programs and all the necessary dependencies to run the Linear Program (LP) of the experiments; ii) extract all the data used for generating the figures and tables; iii) generate the figures and tables; iv) create custom scenarios for their custom experiments.

B. Description

1) Check-list (artifact meta information):

- **Program**: (i) The LP modeling and execution; (ii) Scripts to extract data from the LP solution; (iii) Scripts to generate the figures and data for the tables;
- **Data set**: Solar irradiation data, workload, the LP parameters;
- **Hardware**: Various x86 or x64 CPUs;
- **Experiment workflow**:
  1) Install, and execute experiments;
  2) Extract data from the LP solution;
  3) Generate figures and data for the tables.
- **Output:** (i) Data extracted from the LP solution file (ii) Figures and data for the tables;
- **Experiment customization**: See below;
- **Publicly available?**: Yes.

2) How the artifact can be obtained: The Persistent ID of the artifact is:

https://doi.org/10.25666/dataubfc-2023-02-03

and to ensure that the code will be available for the long term and that the reader will use the same exact version to reproduce the experiments, we stored the code, all the necessary input files and parameters, and the instructions on Software Heritage:

https://archive.softwareheritage.org/swh:1:rev:44d2de89057ff6df4657769e0b14ac2bade57830;origin=https://gitlab.com/migvasc/lowcarboncloud

Finally, all source material can also be downloaded from the git repository associated with the Persistent ID:

https://gitlab.com/migvasc/lowcarboncloud

The artifact is structured by the following main directories:

- `input`: It contains the data that will be used as input for the LP (solar irradiation values, parameters, workload data);
- `script`: contains all the source code that will run the LP, extract the results and generate the plots;
- `results`: It contains the plots (in PDF format) and the data files (CVS file in text format) used to generate them, extracted data from the LP solution and data used to generate the tables presented in the paper.

3) Hardware requirements: Any modern x86 or x64 CPU is appropriate to execute the experiments. The experiments may be executed in parallel in order to reduce the execution time. If you want to run the experiments in parallel, it is recommended that your system have at least 16 GB of RAM (running all the experiments in parallel consumes around 10 GB of RAM). If your system does not meet the minimum requirements for running in parallel, you may still run the experiments in a sequential way. For running the sequential version of the experiments, your system needs at least 4GB of RAM (running each experiment consumes around 2 GB of RAM).

4) Software requirements:

- **Git** (week requirement): a version control system that is being used for storing all the necessary materials to run and analyze the experiments: the code of the simulations, input files, and scripts for extracting the results;
- **Nix** (weak requirement but strongly recommended): a multi-platform packet manager (run in Linux, Windows with WSL, and MacOS) that allows configuring the experiments in a practical way and allows for reproducibility.

If the reader does not want to install Nix, it is recommended to use a Linux distribution (preferably Ubuntu or Debian). Furthermore, it is necessary to install the following programs and packages/libraries:

- Python 3 (version 3.10.8): Used for modeling and executing the LP, and extracting data. Necessary libraries and their versions: PuLP (2.7.0), and argparse (1.1.0);
- R (version 4.1.2): Used for generating the plots. Necessary packages and their versions: tidyverse (1.3.1), gridExtra (2.3.0), patchwork (1.1.1), viridis (0.6.2), stringr (1.4.0), rjson (0.2.20), rlist (0.4.6.2).

Finally, if the reader does not want to install Git, the artifact can also be downloaded from Software Heritage, as stated before.

5) Datasets: All the necessary data sets to execute the experiments are available in the `input` directory. The solar irradiation data was collected from the MERRA-2 service. The main input file is in the `.json` format, containing all the LP parameters and inputs. Regarding the workload file, it is a CSV file in text format, where for each time slot, there is the total CPU cores demand value. More details about these input files are presented in Section *Experiment customization*.

6) Installation: Dowload the artifact (either from Software Heritage or by cloning the Git repository). Installing Nix is optional, however, if the reader decides not to use Nix, it is necessary to install Python, R, and all the packages described in the software requirement section. In the Nix option, this dependency installation was automated. After installing all the necessary dependencies, and if you wish to use Git, you can clone the Git repository to your local computer with the following command: `git clone https://gitlab.com/migvasc/lowcarboncloud` and then enter in the cloned/downloaded repository.

7) Experiment workflow: There are two ways to execute the experiments: in parallel or sequentially. In the first, the total execution time of the experiments will be shorter, however, it has higher hardware requirements (as detailed in the section **Hardware requirements**).

If you are using Nix and want to execute the experiments in parallel, you must use the following command: `bash ./scripts/workflow_parallel_nix.sh` and if you want to run the experiments in a sequential way, execute: `bash ./scripts/workflow_sequential_nix.sh`. This command will build a nix environment with all the necessary dependencies to reproduce the experiments and execute them.

If you are not using Nix, and you want to execute the experiments in parallel, you must use the following command: `bash ./scripts/workflow_parallel.sh` and if you want to run the experiments in a sequential way, execute: `bash ./scripts/workflow_sequential.sh`.

These script files (`workflow_parallel` and `workflow_sequential`) automate the execution of the experiments and have 3 main steps. In the first step, the LP will be executed (using the `low_carbon_cloud.py` program) for each scenario (the years 2018, 2019, 2020, 2021, and the two extra scenarios of only using the grid and only using PVs and batteries). In the second step, after the execution of the LPs, the data for generating the plots and the tables will be extracted (files `extract_data_figures.py`, `extract_table_v_data.py`, `extract_table_viii_data.py`, and `extract_table_vi_vii_data.py`). Finally, in the last step, the figures will be generated using the R script file `plots.r`.

The total time it takes to run the entire workflow may vary depending on the machine's hardware configuration where the experiments will execute. For example, on our test machine equipped with an Intel core i9-11950H CPU and 32 GB of RAM, it took about 1 hour for the parallel version, and about 6 hours for the sequential version.

This execution time is different from the one reported in the paper because, in this artifact, we are using another solver: the CBC solver (Pulp's default solver). In the paper, we used Gurobi, a commercial solver requiring a license (free license for academics), so we are providing this artifact using CBC to avoid costs for the reader who wants to reproduce the experiments.

8) Evaluation and expected result: Once the experiments are completed, a `results` directory will be created. In this directory, the reader will find:

- For each Table *y* present in the paper, there is a respective file `table_y_data.csv` that contains the data used for the table;
- Each scenario of the experiment has its own folder. For example, for the results for the year 2021 there is the folder `results/2021`, and within this folder are the following files:
  - `metrics.csv` contains the values for the metrics Green Energy Coefficient (GEC) and CO_2 Savings, and Data Center utilization (DCU);
  - `solution.csv` is an auxiliary file to store the values of the computed variables of the linear program;
  - `summary_results.csv` contains the values for carbon emissions, which input was used in the experiments, and the runtime;
  - For each Figure *x* present in the paper, there is a file `figure_x.pdf` with the plot, and `figure_x_data.csv` with the data used to generate the figure. In the paper, we only presented the figures for the year 2021, but the figures are generated for other years as well.

For all the Tables and most of the Figures, the results presented in the paper and obtained by reproducing the experiments will be the same or very similar.

The reason why some results may be different is that there are multiple alternative solutions for an LP. In other words, the optimal value of the LP regarding its objective function will be the same, but the variable values might differ. In our case, the computed optimal value for the PV surface area and the capacity of batteries is the same. The difference might occur in the variables regarding how much power to charge/discharge from the batteries, use from the grid, or workload to execute at a specific data center at a specific time slot.

We provide a script `test_output.py` that will compare the results of the experiment execution with the expected results to automate the validation process. The validation will consider the sizing and the total carbon emissions. If the validation succeeds, the following messages will be shown: *"Sizing results validated!! The value obtained is equal to the expected result."* and *"Total emissions results validated!! The value obtained is equal to the expected result."*. To do the validation process, the reader may also compare the output with the tables and figures in the paper and the folder `expected_results`.

9) Experiment customization: This section describes how other scenarios can be executed using the present artifact. Examples of possible scenarios: using other locations for the data centers, workloads, and carbon footprint values for PVs, batteries, and the grid.

The starting point to generate your own scenario is creating a `.json` file that will describe all the parameters and inputs necessary for the Linear Program. This file must be located inside the `input` folder. More details about all the necessary parameters, what they are, and their data type can be found in the `readme.md` file of the artifact.

After creating your input file, execute: `bash scripts/run_custom_scenario_nix.sh example.json` if you are using Nix, or execute `bash scripts/run_custom_scenario.sh example.json` otherwise.

Once the experiment execution is complete, all results will be in the respective results directory. In the previous example, this directory is named `results/example`.