



Integrating Pre-Cooling of Data Center operated with Renewable Energies

Maël Madon, Jean-Marc Pierson

► To cite this version:

Maël Madon, Jean-Marc Pierson. Integrating Pre-Cooling of Data Center operated with Renewable Energies. IEEE International Conference on Green Computing and Communications (GreenCom 2020), Nov 2020, Rhodes Island, Greece. hal-02970488

HAL Id: hal-02970488

<https://hal.science/hal-02970488>

Submitted on 18 Oct 2020

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.

Integrating Pre-Cooling of Data Center operated with Renewable Energies

Maël Madon

Ecole Polytechnique

Palaiseau, France

mael.madon@polytechnique.org

Jean-Marc Pierson

Institut de Recherche en Informatique de Toulouse (IRIT)

Université de Toulouse, Toulouse, France

jean-marc.pierson@irit.fr

Abstract—Recently, integrating renewable energies in data centers has gained interest in the research community. Most researchers focus on the joint management of the IT and electrical infrastructures, few integrate cooling explicitly in their studies. One path that has been mostly forgotten so far is the potential for pre-cooling the data center when renewable energies are available. Recent works address this potential from an economical perspective to avoid peak prices, but none from an ecological perspective to reduce CO_2 emissions. In our work, we fill the gap by investigating to which extent reducing the temperature when renewable energies are available has a potential interest, and how it compares to other temperature control strategies. We build on a strong physical model of heat dispersion and a MILP formulation of the problem at hand, and we propose heuristics to handle the cooling device at best with renewable energy. Finally, we run experiments on real data traces (IT and renewable energy) to derive the conclusions that pre-cooling has indeed an interest for reducing carbon emissions.

Index Terms—Renewable energy; Cooling; Green data center; Pre-cooling

I. INTRODUCTION

As the climate situation is becoming more and more critical, environmental concerns are growing. The Information Technology (IT) sector is no exception. On the issue of data centers, which are large consumers of electricity, researchers have been studying energy efficiency for more than a decade. Authors surveyed energy efficient techniques for cloud data centers [1], [2], multi-tenant data centers [3], some focusing on cooling [4] to cite a few. According to the 2017 Greenpeace report [5], more and more IT companies are committed to a 100% renewable supply. The digital giants are ranked high in Greenpeace company scoreboard with respectively 83%, 67% and 56% clean energy index for Apple, Facebook and Google. These results are generally achieved by huge investments in green power farms or contracts with electricity providers. At the same time, the sector is witnessing the emergence of “green cloud solutions” (see for example [6], [7]).

The interest in including renewable energy for powering data centers has emerged in the recent years. Most of the works focus on the joint management of the IT workload and the electrical infrastructure (centralized in [8] and decentralized in [9] and [10]), some are investigating the benefit of managing

several data centers to balance the load according to the available energy [11]–[13] but few integrate cooling aspects (free cooling in [14] and [15]). However, cooling a data center can represent a large part of the energy expenditure, especially in small and not very optimized data centers. This extra power consumption is often considered as constant over time or proportional to IT load. We argue that cooling could be used to store renewable energy surplus into the data center’s thermal mass in order to anticipate later lack of production. What is the potential of this pre-cooling approach to cope with variability of renewable energies?

While some works exist on using pre-cooling to prevent overheating in time of cooling shutdown [16], only a couple of researches have been conducted combining renewable energy and cooling: [17] investigated three forms of energy storage (thermal tanks, UPS batteries and building thermal mass). [18] used a demand response mechanism to switch off the cooling device when electricity prices are high, while [19] exhibit the interest for data center in demand response program, using energy storage in batteries and fuel cells. The closest to our work is [20] that investigates several policies for tasks management (depending on the renewable energy availability), one being the pre-cooling of the building. This latter gave excellent results in terms of solar utilization and waiting time for tasks. However, they did not focus on the pre-cooling itself which was seen as a secondary aspect in their work.

The main contributions of this work are:

- to simulate the data center temperature evolution based on a thermal model and real data traces for weather and IT load
- to test heuristics for the operation of the cooling device that optimizes the carbon emission according to the availability of renewable energy
- to use a linear solving formulation of the problem to find optimal solutions under maximum and minimum temperature constraints and conclude on the potential of pre-cooling

The rest of this paper is organized as follows: Section II gives related works. Section III details the data center modeling including its elements, the thermal model and the

cooling model. Section IV provides heuristics for the pre-cooling approach. Section V gives the experimental setup while Section VI presents the results of the experiments and discusses them. Section VII concludes the paper and provides perspectives.

II. RELATED WORK

A. Data centers supplied with renewable energies

Many works have integrated co-located renewable sources in the power supply system of data centers. As always with renewable energies, the difficulty stands in the management of the variability of production. To handle this, data centers have a great potential: IT loads can be scheduled in a smart way or even be migrated to another data center located in an area where production is better. Goiri et al. [8] made a pioneer work. They built Parasol, a real prototype green data center powered with solar panels and grid as well as storage units. They developed GreenSwitch, a dynamic scheduler and source selector. GreenSwitch used a formulation as a Mixed Integer Linear Programming problem integrating both workload scheduling and energy management constraints. The savings were from 40% to 100% energy from the grid, depending on the weather conditions. GreenSwitch achieves it by moving deferrable IT tasks, charging the batteries with the power surplus and discharging them when production is low. In [19] Oleksiak et al. have a similar path, adding fuel cells storage to the system and building their management policy with a demand response program by shifting IT load. In the DATAZERO project [21], Grange et al. [9] and Caux et al. [10] proposed more decentralized approaches where tasks scheduling is treated apart from power management. They consider a utility function given by an independent power decision module and test scheduling heuristics for IT tasks with due dates. Grange et al. achieved a reduction of grid consumption up to 49%. [10] and [22] propose to degrade application performances according to the availability of renewable sources.

Liu et al. [14] took into account renewable supply, dynamic pricing and IT workload planning as well as cooling. They considered free cooling, consisting in blowing outside air into the building. A concurrent approach is the one of Habibi Khalaj et al. [15]. They studied 42 locations to assess their production and free cooling potential and gave results on optimal sizing for hybrid power infrastructure. Other works study geographical allocation of tasks on data center having their own on-site or off-site renewable sources [11]–[13], [23].

All these works witness for the potential and feasibility in integrating renewable energy for powering datacenters. Nowadays, renewable energy powered datacenters is already a feasible technological and economical path.

B. Pre-cooling data centers

To the best of our knowledge, only few researches have been done on the pre-cooling of data centers for saving energy. Most

of earlier works focused on the potential of pre-cooling to limit the impact of cooling shutdown, following Khankari [16] who studied the potential of the thermal mass of the data center room in case of shutdown, with respect to the room configuration, the IT density, the enclosing racks arrangement, and the threshold temperatures of the servers. Zhang et al. [17] present TEstore, a cooling strategy exploiting thermal and energy storage to cut energy bills in data centers. They assume a varying price for electrical power during the day and study three forms of energy storage: (i) ice or water-based thermal tanks, (ii) UPS batteries, (iii) building thermal mass. When power price is low, energy can be used to charge the batteries, store cold in thermal tanks or pre-cool the building. When power price is high, the energy stored is used to cool the infrastructure. Their results show that almost 85% of the cost saving is achieved by exploiting the thermal tanks on the long timescale. UPS batteries can be used in the middle term and pre-cooling potential in their model is relatively low: it represents a few percents of the total saving, with only the ability to pre-cool a few minutes before the peak price.

However, Lukawski et al. claim in [18] better results for pre-cooling. They use pre-cooling as a demand response mechanism for reducing coincident peak loads in the power market, i.e. they cut off the cooling device of the data center to save energy during a 15 minutes peak price. The model was tested and validated experimentally. The reasons for the different results with TEstore are twofold: (i) the data center considered has a low floor utilization (below 500W/m²) and (ii) it has no server air containment system. Thanks to this, the thermal mass of the room is large enough related to the heat emitted to provide a thermal buffer allowing extended demand response time. It is not the case in high density facilities. As a result, the candidates for using a pre-cooling mechanism are small data centers or computer rooms.

Zhang et al. and Lukawski et al. don't deal with renewable energies but rather with financial considerations. Recently, Li et al. covered synergies between renewable production and pre-cooling possibilities [20] with a holistic thermal-aware workload and cooling management for the maximization of renewable energy sources. They consider batch (deferrable) jobs and interactive (non-deferrable) jobs. They compare four methods: (i) tasks are executed as soon as they arrive; (ii) load balancing distribution over time where batch jobs are distributed evenly over multiple time slots; (iii) best effort where the batch jobs are scheduled according to the predicted power generation; (iv) thermal-aware workload management where cooling power consumption is additionally taken into account. In the last strategy, if IT doesn't consume everything, renewable energy surplus is used to cool the room. Pre-cooling plays here again a secondary role. The results of their comparison is that they achieve to reach more than 98% solar utilization and less average waiting time with the last strategy for batch jobs than with the second and third strategy.

LIST OF SYMBOLS

η_{pv}	efficiency of PV panels [-]
a	thermal inertia of the building [-]
a_C, b_C	heat pump parameters [-]
a_{COP}, b_{COP}	COP temperature dependency parameters [-]
A_{pv}	area of PV panels [m ²]
C	thermal capacitance of the building [J/°C]
COP	heat pump coefficient of performance [-]
COP_{nom}	heat pump coef. of perf. at nominal conditions [-]
h	time step [s]
I	solar irradiation [W/m ²]
k	time [s]
P_{cons}	data center power consumption [W]
P_{cool}	power consumption of the cooling device [W]
P_{grid}	power drawn from the grid [W]
P_{IT}	power consumption of IT equipment [W]
P_{prod}	PV production [W]
Q_{cool}	heat dissipated by the cooling device [W]
Q_{IT}	heat generated by IT [W]
Q_{nom}	heat pump nominal cooling capacity at state 1 [W]
R	thermal resistance of the walls [°C/W]
r	heat pump state [-]
T	room temperature [°C]
T_{ext}	ambient temperature [°C]
$T_{min}, T_{min}, T_{max}$	thermostatic control temperature [°C]

However, this strategy leads to temperature lower than the recommended ASHRAE temperatures [24] and the goal of maximizing renewable power utilization at any cost could be discussed.

Compared to the existing state of the art, our work isolates only the effect of pre-cooling, respecting the recommended operating temperature, allowing selling surplus of energy on the grid, and opening the door to a more accurate usage of it in other strategies.

III. DATA CENTER MODELING

In this Section, the physical model for the data center is described. Time is discretized into time steps of length h , and we note the variables as functions of the discrete time k .

A. Description of the model components

Following the results from [18] which showed the interest of pre-cooling from an economic perspective, we consider a low-density data center with co-located photo-voltaic (PV) production. The building is typical of a small telecommunication data center, a computer room or internal servers for a company. The power supply system is composed of a PV source delivering at time k a power P_{prod}^k (in Watt) and access to the power grid P_{grid}^k . We relate the power produced by the PV panels to meteorological data using a widely used model [25]–[29] described in Equation (1):

$$P_{prod}^k = A_{pv} \times \eta_{pv} \times I^k \quad (1)$$

where I^k is the solar irradiation (in W/m²) at time k , A_{pv} the area of the panels (m²) and η_{pv} their efficiency.

This electricity is used to power the IT equipment and a cooling system, denoted by P_{IT}^k and P_{cool}^k respectively. Other energy usage such as lighting, power distribution units, backup

batteries or auxiliary facilities is ignored in this model. Thus, if we denote by P_{cons}^k the total power of the data center, we have:

$$P_{cons}^k = P_{IT}^k + P_{cool}^k \quad (2)$$

As a cooling device, we studied as in [18] a geothermal heat pump. This system is more efficient than a conventional air-cooled computer room air conditioning (CRAC) as the heat sink in the ground stays at a more stable temperature than the outdoor air. Indeed, ground temperature remains around 13°C throughout the year, well below the recommended maximum temperature for data centers. Please note that other cooling system consisting in room air circulation could be used in the analysis.

Electricity from the grid is used only when necessary, i.e. when $P_{prod}^k < P_{cons}^k$. In general, we have:

$$P_{cons}^k = P_{prod}^k + P_{grid}^k \quad (3)$$

where P_{grid}^k is the power drawn from the electrical grid at time k .

B. Thermal model

To simulate the thermal behavior of the data center building, we adopted a discrete model used in similar works [18], [30]. The system we study is the whole building considered as a homogeneous thermal mass isolated from other buildings. The heat transfers we take into account are the following:

- the IT thermal load Q_{IT}^k at time k
- the heat dissipated by the cooling device Q_{cool}^k at time k
- the heat transfers through the walls of the building

From the second principle of thermodynamics, the temperature follows Equation (4). It has been tested numerically and experimentally by Lukawski et al. in [18] for data centers.

$$T^{k+1} = a.T^k + (1 - a)[T_{ext}^k + R(Q_{IT}^k - Q_{cool}^k)] \quad (4)$$

where T^k is the room temperature at time k , T_{ext}^k the ambient temperature at time k , R the thermal resistance of the walls (in °C/W) and a a dimensionless parameter capturing the thermal inertia of the building.

$$a = e^{-\frac{h}{R.C}} \quad (5)$$

with C the thermal capacitance (in J/°C).

According to [31], heat dissipation from IT can be approximated by its power consumption, as the power transmitted by computing or other information technology equipment through the data lines is negligible:

$$Q_{IT}^k = P_{IT}^k \quad (6)$$

C. Cooling system

Once more, the cooling model has been adopted from Lukawski et al. and is validated in [32]. The data center acts as a Thermostatically Controlled Load (TCL): the heat pump maintains the temperature within two targets T_{min} and T_{max} by switching on and off. More precisely, the regulation is done thanks to three available states for the pump: off, nominal rate or double rate. Heat dissipated by the cooling device can be expressed as follows:

$$Q_{cool}^k = r^k \times Q_{nom} \times (a_C \cdot T^k + b_C) \quad (7)$$

where Q_{nom} , a_C and b_C are parameters depending on the heat pump model and r^k is the variable representing the state of the pump at time k ($r^k \in \{0, 1, 2\}$).

We name r_0^k the state of the heat pump when using the TCL policy. The thermostatic control is then expressed by:

$$r_0^{k+1} = \begin{cases} 0 & \text{if } T^k \leq T_{min} \\ 2 & \text{if } T^k \geq T_{max} \\ 1 & \text{if } T_{min} < T^k < T_{med} \text{ \& } r_0^k = 2 \\ 1 & \text{if } T_{med} < T^k < T_{max} \text{ \& } r_0^k = 0 \\ r_0^k & \text{else} \end{cases} \quad (8)$$

where T_{min} is the minimal temperature in the room, T_{max} the maximal temperature and T_{med} the threshold temperature allowing where the pump runs at nominal rate, for a smooth transition between extreme rates (off and double rate).

Finally, power consumption of the cooling system can be computed using the coefficient of performance (COP) with the Equations (9) and (10):

$$P_{cool}^k = \frac{Q_{cool}^k}{COP^k} \quad (9)$$

$$COP^k = COP_{nom} \times (a_{COP} \cdot T^k + b_{COP}) \quad (10)$$

where COP_{nom} , a_{COP} and b_{COP} are here again specific to the pump.

IV. APPROACH TO SIMULATE THE PRE-COOLING

We developed in Python a simulation of the data center described in Section III. This entirely modular program allowed us to try different approaches with different data in order to assess the potential and limits of pre-cooling.

The program takes a time window as entry along with the IT load and weather conditions (ambient temperature, solar irradiation, wind speed) over this period. The program outputs the evolution over time of the two computed variables, i.e. room temperature and cooling device power consumption.

A. Thermostatically Controlled Load (TCL)

First, the thermal behavior of the data center over time was simulated according to the equations previously described. By default, the algorithm has nothing to decide: indoor temperature and heat pump state are determined at each iteration based

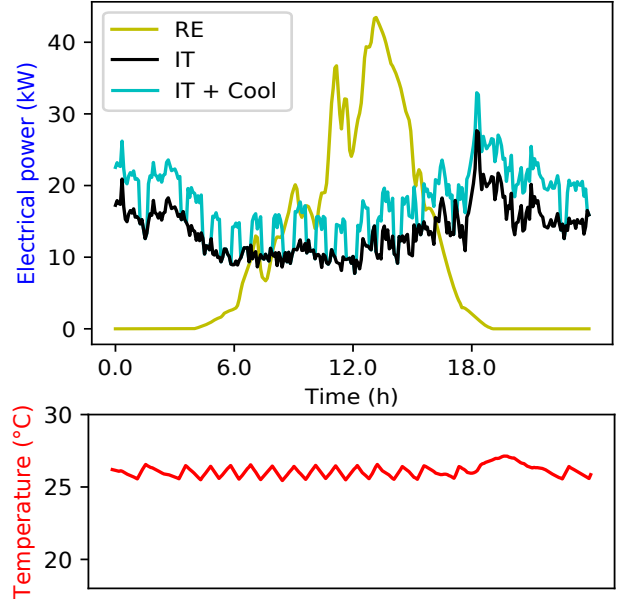


Fig. 1. Standard output of the program for one day. Room temperature in red is thermostatically controlled. Renewable energy production (RE) is in yellow. IT load is in black and cooling power (Cool) is plotted above in blue.

on the values of the model variables at the previous iteration (Equations (4) and (8)).

Figure 1 exhibits for illustration a standard output where the cooling strategy is the basic TCL. In this example the temperature can be maintained under the T_{max} limit only by switching the heat pump between state 0 ($r_0^k = 0$, no cooling) and state 1 ($r_0^k = 1$). The second state of the heat pump is in this case never used. Also, the shape of the room temperature curve helps us to visualize the thermal inertia of the system. Depending on the ambient temperature, it takes about one or two hours to make the temperature drop by one degree with cooling device at state 1 and about a quarter to a tenth of this time to regain it with cooling device off. It means that our building has a rather large thermal inertia which makes it possible to store energy in the thermal mass. This is what we want to exploit with pre-cooling. We will see in Section VI that this thermal mass storage capacity is closely linked to the density of IT equipment per square meter.

B. Naive heuristic

The pre-cooling potential stands in the period where production is greater than consumption, i.e. where the renewable production is above the consumption. We could cool down more the building at that time to store electricity surplus in form of cold. A simple heuristic for the cooling control is the following: as long as renewable production is in surplus, the heuristic makes the maximum use of this energy by increasing the heat pump state. This cooling strategy can be expressed

as:

$$r^k = \begin{cases} r_0^k & \text{if } P_{prod}^k < P_{IT}^k + P_n \\ 1 & \text{if } P_{IT}^k + P_n \leq P_{prod}^k < P_{IT}^k + 2P_n \\ 2 & \text{if } P_{prod}^k \geq P_{IT}^k + 2P_n \end{cases} \quad (11)$$

with r_0^k being the standard cooling strategy presented in Equation (8) and P_n the power of one state of the heat pump (considered constant whatever the temperature) and $P_n = \frac{Q_{nom}}{CO P_{nom}}$.

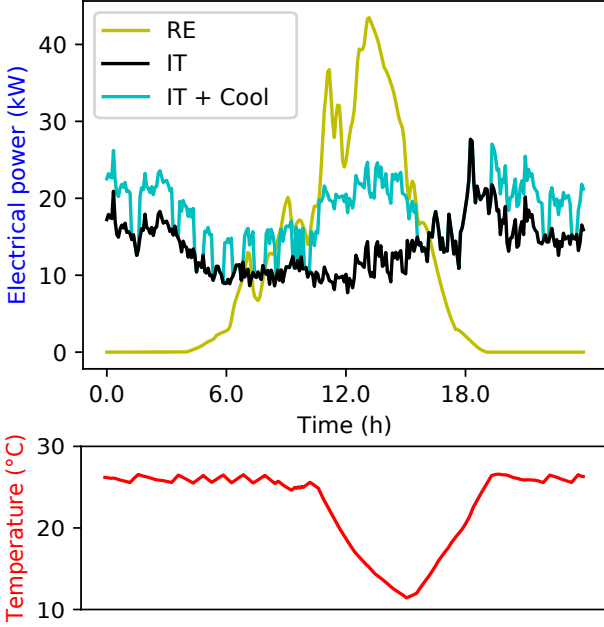


Fig. 2. Thermal behavior and power consumption of the data center with naive pre-cooling heuristic.

Figure 2 shows the results with the naive heuristic for the cooling control in the same conditions as in Figure 1. One can see that it resulted in the heat pump being completely switched off between 4PM and 7PM approximately, thus saving electricity from the grid. It has costed on the downside an overconsumption of the cooling system during the day. Overall, the algorithm trades a loss in the total energy used (the area under the consumption profile, in kWh) to the benefit of a better self-supply. However, it is easy to realize that this naive pre-cooling approach is not optimal. First, it can make the room temperature fall under the ASHRAE recommended temperature as it is the case in this example. Second, cooling at the maximum capacity as soon as energy is available is likely to be unnecessary to achieve the best results. The purpose of the following is therefore to find the best time to start pre-cooling or more generally the best pre-cooling policy.

C. Linear solving

As seen before, the problem of optimizing the data center cooling policy in order to maximize the direct use of renewable production has only one decision variable: the heat pump

state r^k at each time k . Modifying its value at a given time will impact the system during the rest of the simulation. Consequently, the best choice can only be made by taking into account all in one the evolution of the exogenous variables over the time window and every possible decision for r^k at each time k . As it is a non-trivial algorithmic problem, a linear solver has been used. This Subsection first explains the choice for the objective function then details the constraints of the model.

1) *Objective function:* The energy purchased from the grid at time k is equal to the discretized power from the grid P_{grid}^k times the time step h . The first objective function was set to minimize this grid energy:

$$\text{minimize } \sum_{k=0}^{K-1} h \cdot P_{grid}^k \quad (12)$$

If renewable production is in surplus at time k , no power has to be purchased from the grid. Else, grid power can be expressed as the gap between production (P_{prod}^k) and consumption (P_{cons}^k):

$$\begin{aligned} P_{grid}^k &= \max(0, P_{cons}^k - P_{prod}^k) \\ &= \max(0, P_{cool}^k + P_{IT}^k - P_{prod}^k) \end{aligned} \quad (13)$$

This simple objective (Equation 12) however failed to address earlier critics about the naive heuristic: the resulting behavior for the cooling system is similarly too greedy, sometimes overcooling or cooling unnecessarily. Also, the CO_2 emissions impact does not appear in the objective function.

Subsequently, our new approach proposed for the objective function is to introduce a price for the energy self-supplied and a different (higher) one for the electricity bought on the grid. This pricing can represent either the intensity of each of these kilowatt-hours in greenhouse gas emissions or their purchase and resale price on the grid. For simplicity, the terms purchase and resale prices are used in the following. Let p_{buy}^k and p_{sell}^k be those prices. Let's assume by common sense that $p_{buy}^k > p_{sell}^k$. An electron used locally is an electron that cannot be not sold. It is therefore considered as bought at price p_{sell}^k .

Thus comes the objective as:

$$\text{minimize } \sum_{k=0}^{K-1} h \cdot (p_{buy}^k \cdot P_{grid}^k + p_{sell}^k \cdot P_{self}^k) \quad (14)$$

where self-supplied power is denoted P_{self}^k :

$$P_{self}^k = \min(P_{cons}^k, P_{prod}^k) \quad (15)$$

With such an objective, solving the problem will really find a trade-off between pre-cooling and overconsumption while allowing to tune it with the two prices.

2) *Constraints from the physical model:* The constraints in the model are based on the variables and the equations presented in the modeling part (Section III). In order to take into account the problem as a whole, the system contains for each time k the equations and inequalities simulating the model:

$$T^k = \begin{cases} T_{init} & \text{if } k = 0 \\ a \cdot T^{k-1} + (1-a)[T_{ext}^{k-1} + R(Q_{IT}^{k-1} - Q_{cool}^{k-1})] & \text{otherwise} \end{cases} \quad (16)$$

$$Q_{cool}^k = r^k \times Q_{nom} \times (a_C \cdot T^k + b_C) \quad (17)$$

$$P_{cool}^k = \frac{Q_{cool}^k}{COP_{nom} \times (a_{COP} \cdot T^k + b_{COP})} \quad (18)$$

Among the equations presented above, Equations (17) and (18) are not linear. Equation (18) was linearized by neglecting the heat pump COP dependency on T_k . Indeed, thanks to the choice of a geothermal source, the slope of this dependency is relatively low. With our numerical parameters, the COP varies from $COP(18^\circ C) = 2.41$ to $COP(27^\circ C) = 3$. After the numerical resolution by the linear solver, and before analyzing the results and figures, this dependency is taken back in Equation (18) to correct the error introduced.

Equation (17) requires more work. To linearize it, we introduce two continuous variables Q_1^k and Q_2^k and two binary variables r_1^k and r_2^k . r^k is a trinary variable, it can be represented without loss of generality as $r^k = r_1^k + r_2^k$ with $r_2^k \leq r_1^k$. With these notations, Equation (17) becomes:

$$Q_{cool}^k = Q_1^k + Q_2^k = (r_1^k + r_2^k) \times X^k \quad (19)$$

where we introduce $X^k = Q_{nom} \times (a_C \cdot T^k + b_C)$ for simplification. Q_{cool}^k is the sum of two products of a binary variable and the continuous variable X^k . X^k is bounded below by zero and above by a constant M . We can use a classical trick of linear optimization to obtain for all k the set of constraints on the heat pump:

$$\left\{ \begin{array}{ll} Q_{cool}^k = Q_1^k + Q_2^k & \\ r_2^k \leq r_1^k & \\ 0 \leq Q_i^k \leq M & i = 1, 2 \\ Q_i^k \leq M \times r_i^k & i = 1, 2 \\ Q_i^k \leq X^k & i = 1, 2 \\ Q_i^k \geq X^k - M \times (1 - r_i^k) & i = 1, 2 \end{array} \right. \quad (20)$$

3) *Constraints on the cooling device:* In addition to the previous constraints extracted from the physical model, other constraints are added to reach the targeted behavior. Two operating modes are defined: (i) the “free mode” where the solver is let completely free to set the evolution of r^k and (ii) the “TCL mode” where the solver is forced into thermostatic control. While it might seem more appropriate to set the solver in free mode for all the duration of the simulation, we did not for two reasons. First, despite its parameterization, the Gurobi solver was not able to find the optimal solution in

a reasonable time for a 24-hour time window problem with a 5-minute time step. Note that in this configuration, the system has 4032 variables among which 1728 are binaries. Gurobi can find an almost optimal solution in a few seconds but struggles to demonstrate its optimality in more than 20 hours of calculation on a powerful machine. This phenomenon is probably due to the impossibility in our problem to cut branches at the root in the decision tree on the variable r^k , for the interdependence reasons mentioned page 5. The second reason is that the quasi-optimal solutions found by the solver seemed physically unrealistic: in order to save as much energy as possible, the cooling system ended out switching on and off at each time step to keep the temperature as close as possible to the maximum temperature, which would lead to premature cooling device aging. As a result, we chose to trigger free mode only if the production is sufficient:

$$\begin{array}{ll} \text{if } P_{prod}^k > P_{IT}^k & : \text{ free mode} \\ \text{else} & : \text{ TCL mode} \end{array} \quad (21)$$

In the following the constraints for both modes are given.

a) *Free mode:* In this mode the system is let as free as possible. Only upper and lower bounds for the room temperature are added, as in ASHRAE thermal recommendations for all classes of devices [24]: $T_{lo} = 18^\circ C$ and $T_{up} = 27^\circ C$. Hence, the set of constraints for all k :

$$T_{lo} \leq T^k \leq T_{up} \quad (22)$$

b) *TCL mode:* In this mode the constraints (22) are kept. New constraints are added, corresponding to the thermostatic control formalized in Equation (8). Here again, linearization tricks and four new binary variables are used to translate these equations into sixteen new constraints for the system. As it does not present any particular difficulty and for reasons of available space, they will not be presented here.

V. EXPERIMENTAL SETUP

In this Section, we describe the experimental choices and the numerical parameters used. Simulations were held on an 3.20GHz Intel Core i7-8700 processor with 12 cores and 16GB of RAM. The time step was set to $h = 300s$. The commercial Gurobi solver and its API for Python were used for the linear program solving. The simulations were stopped when either the gap between the current solution and the computed bound was less than 0.01% or a time limit of 1000s was reached. In practice, all our solutions were either exact or less than 0.02% away from the computed bound.

A. Data center sizing

a) *Thermal parameters:* The thermal parameters are taken from a model calibration performed in a small 93m² telecommunications data center located in Ithaca (NY) [18]. We reported the values of the parameters in Table I.

TABLE I
THERMAL PARAMETERS TAKEN FROM [18]

Symbol	Parameter	Value
P_{IT}	average IT load	14040 [W]
R	thermal resistance	$4.67 \cdot 10^{-3} [^{\circ}\text{C}\text{W}^{-1}]$
C	thermal capacitance	$15.76 \cdot 10^6 [\text{J}^{\circ}\text{C}^{-1}]$
Q_{nom}	heat pump nominal cooling capacity at state 1	15767 [W]
a_C, b_C	heat pump parameters	0.024, 0.361 [-]
COP_{nom}	heat pump coef. of perf. at nominal conditions	3 [-]
a_{COP}, b_{COP}	COP parameters	0.022, 0.406 [-]
$T_{min,med,max}^*$	thermostatic control temp	25.6, 26.4, 27.2 [$^{\circ}\text{C}$]

b) Power supply: The PV production is modeled in Equation (1). The constants were chosen to allow excess production during the peak period and thus give a margin for pre-cooling: $A_{pv} \times \eta_{pv} = 50\text{m}^2$. It corresponds to $A_{pv} = 300\text{m}^2$ of PV panels with efficiency $\eta_{pv} = 0.17$ for a production of 55kW peak (considering a maximum solar irradiation of $1100\text{W}/\text{m}^2$, standard for Europe and in the data set used in the experiments).

B. External data

a) Meteorological data: In order to obtain realistic data for renewable production, solar irradiation records were downloaded from Solar Radiation Data (SODA) website [33]. The data set also features ambient temperature. It covers the year 2006 in the city of Belfort (France), with a recording time step of one minute. Figure 3 shows the minimal, median and maximal day of production for each season in the data.

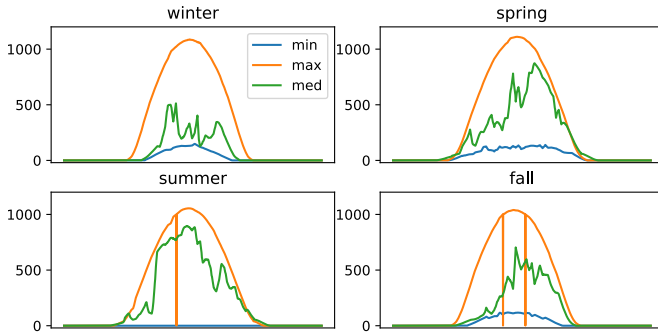


Fig. 3. Minimal, median and maximal day of solar irradiation for each season in Belfort (France), 2006. Each plot corresponds to 24h and the unit is W/m^2

b) IT workload: To model the data center IT load over time, we made use of two different data sets: a generated synthetic IT trace and a real cluster trace. The traces can be seen in Figure 4.

The generated trace comes from a workload generator developed by Da Costa et al.. It is based on the analysis of the Google Cluster Workload Traces and its implementation is described in [34]. The generator is a Python script which takes as input a parameter d representing the average number

of tasks processed by the data center each hour. Thanks to a probabilistic model, it outputs a trace of n points similar to a standard Google trace, each point indicating the number of tasks processed at during each time step. A low value for d was chosen, to imitate a small data center with a low task arrival rate. We assumed that each task requires the same amount of power to run and thus normalized the output to have an average power of $\overline{P_{IT}}$, $\overline{P_{IT}}$ being a constant chosen to match the sizing of the model.

The second data set is a real trace released by Alibaba in 2018 and recorded from 4000 of its servers¹. Here again the data was normalized to $\overline{P_{IT}}$ by identifying CPU load as thermal load. One can notice on Figure 4 the variability of the data during the day: CPU usage reaches a peak around 6PM while staying relatively low between 6 and 12AM with less than half the peak usage. This tendency is visible for each of the 8 days.

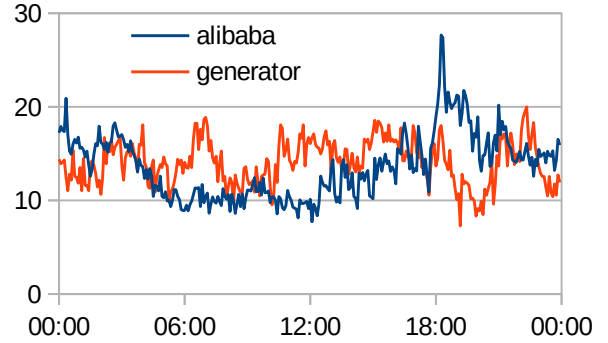


Fig. 4. The two workloads used for the simulations. Units in kW.

c) Pricing: Choosing the pricing for the electricity is fundamental as it will directly impact the objective function in the optimization problem. A high price for grid power compared to self-supplied power will lead to an important pre-cooling. On the opposite, if their ratio approaches 1, the system will have no incentive to use energy when it is available locally, resulting in no pre-cooling being carried out.

While in general the price could be dependent on time (as in Equation (14)), we chose in our study constant values for p_{sell}^k and p_{buy}^k . In order to reflect the environmental impact of IT, carbon intensities were used as prices for electricity. For PV power, the life cycle analysis of the production facility was taken into account. Asdrubali et al. made in [35] a harmonization over 50 life cycle assessments found in the literature. We picked for the PV carbon intensity the median value from their study, equal to $p_{sell} = 29.2\text{gCO}_2\text{eq}/\text{kWh}$. For grid power, the carbon intensity of the energy mix of different European countries was used. The data from the European Environment Agency [36] is reported in Table II.

¹data for 8 days is available at github.com/alibaba/clusterdata

TABLE II
CARBON INTENSITY OF ELECTRICITY GENERATION FOR SOME EUROPEAN COUNTRIES IN 2016

Country	$p_{buy}(gCO_2eq/kWh)$
Germany	440.8
France	58.5
European Union	295.8

VI. RESULTS AND DISCUSSION

A. Illustration of the cooling operation

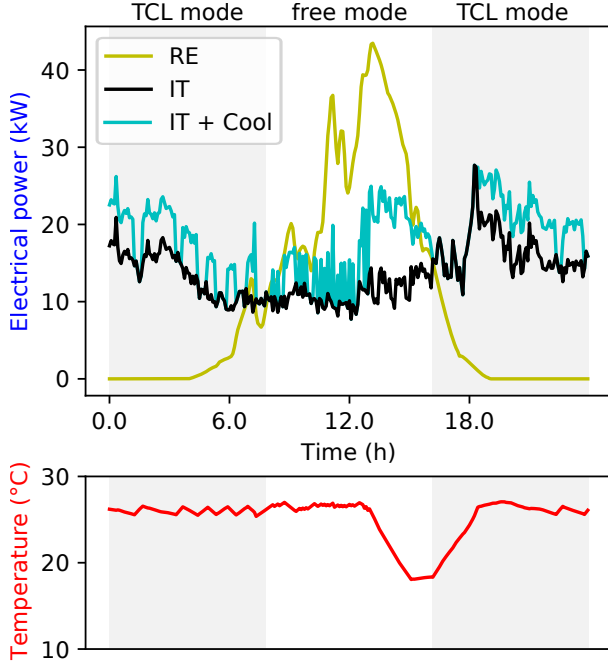


Fig. 5. Thermal behavior and power consumption of the data center for one day obtained by linear resolution. Data: trace Alibaba, median production day for Spring, french energy mix.

Figure 5 illustrates one result from the linear solving. The two modes are visible: when solar production is insufficient the data center is operated in TCL mode, otherwise free mode is triggered (Equation (21)). TCL mode features a smooth operation maintaining the temperature between the two thresholds by alternating between steps with heat pump off and steps with heat pump at state 1. The length of these steps varies depending on the outside temperature and the heat emitted by IT equipment. Heat pump operation is left to the linear solver in free mode. We can observe during the first hours the greedy management commented upon previously (before Equation (21)). Then between 1:00 and 3:05PM the heat pump starts pre-cooling at state 2. To adapt to the production, cooling is lowered to state 1 between 3:05 and 4:10PM. Finally, the heat pump can be completely switched off between 4:10 and 6:25PM. Overall in this example, 1.22% of carbon emissions have been saved compared to the control execution.

Note that this result depends highly on the energy mix. French carbon intensity of electricity is used in this example. However, with the same data but the European mix, almost 4% savings can be reached.

B. Results for different prices

	France	Germany	EU
winter	0.74%	1.85%	1.75%
spring	1.17%	4.16%	3.87%
summer	0.97%	3.89%	3.59%
fall	0.90%	3.15%	2.95%
average	0.95%	3.26%	3.04%

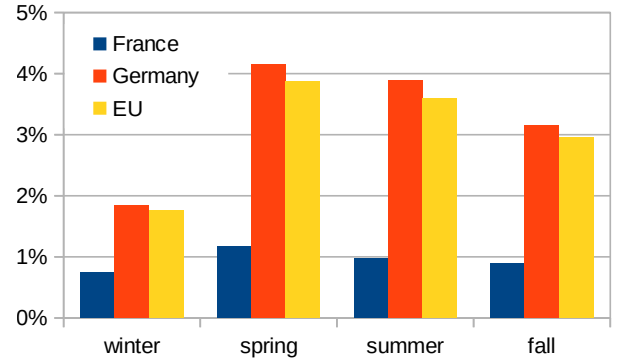


Fig. 6. Gains of pre-cooling in terms of carbon intensity of total energy supply compared to the control experiment for three different energy mix. Results for the generated trace and the median production day of each season (green line in Figure 3)

As mentioned in the previous Subsection, the influence of the grid electricity carbon intensity p_{buy} is decisive because this parameter tunes the incentive to pre-cool the data center. Figure 6 plots the results of the linear optimization (hence only when the heat pump is controlled in free mode) for four different weather data and three different energy mixes: France, Germany and the European Union (carbon intensities in Table II). Even though a more systematic and statistical analysis would be needed, *these results lead us to expect a carbon saving potential of around 1% for France and 3% for the European Union thanks to pre-cooling.*

To put these results into context, we calculated with the simulation that augmenting the temperature set point T_{min} , T_{med} , T_{max} for the thermostatic control by 1°C (resp. 3°C) leads to an overall energy saving of 1.3% (resp. 3.3%) for a full day of operation (see Table III).

In other words, in a country with a low-carbon electricity mix such as France, using a pre-cooling policy in a small data-center to optimize the consumption of locally produced renewable energy could help to save as much carbon emissions as augmenting the temperature set point by 1°C. For a more carbon-intensive electricity mix such as Germany or the European Union, it is comparable to an augmentation by 3°C.

TABLE III
OVERALL ENERGY SAVING FOR THE DATA CENTER OPERATION OBTAINED
BY AUGMENTING THE TEMPERATURE SETPOINT OF THE THERMOSTATIC
CONTROL.

+1°C	+2°C	+3°C	+4°C
1.34%	2.37%	3.30%	4.45%

C. Discussion

a) *Validity*: The thermal modelling used in this work, while been used several times in previous works as explained earlier, is nevertheless an approximate model for heat transfer and dispersion. The main strong assumption is that we consider the whole building as a homogeneous thermal mass. Also, the thermal model does not consider air confinement (hot/cold aisles for instance). More sophisticated modelling using Computational Fluid Dynamic software (for instance the TRNSYS software [32]) would certainly lead to more accurate model.

b) *Limits*: Our approach is linked to the fact that the building must have a sufficient thermal inertia and a low IT power density by square meter. The thermal inertia is linked with the constant a in Equation 5. In this formula, C varies from 0.015 to $0.065 kWh/^\circ C/m^2$ and $1/R$ varies from 0.001 (a very efficient building) to $0.003 kW/^\circ C/m^2$ [37]. In our experiments, with a $93m^2$ data center room, and in the same units, we have $C = 0.046 kWh/^\circ C/m^2$ and $1/R = 0.023 kW/^\circ C/m^2$. Different values may lead to different results: In particular, analyzing the Equation 5, the higher are C and R , the more the thermal inertia a is important. In other words, the more the walls resist to the heat transfer (higher R) and the more the data center room needs energy to increase the heat by one degree (higher C), the more the pre-cooling has potential interest.

Concerning the floor utilization impact, Figure 7 shows the evolution of the savings on CO_2 emissions when the IT density, expressed in W/m^2 , increases. When the floor utilization increases, the savings are decreasing. This is due to the fact that the heat generated needs more energy to be cooled in general. While the pre-cooling is active only on a portion of the time window (see the free mode in the illustration on Figure 5), altogether the savings are less in proportion. Note that in our experiments the density was $150 W/m^2$ ($14kW$ for a $93m^2$ server room).

Running the linear solver needs an accurate prediction for weather conditions. If the production linked with the weather conditions have been under- or over-estimated, it might be the case that the actual need for cooling and the potential for pre-cooling are different from expected. However, the system would still work, with better or worse final energy savings. A detailed formal analysis of the robustness of the results could be conducted as future works.

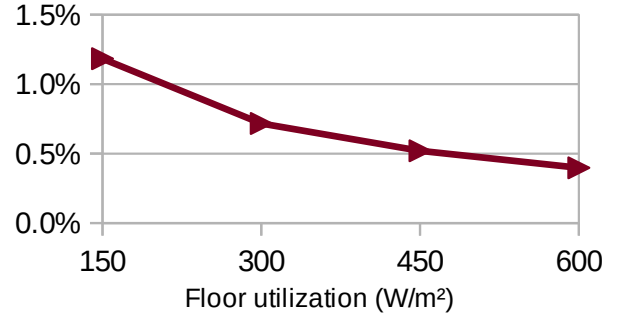


Fig. 7. CO_2 savings thanks to pre-cooling depending on IT floor utilization. Data: average over the four seasonal median days with French energy mix.

VII. CONCLUSION

In this work, the pre-cooling of data center using renewable energies has been modeled as a linear program, and simulated on different IT workload conditions. Benefits in terms of CO_2 emissions have been demonstrated even in regions where the carbon intensity of the Power Grid is low, at the cost of using more energy coming with a lower carbon intensity.

Future works will first include short term experiments with different weather conditions and energy productions (including wind) and different IT workloads to isolate the individual effects of these two inputs on the results. Second we will investigate dynamic pricing where the purchase price changes over time which is the case on the energy market. Third, we will study the impact and the robustness of the heuristic when using online predicted workload and weather conditions rather than offline traces. In parallel, an integration of the pre-cooling in a data center management system including IT scheduling and electrical commitment [21] will be conducted.

ACKNOWLEDGMENT

This work was supported by the French Research Agency under the project DATAZERO (ANR-15-CE25-0012).

REFERENCES

- [1] A.-C. Orgerie, M. D. d. Assuncao, and L. Lefevre, "A survey on techniques for improving the energy efficiency of large-scale distributed systems," *ACM Comput. Surv.*, vol. 46, no. 4, Mar. 2014. [Online]. Available: <https://doi.org/10.1145/2532637>
- [2] T. Mastelic, A. Oleksiak, H. Claussen, I. Brandic, J.-M. Pierson, and A. V. Vasilakos, "Cloud computing: Survey on energy efficiency," *ACM Comput. Surv.*, vol. 47, no. 2, Dec. 2014. [Online]. Available: <https://doi.org/10.1145/2656204>
- [3] S. Malla and K. Christensen, "A survey on power management techniques for oversubscription of multi-tenant data centers," *ACM Comput. Surv.*, vol. 52, no. 1, Feb. 2019. [Online]. Available: <https://doi.org/10.1145/3291049>
- [4] J. Wan, X. Gui, S. Kasahara, Y. Zhang, and R. Zhang, "Air flow measurement and management for improving cooling and energy efficiency in raised-floor data centers: A survey," *IEEE Access*, vol. 6, pp. 48 867–48 901, 2018.
- [5] G. Cook, J. Lee, T. Tsai, A. Kong, J. Deans, B. Johnson, and E. Jardim, "Clicking clean: Who is winning the race to build a green internet?" *Greenpeace Inc., Washington, DC*, p. 102, Jan. 2017.

- [6] Green House Data. Server Colocation | Green House Data. Accessed: 2019-09-18. [Online]. Available: <https://www.greenhousedata.com/colocation>
- [7] EcoCloud360. Secure Microsoft Hosted Exchange Cloud Service. Accessed: 2019-09-18. [Online]. Available: <https://ecocloud360.com/>
- [8] I. n. Goiri, W. Katsak, K. Le, T. D. Nguyen, and R. Bianchini, "Parasol and greenswitch: Managing datacenters powered by renewable energy," in *Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 51–64. [Online]. Available: <https://doi.org/10.1145/2451116.2451123>
- [9] L. Grange, G. Da Costa, and P. Stolf, "Green IT scheduling for data center powered with renewable energy," *Future Generation Computer Systems*, vol. 86, pp. 99–120, Sep. 2018.
- [10] S. Caux, P. Renaud-Goud, G. Rostirolla, and P. Stolf, "IT Optimization for Datacenters Under Renewable Power Constraint," in *Euro-Par 2018: Parallel Processing*, M. Aldinucci, L. Padovani, and M. Torquati, Eds. Cham: Springer International Publishing, 2018, vol. 11014, pp. 339–351.
- [11] D. Hatzopoulos, I. Koutsopoulos, G. Koutitas, and W. V. Heddeghem, "Dynamic virtual machine allocation in cloud server facility systems with renewable energy sources," in *2013 IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 4217–4221.
- [12] D. Laganà, C. Mastroianni, M. Meo, and D. Renga, "Reducing the Operational Cost of Cloud Data Centers through Renewable Energy," *Algorithms*, vol. 11, no. 10, p. 145, Oct. 2018.
- [13] A. Khosravi, A. Nadjaran Toosi, and R. Buyya, "Online virtual machine migration for renewable energy usage maximization in geographically distributed cloud data centers," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 18, p. e4125, Sep. 2017.
- [14] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and cooling aware workload management for sustainable data centers," in *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 175–186. [Online]. Available: <https://doi.org/10.1145/2254756.2254779>
- [15] A. Habibi Khalaj, K. Abdulla, and S. K. Halgamuge, "Towards the stand-alone operation of data centers with free cooling and optimally sized hybrid renewable power generation and energy storage," *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 451–472, Oct. 2018.
- [16] K. Khankari, "Thermal mass availability for cooling data centers during power shutdown," *ASHRAE Transactions*, vol. 116, 2010.
- [17] Y. Zhang, Y. Wang, and X. Wang, "Testore: Exploiting thermal and energy storage to cut the electricity bill for datacenter cooling," in *Proceedings of the 8th International Conference on Network and Service Management*, ser. CNSM '12. Laxenburg, AUT: International Federation for Information Processing, 2012, p. 19–27.
- [18] M. Lukawski, J. W. Tester, M. C. Moore, P. Krol, and C. L. Anderson, "Demand Response for Reducing Coincident Peak Loads in Data Centers," in *HICSS*, Jan. 2019, p. 11. [Online]. Available: <http://hdl.handle.net/10125/59776>
- [19] A. Oleksiak, W. Piatek, K. Kuczynski, and F. Sidorski, "Reducing energy costs in data centres using renewable energy sources and energy storage," in *Proceedings of the 5th International Workshop on Energy Efficient Data Centres*, ser. E2DC '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: <https://doi.org/10.1145/2940679.2940684>
- [20] Y. Li, X. Wang, P. Luo, and Q. Pan, "Thermal-Aware Hybrid Workload Management in a Green Datacenter towards Renewable Energy Utilization," *Energies*, vol. 12, no. 8, p. 1494, Apr. 2019.
- [21] J.-M. Pierson, G. Baudic, S. Caux, B. Celik, G. Da Costa, L. Grange, M. Haddad, J. Lecuivre, J.-M. Nicod, L. Philippe, V. Rehn-Sonigo, R. Roche, G. Rostirolla, A. Sayah, P. Stolf, M.-T. Thi, and C. Varnier, "DATAZERO: DATAcenter with Zero Emission and Robust management using renewable energy," *IEEE Access*, vol. 7, pp. 103 209–103 230, Jul. 2019. [Online]. Available: <http://doi.org/10.1109/ACCESS.2019.2930368>
- [22] C. Dupont, M. Sheikhalishahi, F. M. Facca, and F. Hermenier, "An energy aware application controller for optimizing renewable energy consumption in data centres," in *Proceedings of the 8th International Conference on Utility and Cloud Computing*, ser. UCC '15. IEEE Press, 2015, p. 195–204.
- [23] B. Camus, F. Dufossé, and A.-C. Orgerie, "The sagitta approach for optimizing solar energy consumption in distributed clouds with stochastic modeling," in *Smart Cities, Green Technologies, and Intelligent Transport Systems*, B. Donnellan, C. Klein, M. Helfert, O. Gusikhin, and A. Pascoal, Eds. Cham: Springer International Publishing, 2019, pp. 52–76.
- [24] ASHRAE, "Data center power equipment thermal guidelines and best practices," Jun. 2016, publisher: ASHRAE Technical Committee 9.9 Mission Critical Facilities, Data Centers, Technology Spaces, and Electronic Equipment. [Online]. Available: https://tc0909.ashraetcs.org/documents/ASHRAE_TC0909_Power_White_Paper_22_June_2016_REVISED.pdf
- [25] G. Tina and S. Gagliano, "Probabilistic analysis of weather data for a hybrid solar/wind energy system," *International Journal of Energy Research*, vol. 35, no. 3, pp. 221–232, 2011. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/er.1686>
- [26] Mohammad Masih Sediqi, Masahiro Furukakoi, Mohammed E. Lotfy, Atsushi Yona, and Tomonobu Senju, "Optimal Economical Sizing of Grid-Connected Hybrid Renewable Energy System," *Journal of Energy and Power Engineering*, vol. 11, no. 4, Apr. 2017.
- [27] S. Sinha and S. S. Chandel, "Review of recent trends in optimization techniques for solar photovoltaic–wind based hybrid energy systems," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 755–769, Oct. 2015.
- [28] M. Bortolini, M. Gamberi, and A. Graziani, "Technical and economic design of photovoltaic and battery energy storage system," *Energy Conversion and Management*, vol. 86, pp. 81–92, Oct. 2014.
- [29] J. K. Kaldellis, D. Zafirakis, and E. Kondili, "Optimum autonomous stand-alone photovoltaic system design on the basis of energy pay-back analysis," *Energy*, vol. 34, no. 9, pp. 1187–1198, Sep. 2009.
- [30] J.-M. Pierson, P. Stolf, H. Sun, and H. Casanova, "MILP formulations for spatio-temporal thermal-aware scheduling in Cloud and HPC data-centers," *Cluster Computing*, Apr. 2019.
- [31] N. Rasmussen, "Calculating Total Cooling Requirements for Data Centers," *White Paper #25*, vol. American Power Conversion (APC), p. 8, 2007.
- [32] D. P. Zurmühl, M. Z. Lukawski, G. A. Aguirre, W. R. Law, G. P. Schnaars, K. F. Beckers, C. L. Anderson, and J. W. Tester, "Hybrid geothermal heat pumps for cooling telecommunications data centers," *Energy and Buildings*, vol. 188–189, pp. 120–128, Apr. 2019.
- [33] SoDa, "Solar Radiation Data," Accessed: 2019-04-15. [Online]. Available: <http://www.soda-pro.com/>
- [34] G. Da Costa, L. Grange, and I. De Courchelle, "Modeling and generating large-scale Google-like workload," in *2016 Seventh International Green and Sustainable Computing Conference (IGSC)*. Hangzhou, China: IEEE, 2016, pp. 1–7.
- [35] F. Asdrubali, G. Baldinelli, F. D'Alessandro, and F. Scrucca, "Life cycle assessment of electricity production from renewable energies: Review and results harmonization," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1113–1122, Feb. 2015.
- [36] European Environment Agency, "CO2 emission intensity," Accessed: 2019-07-04. [Online]. Available: <https://www.eea.europa.eu/data-and-maps/daviz/co2-emission-intensity-5>
- [37] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, May 2009.