

Modeling CO₂ Emissions to Reduce the Environmental Impact of Cloud Applications

Cinzia Cappiello^(✉), Paco Melià, and Pierluigi Plebani

Dipartimento di Elettronica, Informazione e Bioingegneria,
Politecnico di Milano, Milan, Italy
{cinzia.cappiello,paco.melia,pierluigi.plebani}@polimi.it

Abstract. Cloud computing has important impacts on the environment: data centers – that represent the physical infrastructure where the cloud resources run – are not always designed as green entities, as they consume large amounts of energy (in form of electric power or fuel) often producing significant amounts of CO₂ emissions. Such emissions depend on the energy sources used by the data centers and may vary over time with respect to the location in which the data center is operating. To decrease the carbon footprint of cloud computing, the selection of the site where to deploy an application, along with the decision of when to start the execution of the application, should be based not only on the satisfaction of the traditional QoS requirements but also on the energy-related constraints and their dynamics over time.

Goal of this paper is to propose a CO₂ emission model able to support emission forecasting, especially for data centers that are based on electricity from the national grid. The proposed emission model can be used to improve the decisions on where and when to deploy applications on data centers in order to minimize CO₂ emissions.

Keywords: CO₂ emissions · Cloud computing · Sustainability

1 Introduction

A substantial part of the energy demand of the IT sector is associated with the functioning of data centers: facilities responsible for the storage and processing of information which constitute the infrastructural basis of cloud computing. Due to the increasing adoption of cloud-based solutions, the role of data centers is becoming more and more crucial. Consequently, the energy required by data centers is now so high that it has been estimated to be comparable with the overall energy consumption of countries like Japan and India, and would rank 5th among world countries [6].

Several approaches have been proposed in the last years with the goal of reducing the energy consumption of data centers. As a significant fraction of the energy consumed in a data center is not consumed for running servers and network devices needed to execute the deployed application, but goes to supplementary elements (e.g., cooling system, lightning), most of the proposed approaches

have focused on reducing this fraction of energy. Some other approaches have focused on optimizing the design, the deployment, and the execution of cloud applications [5] in order to minimize the waste of resources, find the servers with the lowest energy consumption, or distributing the workload looking at the amount of energy consumed by the different servers.

In this work we want to adopt a different perspective by looking at CO₂ emissions instead of at the pure energy consumption. Indeed, the actual environmental impact of energy usage very much depends on the specific energy mix used to generate electric power. In fact, as the ability of different power generation technologies to follow daily, weekly and seasonal fluctuations of power demand varies widely (with large nuclear or coal-fueled plants requiring hours, if not days, to change their power output, and natural gas turbines or hydro-electric plants being able to respond in minutes), the composition of the energy mix is subject to important variation over different time horizons. CO₂ emission factors hence change consequently, according to the different power generation sources running in different moments of the day and in different seasons. For this reason, the optimization of cloud applications requires appropriate tools to forecast the dynamics of the energy mix and the relevant emission factors over time, and to identify the best schedule to deploy cloud applications.

The goal of this paper is to propose a modeling framework to predict short and medium-term fluctuations of (i) electric energy consumption at the national scale, (ii) the composition of the energy production mix by energy source, and (iii) the consequent CO₂ emission factor. We demonstrate the framework by applying it to the assessment of a federated cloud infrastructure. We use data from two European countries, i.e., France and the United Kingdom, for which time series of energy production (disaggregated by generation source) and consumption are available in real time at high temporal resolution. We eventually show how the availability of accurate CO₂ emission forecasts can support IT systems managers in pursuing a greener deployment of cloud applications.

The paper is structured as follows. Section 2 reviews the state of the art of the literature on the topic to highlight the novelty of the proposed approach. Section 3 illustrates the modeling framework and demonstrates it through the application to the two case studies (France and the UK). Section 4 describes how the models developed in Sect. 3 can be used to select the most appropriate site for application deployment on the basis of predicted CO₂ emissions.

2 Related Work

Assessment, measurement and improvement of energy efficiency in data centers and clouds have been important mainstreams in research in recent years [16]. The work proposed in this paper also focuses on the improvement of the sustainability of data centers, but it aims to reduce the CO₂ emissions by adapting the applications running on federated cloud environments. The importance of CO₂ emissions has been already considered in the literature. Some contributions focus on how to use efficiently available renewable resources, while avoiding

peak demand of energy from electricity providers. In [1], authors propose the usage of Geographical Load Balancing (GLB) to shift workloads and avoid peak power demands. This requires to predict both the incoming workload and the peak demand to the network. The algorithm is implemented as a network flow optimization problem. A similar approach is discussed in [11], which uses both workload shifting and local power generation for avoiding peak load demands on the energy network. The importance of considering the type of energy sources has been addressed also in [2], which proposes an integrated framework for sustainable clouds where information on data centers, communication networks and energy sources is considered. In [15], the authors focus on the optimization of power generation with respect to carbon emissions. Also [13] focuses on assessing the carbon footprint of cloud computing services. In this paper we aim to provide a tool to forecast carbon emissions and/or to suggest the most suitable deployment time to improve the sustainability of cloud applications. Mathematical models to forecast electric power demand over time horizons spanning from few minutes ahead (very short-term forecasts) up to a decade ahead (long-term forecasts) are key to support operations and planning of power systems, and have been the subject of research since the late 1960 [12]. In the early 2000s, Alfaraz and Nazeeruddin [4] carried out an exhaustive review of the vast range of literature produced on the subject in the previous fifty years, and identified nine major categories of load forecasting techniques: multiple regression; exponential smoothing; iterative reweighted least-squares; adaptive load forecasting; stochastic time series; ARMAX models based on genetic algorithms; fuzzy logic; neural networks; and knowledge-based expert systems. In the last decade, other methods have been developed and tested (see, e.g., [9,10]), further improving forecasting reliability.

3 Energy Mix Analysis

Greener choices in the deployment of cloud applications should be performed by considering not only the typical quality aspects (e.g., response time, availability, security) but also green requirements which involve both energy consumption and CO₂ emissions. Focusing on the latter, evaluation of CO₂ emissions is based on emission factors (gCO_{2e}/kWh) provided by national grids. Emission factors largely vary from country to country. For example, if we consider France and the United Kingdom, technical reports describe that the country with the lowest carbon intensity is France, whose power generation is mainly based on nuclear plants. Estimated emission factors for France range between 62 [3] and 146 [8] gCO_{2e}/kWh. In contrast, United Kingdom energy is more carbon-intensive, with emission factors estimated to range between 567 [7] and 658 [8] gCO_{2e}/kWh.

As our goal is to deploy an application in a federated cloud environment, calculating and predicting emission factors for each of the sites included in the federation are crucial aspects in our approach. Indeed, knowing in advance which will be the emission factors of the countries in which the data centers belonging to the federation are established would allow us to calculate how CO₂ emissions

may vary with respect to the location in which the application will be deployed and the time at which it will be executed. To this aim, we started from historical values about power generation disaggregated by energy source that some countries publish via public web sites. In particular, the French energy mix can be retrieved through the information service *éCO2mix* available on the RTE website¹. Such service shows electricity demand, electricity generation classified by source and cross-border commercial exchanges (imports/exports). Data are automatically updated every 15 min. Similar information is available for the UK. Real-time and historic data about the energy generation in the UK are available through the BMRS (Balancing Mechanism Reporting System) website². For this web site data are updated every 5 min.

Having the values from these two web sites, we constructed a model that reproduces in a simplified, but sufficiently precise way the analyzed systems, providing a tool to forecast CO₂ emissions and taking greener decisions when deploying cloud-based applications.

To build the model, we adopted a traditional approach and went through the following sequential phases:

- *Analysis of the problem*: it is necessary to observe the problem in order to understand the goals of the model and the data that have to be retrieved.
- *Conceptualization*: a model aims to provide a representation of a real world scenario. This phase focuses on guaranteeing the accuracy and completeness of the model. In fact, the model should be a simple, concise and correct view of the reality and should include all the elements that are considered as relevant. However, the model should not be too complex: complexity often implies a higher computational cost (i.e., longer execution time).
- *Calibration*: after gathering a sufficient amount of data, the calibration phase aims to estimate the parameters included in the model. For this reason, such phase is also called *parameterization* and can be performed by using several methods. We used the most common one: the Least Squares Method that minimizes the sum of the squares of the errors.
- *Validation*: the goal of the validation is to consider a new dataset and verify that the calibrated model is able to explain data trends and characteristics. If the results of the validation are not satisfactory, it is possible to enrich the dataset used for the calibration or go back to the conceptualization phase and change the model.

3.1 Analysis of the Problem

As already mentioned, the model to estimate CO₂ emissions has been designed by observing the available data on power production of France and the UK for two years: 2013 and 2014. As depicted in Fig. 1, the French energy mix is mainly composed of nuclear sources, while hydroelectric plants are the second most

¹ <http://www.rte-france.com/fr/>.

² <http://www.bmreports.com/>.

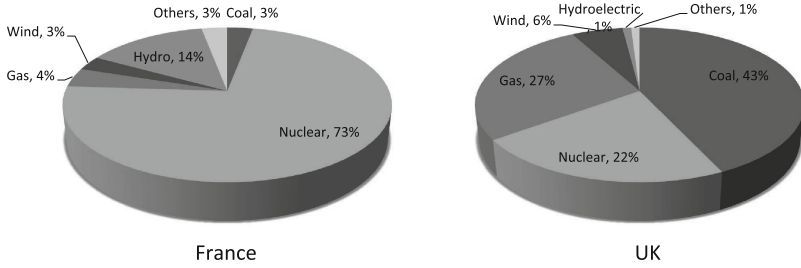


Fig. 1. Energy mix in France and the UK in 2013

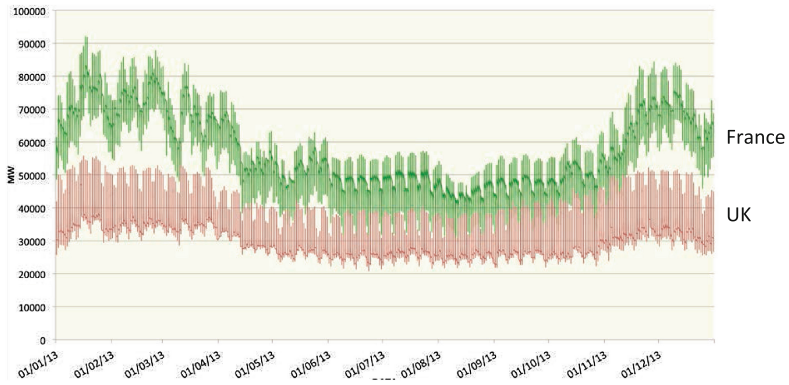


Fig. 2. Energy consumption in France and UK in 2013

important energy source, mainly used to dampen the fluctuations of nuclear production. The UK has a more diversified energy mix: 91 % of the energy production comes from coal, nuclear and gas. With respect to renewable sources, it is worth to notice that the UK relies on wind for 6 % of the whole energy production. This is due to the fact that the typical English weather is mainly windy especially in the cities close to the sea. Comparing the total production of the two countries (see Fig. 2), it is immediately apparent that, on average, the production of France is higher. This is not only due to the difference in population (nearly 67 millions in France against 60 millions in the UK), but also to the high production of thermonuclear energy in France, a considerable portion of which is exported to the neighboring countries.

3.2 Conceptualization

Based on the data gathered from the already mentioned web sites, Fig. 2 shows a comparison between energy consumption patterns in France and the UK during 2013. Regardless of the country, it is clear that some patterns occur periodically at different levels: from hour to hour, from day to day, from season to season.

To find a rationale behind this behavior, we analysed the data and we took into account the following time-variant elements that may be correlated with the energy consumption:

- *Temperature (T)*: considering the climate of the two countries, we assumed that higher consumption levels are related to lower temperatures (e.g., for heating).
- *Daylight Hours (DH)*: the electricity consumption raises in the periods characterized by a smaller number of hours of light.
- *Average seasonal Trends (Avg)*: we observed regular seasonal trends in weekdays data.
- *Power generation of close instants of time (P)*: we observed a clear autocorrelation between subsequent weekdays and among the same days of different weeks (e.g., every Monday or every Sunday).

On the basis of these considerations, we drafted a general model as:

$$P(t+1) = a \cdot f(P) + b \cdot f_1(T) + c \cdot f_2(DH) + d \cdot f_3(Avg) + e \cdot f_4(error) + error(t+1)$$

Analyzing such a model it is possible to notice that it can be formally represented by using a PARMAX model that is composed of the following parts:

- *The P (Periodic) part*: it is related to the time-variant parameters, that are parameters that have different values depending on the time in which they are considered. In the model this part includes the daily and seasonal parameters that vary on the basis of the day and season of the year.
- *The AR (Auto Regressive) part*: it links the estimated value with the previous values. In this case this part considers the emissions of the day and of one week before.
- *The MA (Moving Average) part*: is associated with the residual information that is the prediction error at previous time steps.
- *The X (eXogenous) part*: it is used to model the information contained in external variables (e.g., temperature and daylight hours)

For each energy source (e.g., nuclear, hydroelectric, coal) used by the two countries, a model has been defined. Due to space limitations, we do not list here all the models defined, but we present only the models of the most important sources for the two countries.

Model for Nuclear in France. French power production mainly relies on nuclear plants. Looking at the power production of this energy source over the year (i.e., 2013), it is possible to notice that there are recurrent seasonal and daily patterns: the production is higher in winter and lower in summer, and also during the day there is an oscillatory behaviour, which shows that the production is higher during the daytime while it decreases at nighttime. Note that the seasonal trend mainly depends on temperature, which is clearly correlated to power production as shown in Fig. 3.

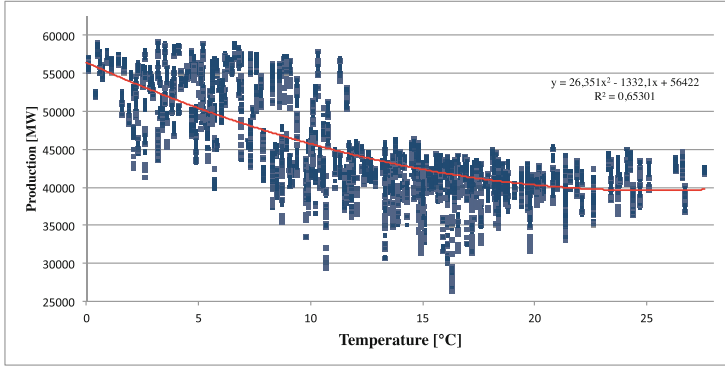


Fig. 3. Correlation between nuclear production and average daily temperature

Considering that time is expressed in hours, the formal model that we have defined to estimate the nuclear power consumption (N) of the next hour is:

$$N(t+1) = \alpha_1 \cdot N_{t+1}^* + \alpha_2 \cdot N_{t-5}^* + \alpha_3 \cdot N_{t-23} + \alpha_4 \cdot \zeta_{N_{t-23}^*} + \alpha_5 \varepsilon_{t-167} + \varepsilon_{t+1}$$

where:

N_{t-23} = nuclear power consumption one day before (same time)

ε_{t-167} = error of the model one week before (same day and same time)

$\alpha_{1...5}$ = coefficients to be estimated

$\zeta_{N_{t-23}^*}$ = error produced by N^* at the time instant $t - 23$ (the same time of the day before)

N^* is the estimation of the value that has been formalized as:

$$N_t^* = (\tau_1 \cdot T_{t-24}^2 + \tau_2 \cdot T_{t-24} + \mu_1) + \Omega_t^h + \Omega_t^{h,s} + \Omega_t^{wd}$$

T_{t-24} = average temperature of the day before

$\tau_1 = 26.351, \tau_2 = -1332.1, \mu_1 = 56422$ = fixed coefficients that have been estimated through correlation

Ω_t^h = coefficient that is dependent on the time

$\Omega_t^{h,s}$ = coefficient that is dependent on the time and season

Ω_t^{wd} = coefficient that is dependent on the weekday

In our model the estimation N_t^* has to be calculated considered as parameters $t+1$ and $t-5$.

Model for Coal in the UK. Power production in the UK mainly relies on coal. Looking at the power production of such energy source over the year (i.e., 2013), it is possible to notice also in this case daily and seasonal trends and a correlation with temperature. The formal model obtained for this source is:

$$C(t+1) = \alpha_1 \cdot C_{t+1}^* + \alpha_2 \cdot C_{t-23} + \alpha_3 \cdot C_{t-167} + \alpha_4 \cdot \varepsilon_{t-23} + \varepsilon_{t+1}$$

where:

C_{t-23} = coal power consumption one day before (same time)

C_{t-167} = coal power consumption one week before (same day and same time)

ε_{t-23} = error of the model one day before (same time)

$\alpha_{1..4}$ = coefficients to be estimated

C^* is the estimation of the value that has been formalized as:

$$C_{t+1}^* = (\beta_1 \cdot D_t^2 + \beta_2 \cdot D_t + \mu_1) + (\tau \cdot T_{t-23} + \mu_T) + \Omega_t^{h,s} + \Omega_t^{wd}$$

D_t = number of days from the beginning of the year

T_{t-23} = average temperature of the day before

$\beta_1, \beta_2, \tau, \mu_1, \mu_T$ fixed coefficients that have been estimated through correlation

$\Omega_t^{h,s}$ = coefficient that is dependent on the time and season

Ω_t^{wd} = coefficient that is dependent on the weekday.

3.3 Calibration

The calibration of the defined models has been performed with the Least Squares Method. Figure 4 shows calibration results for the two models presented above. Considering the French model, it is possible to notice that the more relevant variable is the value of the power production recorded 24 h earlier. In the British scenario, three variables have a significant role in the model and, in particular, the values recorded 24 h and a week earlier and the estimation of coal production based on the season and temperature.

Parameters	Value	Parameters	Value
α_1	0.294	α_1	0.296
α_2	0.020	α_2	0.339
α_3	0.686	α_3	0.369
α_4	0.108	α_4	0.332
α_5	0.382		

(B) Coal- UK

(a) Nuclear - France

Fig. 4. Calibration for the French nuclear and British coal models

3.4 Validation

The validation phase focuses on two steps: (i) the comparison between the estimated and real values for 2014, (ii) the aggregation of the different models for each country and (iii) the validation of the aggregated model on data for 2014.

Figure 5 shows the performances of the calibrated models of nuclear power production in reconstructing data from 2014. In fact, the correlation between observed data and model predictions is 0.945 (and thus $R^2=0.893$).

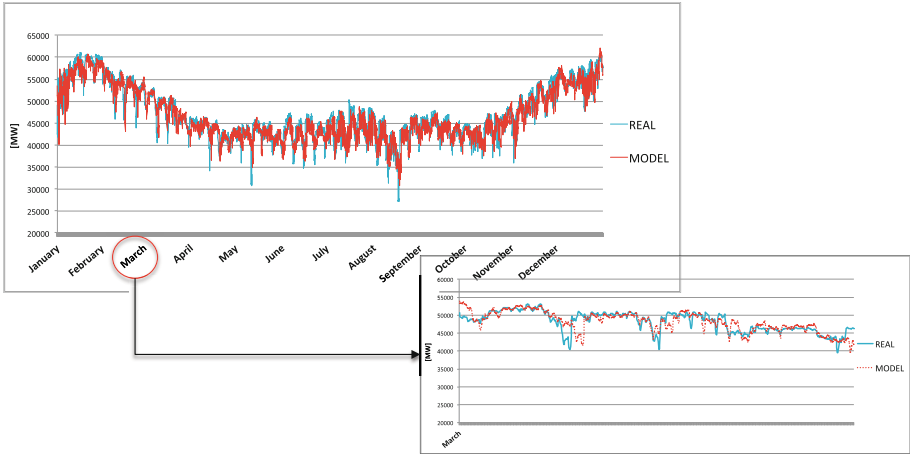


Fig. 5. Validation of the French nuclear model with 2014 data

Good results have been obtained also for the UK: the model for coal has a correlation with the real data $R=0.944$ and $R^2=0.891$.

In order to estimate CO₂ emissions coefficients for each country, it is necessary to aggregate the different models defined for the different energy sources. The aggregated model has been also validated. Results for France and the UK are reported in the following:

- *France*: Correlation model - real data= 0.971, $R^2=0.942$
- *UK*: Correlation model - real data= 0.940, $R^2=0.884$

4 CO₂-Driven Site Selection

To demonstrate how the approach presented in this paper can be applied to the estimation of CO₂ emissions, we refer to a scenario involving a federated cloud infrastructure. More in detail, we assume that several cloud platforms established in different countries around the world constitute a federation. This means that agreements between the owners of these platforms exist in order to optimize the usage of the installed resources (e.g., VMs, storage). As a result, migration of VMs among the sites, as well as the possibility to control the execution of the applications on top of them are all possible actions.

From the application perspective, we consider a real HPC application in the ecology domain [14] shown in Fig. 6 adopting BPMN notation. Without entering into details, the application starts with an initial setup (activity A1). The work is then split into several instances composed of two activities: data loading (A2) and computation (A3). Once all the instances are terminated, the partial results are aggregated (A4) to provide the result to the final user. We assume that one VM is required for A1 and A4, while for A2 and A3 the number of VMs may change according to the number of iterations required.

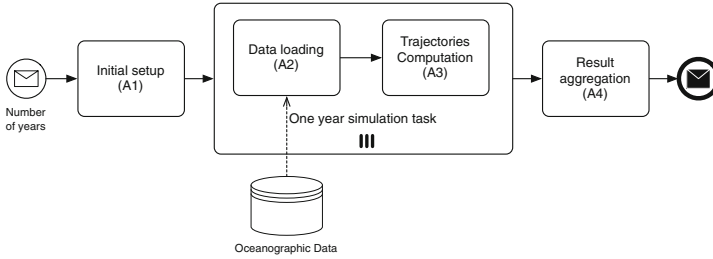


Fig. 6. Running example.

Along with the usual constraints about the VMs expressed in terms of number of cores, amount of memory, or storage, developers can also specify constraints on VM locations. Such constraints can be justified by legal issues (e.g., the data managed during A4 must not be moved to the USA) or to increase performances (e.g., A1 and A2 communicate very frequently and exchange a significant amount of data so it is better to put them at the same location).

Starting from this example, we want to show how the CO₂ emission model described in the previous sections can be exploited to decide when and where it is preferable – to minimize CO₂ emissions – to deploy and run the application. Since the only countries for which the CO₂ emission model has been produced in this paper are France and the UK, the following discussion refers to deployments that can occur on such locations. Similarly, as the data used to validate the CO₂ models refers to the year 2014, all the examples refer to this period.

As we are considering an HPC application, it is reasonable to assume that the same application has been already executed in the past. For this reason, we can also assume that some information about the energy consumed and the response time of the application when running in the UK or in France is already available. Based on our real experiments, we have the following situation [5]:

- UK: Response time = 17.07 h; Energy = 197.87 Wh.
- France: Response time = 13.50 h; Energy = 30.645 Wh.

These numbers refer to the scenario in which one VM is assigned to each of the activities and there is no concurrent iterations of activities A2 and A3. This assumption does not hamper the validity of our approach as having more instances of A2 and A3 simply reduces the response time regardless of the site in which the applications is running. The difference in response time and the energy consumed by the application depends on the characteristic of the physical machine installed on the two locations. More precisely, the British data center is equipped with less recent machines, so they are less performant and do not have low-power processors installed. Conversely, the French data center has been established more recently, with physical servers implementing several techniques to reduce power consumption.

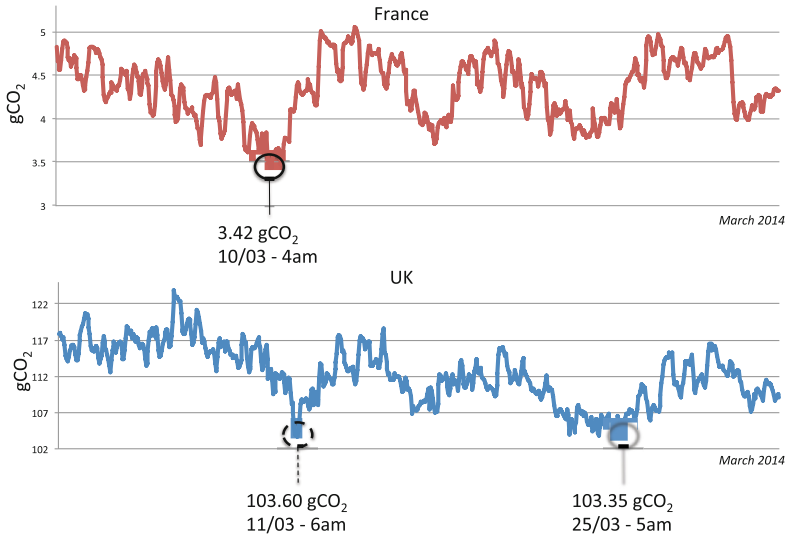


Fig. 7. Analysis of CO₂ emissions related to the considered application in the UK and France

Assuming that today is March 1st, 2014 and the application needs to run in 30 days, we use the proposed models to estimate the energy mix and, based on the energy consumed by the application, the related CO₂ emissions.

Figure 7 shows the result of this computation considering the two possible deployments: i.e., on the UK or on France. These CO₂ emission trends can be used to figure out if it is better to deploy the VMs on France or the UK, and when the application has to start, with the final goal of reducing the CO₂ emissions. Based on them, it is easy to understand that deploying the application in France is always the best choice: CO₂ emissions are always lower (about 4 gCO_{2e}) than in case the application run in the UK (about 110 gCO_{2e}). Stated this, the best time to run the application is on March 10th at 4 am, as the estimated CO₂ emissions are predicted to be 3.42 gCO_{2e}. In case, for some reason, the deployment must be done on the UK, then the lowest CO₂ emission is expected to occur on March 11th at 6 am or on March 25th at 5 am.

5 Concluding Remarks

This paper highlights the importance of CO₂ emissions in the deployment of applications. In particular, we show the way in which it is possible to build models to estimate future trends in emissions in order to suggest the most suitable deployment time able to improve the sustainability of the application. The validation scenario based on real data publicly available on the energy mix in France and UK shows how energy savings can be obtained by following a particular deployment strategy.

Acknowledgment. We would like to thank Valeria Crespi, Michele Del Vecchio, Alessandro Gentile and Marco Tangi for their valuable contribution in the definition of the different models.

References

1. Abbasi, Z., Pore, M., Gupta, S.K.S.: Impact of workload and renewable prediction on the value of geographical workload management. In: Klingert, S., Hesselbach-Serra, X., Ortega, M.P., Giuliani, G. (eds.) EDC 2013. LNCS, vol. 8343, pp. 1–15. Springer, Heidelberg (2014)
2. Addis, B., Ardagna, D., Capone, A., Carello, G.: Energy-aware joint management of networks and cloud infrastructures. *Comput. Netw.* **70**, 75–95 (2014)
3. ADEME: Guide des facteurs d'émissions, Version 6.1. Chapitre 2, Facteurs associés à la consommation directe d'énergie. Technical report (2010)
4. Almehaie, E., Soltan, H.: A methodology for electric power load forecasting. *Alexandria Eng. J.* **50**(2), 137–144 (2011)
5. Cappiello, C., Ho, N., Pernici, B., Plebani, P., Vitali, M.: Co2-aware adaptation strategies for cloud applications. *IEEE Trans. Cloud Comput.* (2015), (pre-print)
6. Cook, G.: How Clean is Your Cloud? Technical report, Greenpeace, April 2012
7. Department for Environment, Food, Rural Affairs: Guidelines to Defra/DECC's GHG Conversion Factors for Company Reporting: Methodology Paper for Emission Factors. Technical report (2012)
8. European Commission: European Commission How to develop a Sustainable Energy Action Plan (SEAP). Technical report (2010)
9. Friedrich, L., Afshari, A.: Short-term forecasting of the Abu Dhabi electricity load using multiple weather variables. *Energy Procedia* **75**, 3014–3026 (2015)
10. Garca-Ascanio, C., Mat, C.: Electric power demand forecasting using interval time series: a comparison between VAR and iMLP. *Energy Policy* **38**(2), 715–725 (2010)
11. Liu, Z., Wierman, A., Chen, Y., Razon, B., Chen, N.: Data center demand response: avoiding the coincident peak via workload shifting and local generation. *Perform. Eval.* **70**(10), 770–791 (2013)
12. Matthewman, P.D., Nicholson, H.: Techniques for load prediction in the electricity-supply industry. *Proc. Inst. Electr. Eng.* **115**(10), 1451–1457 (1968)
13. Maurice, E., Dandres, T., Moghaddam, R.F., Nguyen, K.K., Lemieux, Y., Cheriet, M., Samson, R.: Modelling of electricity mix in temporal differentiated life-cycle-assessment to minimize carbon footprint of a cloud computing service. In: *ICT for Sustainability 2014 (ICT4S-14)*, 25 August, 2014, Stockholm, Sweden (2014)
14. Melià, P., Schiavina, M., Gatto, M., Bonaventura, L., Masina, S., Casagrandi, R.: Integrating field data into individual-based models of the migration of European eel larvae. *Mar. Ecol. Prog. Ser.* **487**, 135–149 (2013)
15. Mirzaesmaeeli, H., Elkamel, A., Douglas, P., Croiset, E., Gupta, M.: A multi-period optimization model for energy planning with CO2 emission consideration. *J. Environ. Manag.* **91**(5), 1063–1070 (2010)
16. Vitali, M., Pernici, B.: A survey on energy efficiency in information systems. *J. Coop. Inf. Sys.* **23**(3), 1–38 (2014)