Probabilistic Power Demand Model For Colocation Data Centers

Gustavo E. da Silva
Department of Systems and Energy
State University of Campinas
Campinas, Brazil
g236236@dac.unicamp.br

Ricardo Torquato
Power Distribution Systems
ERA Energy Research and Analytics
Campinas, Brazil
ricardo.torquato@eraenergy.com.br

Walmir Freitas
Department of Systems and Energy
State University of Campinas
Campinas, Brazil
walmir@unicamp.br

***Abstract*—The recent expansion of data center grid connections has raised concerns about their impact on distribution systems, creating the need for proper power demand models. Existing models rely on the knowledge of specific system parameters and variables and, as such, cannot be generalized to other facilities. In response, this paper identifies the key uncertainties in the operation of colocation data centers and uses this information to develop a probabilistic power demand model for these loads. The model relies only on characteristics readily available in practice such as the capacity of the data center main components (servers, power supply and cooling systems) and, therefore, can be used to represent different colocation data centers. The proposed formulation is validated with real power demand measurements from six data centers in Brazil and presents good accuracy both in terms of instantaneous power demand (average 9% error) and monthly energy consumption (average 5% error).**

***Index Terms*—Colocation data centers, energy consumption, power demand, probabilistic model, real data validation.**

I. Introduction

From 2020 to 2024, the global energy demand for data centers grew from 250 TWh to 420 TWh. In 2024, the major energy consumers on data center sites were the United States and China, responsible for 43% and 24% of the total global consumption, respectively. Europe consumed 16%, Asia, excluding China, was responsible for about 10% and the rest of the world was responsible for 7% of the global energy consumed by this type of load [1]. In 2025, public data centers were expected to amount to 6111 ventures worldwide, presenting a global Compound Annual Growth Rate of 6.8%, mainly driven by Asia Pacific, Middle East, Africa and South America. This number is expected to grow to 8378 sites worldwide by 2030 [2]. In Soth America, the Brazilian scenario stands out, especially due to the availability of renewable resources, internal computational demand and reliable power grids. The total demand of data centers connected directly to the Brazilian transmission system in 2025 was 826 MW, and this demand is expected to exceed 20 GW by 2035 [3].

Among the distinct types of data centers, colocation facilities have experienced a significant rise, corresponding to 90% of global installations with 5544 centers and 64% of the total Brazilian sites in 2025 [4]. These ventures offer flexibility and cost-reduction to the user by allowing the outsource of data infrastructure needs. In colocation data centers, enterprises and users can rent server space while the facility holder provides the necessary infrastructure and maintenance.

Due to the massive loads that these sites represent and the increased number of recent connections, understanding the behavior of their power demand and energy consumption became essential to map their impact on the electrical system as well as on other customers, ensuring robust system planning and operation practices. Due to the high power factor nature of the load (typically over 0.98), active power demand is the most relevant component of the overall demand. In this context, existing literature such as [5], [6] provide computational models of the main data center components (also named sub-systems), but these models do not cover the complete infrastructure and do not provide typical values for the model parameters. Other formulations such as [7], [8], despite presenting a complete model, are highly dependent on facility-specific parameters and usage data and, as such, cannot be used for other data centers. Therefore, they cannot be directly adopted by utility engineers and system planners in their studies, which must consider multiple data center loads.

In this context, this paper proposes a generalized approach to model colocation data centers when detailed information of their internal infrastructure is not available. The key idea is to identify the main uncertainties of the demand profile of internal components, such as the server usage pattern, and model them as random variables, to obtain a probabilistic model of the data center. In addition to modeling consumption uncertainties, this work quantifies variables that have not been previously quantified in the existing literature, such as the series power loss coefficient and the idle power demand of the power conditioning system equipment. A detailed validation study is also conducted with real measurement data from Brazilian colocation data centers.

This paper is organized as follows. The formulation of the power demand of the main sub-systems, the random variables adopted to represent demand uncertainties, and the process to generate the simulation scenarios are presented in Section II. Section III presents the model application and validation considering measurement data from six real data centers. Section IV outlines the main conclusions.

II. Probabilistic Consumption Model

Data center power consumption is typically segmented into four main parts: IT load, power conditioning, cooling and

This work was supported in part by the National Council for Scientific and Technological Development (CNPq), grants #88887.201666/2025-00 and #310934/2023-0, in part by the São Paulo Research Foundation (FAPESP), grant #2021/11380-5, and in part by CPFL Energia, through the ANEEL R&DI program, under grant PD-00063-3091/2023.

979-8-3315-8138-1

miscellaneous systems. Figure 1 presents a data center electric diagram with its main components. Each of these parts is modeled in detail in this section.

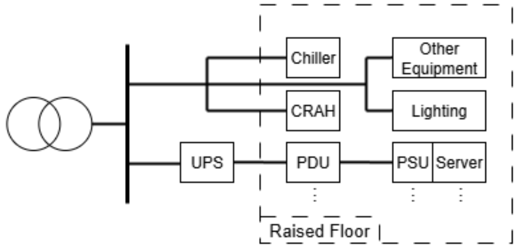

Figure 1. Data center electric diagram.

A. Load Consumption Models

1) IT load

The IT equipment is responsible for data processing in a data center. Due to its importance, this load segment is labeled as critical and comprises at least half of the power demand on modern data centers [5]. Server units represent most of the IT load and its power consumption is highly dependent on CPU utilization, u. The relation between power demand and utilization can be modeled as a linear relation as shown in (1),

$$P_{server} = P_{server}^{idle} + (P_{server}^{max} - P_{server}^{idle}) \cdot u_{server} \quad (1)$$

where P_{server}^{idle}, P_{server}^{max} are the average power values when the server is idle and fully utilized respectively, and u_{server} is the server utilization varying from 0 to 1. This formulation is the simplest and most applicable on distinct types of servers [9]. As the idle server power demand is typically close to 60% of the full data center power demand [10], (1) can be simplified to (2),

$$P_{server} = P_{server}^{max} \cdot (0.6 - 0.4 \cdot u_{server}) \quad (2)$$

where the only parameters are the total server power, that is known or at least predictable, and the server utilization.

2) Power conditioning system

The power conditioning system is composed of Uninterruptible Power Supplies (UPS), Power Delivery Units (PDU) and Power Supply Units (PSU), with the latter located at the rack level. The power demand of this system is mainly composed of PDU and UPS losses. PDU losses depend on the square of the server demand, whereas UPS losses have a linear relationship with server demand as shown in (3) and (4),

$$P_{PDU}^{losses} = P_{PDU}^{idle} + \varphi_{PDU} \cdot (P_{server})^2 \quad (3)$$

$$P_{UPS}^{losses} = P_{UPS}^{idle} + \varphi_{UPS} \cdot (P_{server} + P_{PDU}^{losses}) \quad (4)$$

where P_{PDU}^{idle}, P_{UPS}^{idle} represents the PDU and UPS power demand in idle state, respectively, while φ_{PDU}, φ_{UPS} are the series power loss coefficients in the devices.

Even though these parameters are not quantified in existing literature, their values can be estimated based on the typical losses in these components and in the power demand ratio of these sub-systems relative to the servers found in [11] and [5], respectively. TABLE I presents the PDU and UPS parameters obtained from the above assumptions. In this work, these parameters remain constant for all colocation data centers.

TABLE I. PDU AND UPS PARAMETERS

Parameter	***Value***
$\mathbf{P_{PDU}^{idle}}$	$\mathbf{0.016095 \cdot P_{server}^{max}}$
$\boldsymbol{\varphi_{PDU}}$	$\mathbf{0.016095/P_{server}^{max}}$
$\mathbf{P_{UPS}^{idle}}$	$\mathbf{0.0448 \cdot P_{server}^{max}}$
$\boldsymbol{\varphi_{UPS}}$	$\mathbf{0.05219}$

3) Cooling system

The cooling process is the major power consuming load different from the IT load. Presently, there are several cooling alternatives available, although the air-cooling technique is the most common. The air-cooling system consists of two main elements: the Computer Room Air Handler (CRAH), which manages the cold air into the racks and absorbs the hot airflow from the server room; and the chiller plant which exchanges the heat collected by the CRAH units to the external environment and provides cold air for the cooling process.

CRAH unit power demand depends on the idle state power losses, P_{CRAH}^{idle}, and the electrical power required for server heat absorption as in (5), (6). The power required for heat transfer depends on the maximum standard airflow volume (typically 14000 m³/h), server utilization (u_{server}) and heat transfer efficiency (η_{heat}). The total power consumed for heat absorption by the CRAH unit is shown in (6), where the idle state loss is typically 7% to 10% of the maximum server power.

$$P_{heat} = 1.33 \cdot 10^{-5} \cdot \frac{P_{servers}^{max}}{\eta_{heat}} \cdot f_{max} \cdot u_{server} \quad (5)$$

$$P_{CRAH} = P_{CRAH}^{idle} + P_{heat} \quad (6)$$

Meanwhile, chiller plant power consumption is modeled in (7), depending on the equipment demand on the raised floor,

$$P_{chiller} = \frac{P_{server} + P_{PDU} + P_{CRAH} + P_{light} + P_{other}}{\alpha_{chiller}} \quad (7)$$

where $\alpha_{chiller}$ is the coefficient of performance of the chiller, which is dependent on the external data center temperature. Equation (8) presents the $\alpha_{chiller}$ relation to external temperature, $T_{external}$, obtained from the linear regression on data provided in [12].

$$\alpha_{chiller} = -0.079 \cdot T_{external} + 5.891 \quad (8)$$

979-8-3315-8138-1

4) Miscellaneous systems

Besides IT, power conditioning and cooling loads, data centers usually own other power consuming devices such as lighting, security, control and monitoring systems. Typically, lighting devices consume about 1% and the other miscellaneous systems consume about 10% of total facility demand [5]. As PDU, UPS and CRAH units together demand about 75% of the total power, lighting and other equipment power consumption are modeled as shown in (9), (10).

$$P_{light} = 0.01 \frac{(P_{server} + P_{PDU} + P_{UPS} + P_{CRAH})}{0.75} \quad (9)$$

$$P_{other} = 0.1 \frac{(P_{server} + P_{PDU} + P_{UPS} + P_{CRAH})}{0.75} \quad (10)$$

B. Main Random Variables

This section discusses and models the main uncertainties of the data center demand.

1) Server usage

The server utilization is a highly variable and unknown parameter which depends on the users access and the service provided by the data center. Typically, the IT load operates on low utilization levels, and the probability of use decays exponentially as the degree of utilization increases as shown in Figure 2 with data from [10].

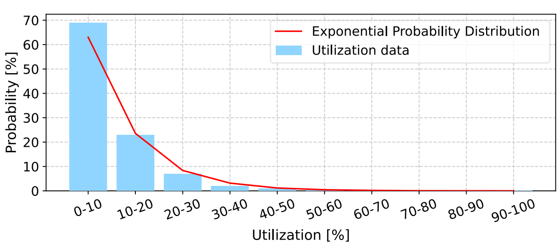

Figure 2. Probability of server utilization

Such behavior is usual at colocation centers which presents an average utilization of 15% [13]. Considering these factors, the utilization level may be modeled as a random variable following an exponential probability distribution with a mean of 15%, as in (11),

$$u(t) = Exp(\lambda = 0.15) \quad (11)$$

where the exponential probability density function is given in (12),

$$f_u(x) = \lambda \cdot e^{-\lambda x}, \qquad \lambda = 0.15 \quad (12)$$

2) CRAH idle consumption

The equipments of the cooling system (and, consequently, their model parameters) may vary from one enterprise to another. CRAH idle power consumption is one of these parameters that depend on the air handler unit characteristics and should be modeled as a random variable. Even though this variable may assume different values, typically the idle state loss is about 7% to 10% of the maximum server power [5]. Therefore, P_{CRAH}^{idle} can be defined as a random variable uniformly distributed as shown in (13),

$$P_{CRAH}^{idle} = Unif([0.07 \cdot P_{servers}^{max}, 0.1 \cdot P_{servers}^{max}]) \quad (13)$$

3) CRAH heat transfer efficiency

Following CRAH idle power consumption, the CRAH heat transfer efficiency is also dependent on the cooling system equipment. Therefore, for an unknown system configuration, it may assume a random value. Despite this random behavior, this efficiency commonly varies from 70% to 90% in real systems. As a result, η_{heat} can be modeled as a uniformly distributed random variable as shown in (14),

$$\eta_{heat} = Unif([0.7, 0.9]) \quad (14)$$

C. Generation of Load Consumption Scenarios

Data center users, in reality, can request every kind of service at any time and they may use the applications during random periods. Aligned with the requests, data center equipment parameters can vary from one facility to another. As a result, the power demand presented by the venture should be modeled considering a combination of random variables. The methodology proposed in this paper is to use the Monte Carlo simulation approach to create power demand scenarios. It creates several different scenarios with the variable parameters for each instant in order to analyze all the possible cases, and this information is later aggregated to obtain overall insights into the data center behavior and its impact on the power system and on other customers.

One scenario, for example, represents the server average utilization of 15% when the CRAH unit demands an average 7% of the server maximum consumption to run in idle state with a heat transfer efficiency of 80% at the instant of 08:15 am. Once the scenario is established, all the power demand of all data center components becomes deterministic and known.

In time series simulation, each time stamp is represented by a snapshot. Therefore, for each instant, several random scenarios are created to represent the possible data center consumption. The possible scenarios are not fully random but skewed according to the probability functions chosen to represent each random variable. The server utilization, for example, is more likely to adopt low values, limited from 0% to 100%, with an average of 15%, while the CRAH unit idle demand can be 5% to 10% of the total server power with a heat transfer coefficient varying from 70% to 90%. By considering all the probabilistic factors, various cases can be established.

The above multi-scenario study can be performed for each 15-minute interval over a month period. A total of 2688 to 2976 snapshots (depending on the month simulated) are produced. In order to analyze all the simulated cases, the probabilistic results are condensed in three groups. The first group features the 5th percentile of demand values (lowest power demand). The second group corresponds to the average power demand of all

979-8-3315-8138-1

scenarios. The last group corresponds to the 95th percentile of power demands (highest power demand). Considering all these assumptions, preliminary studies had shown that 1500 random simulations are enough to cover the possible scenarios and establish the data center power consumption.

III. Model Validation

This section presents comparisons of results provided by the proposed model with results found in literature and with real measurements from six Brazilian data centers.

A. Single Facility Simulation

The developed model was applied to represent a colocation data center located in southeast of Brazil. The facility has a peak total demand of 27 MW and is connected to a high voltage distribution system through a private substation. To validate the proposed model, simulation results are compared with real active and reactive power demand measurements at the data center obtained for a month period. The proportion of power demand due to each data center component obtained with the proposed model is shown in TABLE II, where it is also compared with typical values from real facilities encountered in literature. Results confirm that the demand composition for each component in the model is in good agreement with values reported in literature.

TABLE II. Power consumption proportionality in the model and other references

Load component	*Model*	*[5]*	*[6]*	*[7]*	*[8]*
IT load	53%	45%	26%	55%	56%
Power conditioning system	8%	8%	11%	7%	8%
Cooling system	28%	38%	50%	31%	30%
Lighting	2%	1%	3%	-	1%
Other	9%	8%	10%	6%	5%

Furthermore, the model application results in the data center load shape for the tested period. Figure 3 shows the data center real (measured) power demand, and the estimated instant power demand during September 2024 considering the 5% lowest, 95% highest and average demand scenarios.

The four curves outline the same general behavior pattern. Therefore, the developed model can estimate the time-dependent power demand profile, capturing the high and low demand periods. In addition, one can note that the average

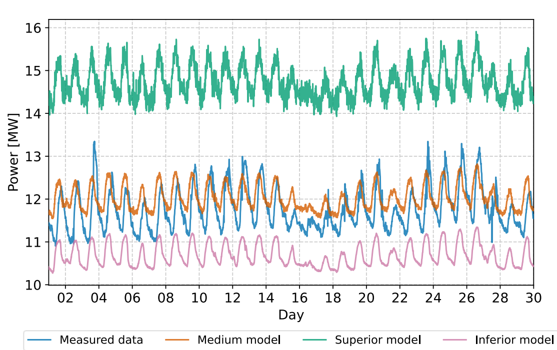

Figure 3. Data center real and estimated instant power demand.

demand indeed matches real measured demand, and the lower (5th) and upper (95th) demand percentiles can effectively represent extreme limits of the actual data center power demand. If desired for a specific purpose, other percentiles (such as 90th or 80th) could still be used to provide results closer to the upper limit of the actual measurement. Overall, this result indicates that the proposed model properly represents the data center time-dependent demand profile.

B. Model Quantitative Validation

For quantitative validation, the described model was applied to six different colocation data centers located in Brazil's Southeast region. For each facility, measured active power values from its substation were collected for a period ranging from October 2023 to August 2025 at every 15 minutes. The instantaneous power for each venture was estimated by the developed model through a Monte Carlo study. The simulations were conducted considering real measured temperature data at each plant's location. As input variable, the maximum servers power was set as 75% of the 95th percentile of measured monthly power demand, so spikes from measurement errors are not considered. For each power demand value, the percentage error (PE) was calculated as shown in (15). The percentage error from inferior (5th percentile), medium (average) and superior (95th percentile) models were aggregated at the histogram shown in Figure 4.

$$PE = \frac{P_{estimated} - P_{measured}}{P_{measured}} \cdot 100\% \qquad (15)$$

The model comparison was also verified by the mean absolute percentage error (MAPE), calculated for the three results of the proposed model and shown in TABLE III.

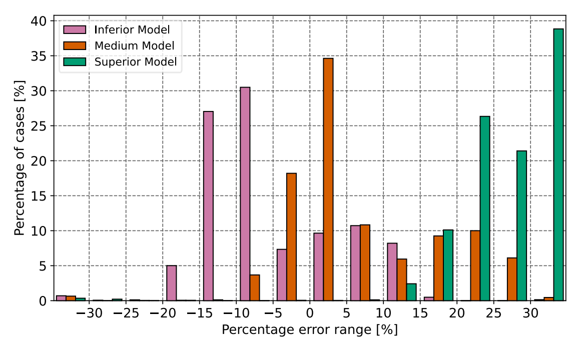

Figure 4. Histogram of models percentage error

TABLE III. MAPE for instantaneous power demand estimation

Model	*MAPE*
Inferior	9.45%
Medium	9.05%
Superior	31.45%

The medium model formulation presents the most accurate results, with a MAPE of 9.05%, followed by the inferior formulation with a 9.45% MAPE. The least accurate model was the superior one, with a MAPE of 31.45%. This superior model should be used when utility engineers are investigating more conservative (stressed) operating scenarios.

979-8-3315-8138-1

The percentage errors presented demonstrate the model precision and accuracy in power demand estimation for colocation data centers, especially the medium formulation.

C. Monthly Energy Estimation

Another application of the developed model is to estimate monthly energy consumption based on a single day measurement. The methodology was applied to the same six data centers previously described but considering the maximum power demand of the servers is 75% of the 95th percentile of the total measured power demand at the first day of the month.

The Monte Carlo study already described was conducted and the monthly energy consumed by each facility was estimated by the medium formulation and compared with the real energy consumption available at the utility commercial system. Figure 5 presents a comparison of the estimated and measured monthly energy consumption from all tested facilities with an error margin of 10%.

The MAPE from this estimate is 5.06%, which indicates that the proposed approach can also be used for estimating the energy consumption of future months.

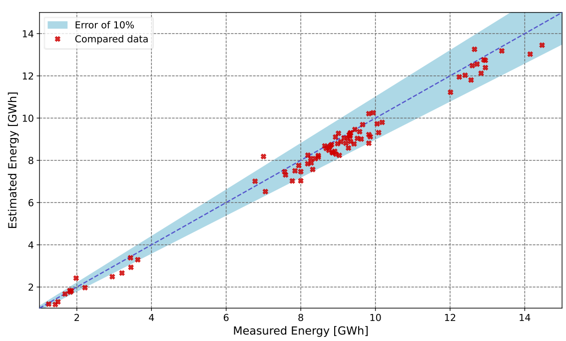

Figure 5. Monthly Energy estimated and measured comparison.

IV. CONCLUSION

This paper presents a complete power demand model for colocation data centers. The formulation is based on the modular analysis of the facility components and the interdependence among the load elements. Different from existing works, this one considers the probabilistic nature of the infrastructure parameters, allowing the power estimation for a wide range of colocation sites taking as input data the maximum server power and the external site temperature. Based on the analysis, the IT load and the cooling systems are the main power consumers, responsible for over 80% of the power demand. This result implies that these systems are the most critical on energy management and need to be analyzed in detail in power system planning and operation studies.

The ambient temperature at the site presents a significant impact on the power demand profile, being a major factor responsible for the daily variation of the load. In addition to the external temperature, server utilization is also responsible for the demand variation.

Overall, the developed model has presented solid power demand estimation validated with real demand measurement data. The mean absolute percentage error from the medium model simulation was about 9% for the instantaneous demand profile and 5% for the monthly energy consumption values for the six real sites evaluated. These results demonstrate that the proposed approach properly predicts both the demand profile and monthly energy consumption. Therefore, the developed probabilistic power demand model for colocation data centers can help planning and operation engineers assess their impact on the electrical system and on other loads.

It is worth noting that the proposed model is applicable to general colocation data centers and relies on multiple constants that segregate the consumption of each data center subsystem. Thus, an important direction for future work is to develop an approach to determine these constants based on information available in practice, so that the model can be fine-tuned to provide better demand estimates for specific data centers.

V. REFERENCES

[1] IEA, “Energy demand from AI – Energy and AI – Analysis - IEA,” IEA, 2025. https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai

[2] Y.-K. Wong, “How Many Data Centers Are There and Where Are They Being Built?,” *Abiresearch.com*, Jul. 16, 2024.

[3] “Process Panel – Access to the Basic Network MME,” Ministry of Mines and Energy, 2025. https://www.gov.br/mme/pt-br/assuntos/secretarias/sntep/acesso-rede-basica/painel-acesso-a-rede-basica-mme (Accessed Sep. 16, 2025) (in Portuguese).

[4] Ministry of Development, Industry, Business and Services (MDIC) and Brazilian Agency for Industrial Development (ABDI), 'Strategy for the Implementation of Public Policy for Data Center Attraction,' Brasília, Brazil, 2023" (in Portuguese)..

[5] K. M. U. Ahmed, M. H. J. Bollen and M. Alvarez, "A Review of Data Centers Energy Consumption and Reliability Modeling," in IEEE Access, vol. 9, pp. 152536-152563, 2021.

[6] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 732-794, Firstquarter 2016

[7] R. Rahmani, I. Moser, and M. Seyedmahmoudian, ‘‘A complete model for modular simulation of data centre power load,’’ 2018, arXiv:1804.00703.

[8] K. M. U. Ahmed, J. Sutaria, M. H. J. Bollen and S. K. Rönnberg, "Electrical Energy Consumption Model of Internal Components in Data Centers," 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, 2019, pp. 1-5.

[9] X. Ma, C. Zhang, S. Li and X. Yang, "A Data-Driven Based Energy Consumption Modeling for Heterogeneous Servers in Data Centers," 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), Taiyuan, China, 2021, pp. 3109-3114.

[10] D. Meisner, B. T. Gold, and T. F. Wenisch, "PowerNap: eliminating server idle power," ACM SIGARCH Computer Architecture News, vol. 37, no. 1, pp. 205–216, Mar. 2009.

[11] S. Itoh, Y. Kodama, T. Shimizu, S. Sekiguchi, H. Nakamura, and N. Mori, ‘‘Power consumption and efficiency of cooling in a data center,’’ in Proc. 11th IEEE/ACM Int. Conf. Grid Comput., Oct. 2010, pp. 305–312.

[12] J. Salom, L. Siso, and M. Sansigre, “Cost-effective renewable cooling on green data centres,” Citeseer, Tech. Rep., 2014.

[13] C. Zaloumis, “Are your data centers keeping you from sustainability,” Ibm.com, Jun. 07, 2022.