Pesquisa Operacional (2023) 43: e271290 p.1-28
doi: 10.1590/0101-7438.2023.043.00271290

Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
ARTICLES

PERFORMANCE ANALYSIS OF A BRAZILIAN CALL CENTER WITH IMPATIENT CUSTOMERS USING M/G^c/1+G AND M/G/c+G QUEUING MODELS

Sidney Carlos Ferrari[1] and Reinaldo Morabito[2*]

Received January 19, 2023 / Accepted January 24, 2023

ABSTRACT. This study analyzes performance measures in a Call Center with impatient customers, who may leave the system before being served. The abandonment of the waiting line is a measure of the subjective perception of the quality of the services offered, being one of the main concerns of the managers of Call Center systems. Based on data extracted from the case study of a Brazilian company, it is shown that the analytical queuing models M/G^c/1+G and M/G/c+G, with generic distributions to represent service times and generic distributions (particularly mixed distributions) to represent patience times, can be effective to evaluate the congestion problem of this Call Center. The sampling performance measures of this case are compared with the measurements obtained through the M/G^c/1+G and M/G/c+G models, using non-mixed and mixed distributions based on Exponential, Fatigue Life and Normal distributions to represent customers' patience. We are not aware of other studies in the literature in this line of research. The results indicate that, in general, the use of analytical queuing models with abandonment, exploring more elaborate distributions to model service times and mixed distributions for patience times, are more accurate in evaluating this Call Center than other queuing models with abandonment, such as the M/M/c+M, M/M/c+G and M/G/c+M.

Keywords: call center, contact center, impatient customers, queuing models with abandonment, mixed distributions, case study, congestion analysis.

1 INTRODUCTION

Call Centers (Contact Centers) have been the object of several pieces of research in recent decades. Through Call Centers, organizations aim to better know their customers and offer what they need, satisfying their expectations. They have been analyzed, predominantly, from the quantitative point of view, using appropriate queuing models to represent them (Mandelbaum and Zeltyn, 2005). A concern in the modeling of these systems is to incorporate the customers' impatience with the purpose of making these models more adjusted to the reality of the Call Center.

*Corresponding author
[1]Federal University of São Carlos, Department of Production Engineering, São Carlos SP, Brazil – E-mail: ferrarisc@gmail.com – http://orcid.org/0000-0001-9434-9323
[2]Federal University of São Carlos, Department of Production Engineering, São Carlos SP, Brazil – E-mail: morabito@ufscar.br – http://orcid.org/0000-0002-3948-305X

Call Center users always have a limited amount of time to wait (their patience time) until they are served. When they are not served during this time, they leave the system, getting no response to their requests. Such abandonment becomes an important operational performance measure because users after waiting for service and leaving the system before being served subjectively declare that it is not worth waiting for the services offered. On the other hand, customers who do not leave the system only express their perception about the services offered if they are asked after the end of the service.

Some measures of performance of a Call Center can be evaluated through stochastic models of queuing theory. In these models, the users are people who call the system to solve a problem or to obtain information, the servers are the agents of the system that serve the users (or, still, they are communication equipment) and, finally, the queues are composed by users who wait to be served. Incorrect sizing of the number of agents can increase the waiting time in the queues, causing the phenomenon of user abandonment. Therefore, not considering abandonment produces distortions in the system information that are important for decisions to be taken by managers, in order to achieve a quality level of service subject to a specified budget. An important performance measure in these systems is the probability of the user leaving the system and this occurs when the user hangs up the telephone (or the computer) before being served by an agent. Other measures associated to the abandonment process and also considered important are the waiting time (measured by its average) and the probability of waiting, both related to the time the user waits to be served (Lima, 2012).

Several queuing models incorporating abandonment have been studied in the literature (Bacelli and Hebuterne, 1981; Whitt, 1993; Mandelbaum et al., 2001; Brown et al. 2002; Mandelbaum and Zeltyn, 2005), but these models consider a single-family class of parametric distributions to model abandonment, the most common being the Exponential distribution, Weibull and Erlang. However, due to the distinct characteristics of the user population found in the Call Center, generally composed of a mixture of groups with different characteristics, each one having its own pattern for the time of patience, it may be interesting to use mixed distributions of probability to better absorb these population nuances. A mixed distribution (or mixture), according to Forbes et al. (2011), is understood as a probability distribution that is a combination of two or more specific probability distributions, such as a Normal distribution and an Exponential distribution.

These mixed models can represent the presence of subpopulations within a general population (Oliveira, 2009; Mandelbaum and Zeltyn, 2012). Subpopulations can be represented by groups of customers who have their own pattern for patient time, forming a population with a mixture of these groups with different characteristics. These considerations justify the concern to consider the abandonment in the mathematical modeling of the Call Center and to analyze the patterns of the time of abandonment of the customers and their implication in the performance measures of the system. Another feature observed in real data extracted from Call Centers is the fact that service times in general do not fit the Exponential distribution, traditionally used in the analytical queuing models to represent Call Centers. In this case, analytical models with a generic distribution type M/G/c+G to model the service times of their users seem to be more adequate to represent call centers than M/M/c+M and M/M/c+G models.

Based on these considerations, an interesting research topic to improve the performance of M/G/c+G queuing models used in Call Center sizing is to investigate and use mixed probability distributions to model abandon times of users. In model M/G/c+G, M indicates that the arrival process is Poisson, c is the number of servers, the first G indicates that the duration of each service has a generic distribution and the last G, that the patience times also have generic distributions. When using a mixed distribution to represent the patience times in this model, it is assumed that the population can be heterogeneous, composed by two or more independent subpopulations, but that it is not possible to distinguish individuals of different types in the population, and the patience times of all individuals are considered independent and identically distributed (iid) random variables.

In this article, a research question to be investigated is whether the analytic queuing models M/G/c+G and M/G^c/1+G (the latter being a single-server system, but with the service rate of the server increased in c times), with generic distributions to model the service times and incorporating user abandonment based on generic distributions (particularly mixed) to model the patience times, are effective to represent the congestion problem in the Call Center of a case study. This article does not particularly deal with formally developing new mathematical models of congestion, which are already relatively well developed by queuing theory. Our main concern is to study this subject in depth and to make an empirical complement to this theory, through scientific analysis based on real data, by a case studied in a Brazilian company's Call Center (which develops specialized software for the management of pharmacies and drugstores), in order to compare the performance of theoretical models M/G^c/1+G and M/G/c+G. In this case study, an experimental model of discrete simulation, appropriately representing the company's Call Center system, is also used to compare and verify the accuracy of the analytical models used in the original and alternative configurations of the system, considering variations in the number of attendants. The probabilities of abandonment obtained by the simulation and the analytical queuing models are compared and the level of service is obtained, with the purpose of evaluating the sensitivity of the models. The results indicate that these models can be useful for Call Center managers to better configure their systems and manage their operations, as well as more effectively plan the future.

These studies have been little explored in the literature and it is understood that they can bring knowledge advances by means of the evaluation and validation of queuing models with abandonment for application in Call Centers, incorporating generic distributions to model service times and mixed generic distributions to represent the patience of the users. Abandonment in the existing literature, in general, has been modeled through a single class of probability distribution. It has been little discussed from the point of view of mixed probability distributions, which may be more adequate to capture the heterogeneity of each customer's time of patience.

Our aim in this experimental study is to show that the use of the analytical queuing models with abandonment, exploring more elaborated distributions to model service times and mixed distributions for patience times, is effective for the performance evaluation of this Call Center.

This article is structured as follows. In Section 2, we present a brief review of the literature on the M/G^c/1+G and M/G/c+G queuing models for Call Centers. In Section 3, we describe the case of the Call Center of the company, object of study of this work, presenting in detail the system, the statistical analyzes, the analysis of the processes of user arrival, service and abandonment of the data collected. In Section 4, we present the analysis of the results obtained and the evaluation of an alternative scenario. Finally, in Section 5, the conclusions of the study are presented and some perspectives for future research are discussed.

2 CALL CENTER QUEUING MODELS

Some analytical queuing models are briefly reviewed in this section. Queuing models with abandoning users are reviewed, especially those that consider a generic distribution for patience times. Another aspect to be considered is the fact that the service times of the users in some Call Centers do not fit perfectly with the Exponential model, leading to the more general system, with a generic probability distribution also for the service times, which can offer better performance to approach the reality of Call Centers. Motivated by the real Call Center data of the company being studied, the queuing models that have service times adjusted by a generic distribution and the patience times also modeled by a generic distribution, particularly mixed, have shown more accuracy to represent the Call Center and were explored in this article.

The abandonment process, described through mixed distributions, has been studied in the literature, proving to be advantageous in relation to those that were not described by mixed distributions. Oliveira (2009) and Mandelbaum and Zeltyn (2012) used several mixed distributions to model the abandonment process that obtained a better fit for patience time than non-mixed distributions. However, to the best of our knowledge, there are no experimental studies in literature that consider mixed distributions into analytical queuing models with abandonment, and that evaluate the results of these models with the actual data extracted from a Call Center. A recent related study can be found in Ferrari and Morabito (2020), but considering exponentially distributed service times (M/M/c+M and M/M/c+G models) rather than generic service time distributions.

Customers' patience times are censored data and can be estimated using Survival Analysis (Cox and Oakes, 1990). In this sense, Palm (1953) estimates the hazard rate of patient customers using a competitive hazard model. The same author also postulates that the hazard rate of the time willing to wait is proportional to the customer's irritation due to waiting. Still in this sense, Aalen and Gjessing (2001) warns that the hazard rate of the population may not represent that of the individuals. It examines the validity of a queuing theory rule that relates the average waiting time in the queue to the probability of abandonment. This "rule" is inferred in Baccelli and Hebuterne (1981) and Zohar et al. (2002). It applies to models with exponentially distributed patience times. However, this rule holds for the analyzed data whose patience distribution is non-exponential. A theoretical explanation of these empirical observations is also sought.

In the development of this study, we used the M/G^c/1+G and M/G/c+G queuing models of the literature to represent the case of the company's Call Center. The choice for these queuing models is justified by the best adjustment of non-exponential distributions to represent service times and patience times, obtained through the statistical analysis applied to the real data of the object of study considered. To make this article more self-contained, the following is a summary of the important results of these queuing models of the literature, containing a brief presentation of the characteristics and performance measures that incorporate the patience time in their equations.

2.1 The M/G^c/1+G queuing model

In this queuing system with a single service channel and with the possibility of the customer leaving the queue before being served, it is assumed that the arrivals to the system occur according to a Poisson process with arrival rate λ and with a FIFO (First In, First Out) attendance discipline. The random variable B represents the service times with a generic distribution function $B(x)$. In the same way, the random variable R denotes the times to abandonment (patience), following a generic distribution $R(x)$. It is considered that both random variables are iid and also that they are independent of each other and also of the arrival process. The equations of this queuing model, defined by Iravani and Balcioglu (2008) and used in this work to obtain the results of the system performance measures, are:

Probability of the system being empty:

$$P_0\left\{1+\Delta\left[\frac{f_1(0)+f_1(MAX\Delta)}{2}+\sum_{n=1}^{MAX-1} f_1(n\Delta)\right]\right\}=1 \tag{1}$$

Probability of abandonment:

$$P_A = 1-\frac{1-P_0}{\rho} \tag{2}$$

Waiting probability of the customers served:

$$P(0<W_S\leq x)=\frac{\Delta P_0}{1-P_A}\left\{\frac{f_1(\Delta)\overline{R}(\Delta)+f_1(n_x\Delta)\overline{R}(n_x\Delta)}{2}+\sum_{n=2}^{n_x-1} f_1(n\Delta)\overline{R}(n\Delta)\right\} \tag{3}$$

Average waiting time of the customers served:

$$E\left[W_S^k\right]=\frac{\Delta P_0}{1-P_A}\left\{\frac{\Delta^k f_1(\Delta)\overline{R}(\Delta)+(MAX\Delta)^k f_1(MAX\Delta)\overline{R}(MAX\Delta)}{2}+\sum_{n=2}^{MAX-1}(n\Delta)^k f_1(n\Delta)\overline{R}(n\Delta)\right\} \tag{4}$$

where $\rho=\frac{\lambda}{\mu}$ (μ is the service rate) and $\bar{F}(t)$, $\bar{R}(t)$ are the complementary distribution functions of the virtual waiting time and abandonment, respectively. The remaining notation of Equations (1)-(4) and further details of the mathematical developments to obtain these equations are presented in Appendix A (more details can be obtained in Iravani and Balcioglu (2008), Ferrari

(2016) and the references therein). Please note that when using this single-server model to represent the system of the company's case study with c servers, we consider the service rate of the server increased in c times.

2.2 The M/G/c+G queuing model

This multiple-server system with abandonment, with c identical servers operating in parallel and independently of each other, receives customers arriving according to a Poisson process with arrival rate λ and service times distributed according to a generic distribution. At the moment of the customer arrival, if he/she finds at least one of the servers free, he/she is served immediately; otherwise, he/she waits in a queue. The service time is described by a random variable B with generic probability distribution function $B(x)$. While waiting in the queue, the customer can abandon the system with a generic probability distribution function $R(x)$. It is assumed that variables B and R are independent of each other and also of the arrival process. Iravani and Balcioglu (2008) proposed the following approximations for the performance measures of this system, which are used in this work:

Intensity of traffic:

$$\rho_{M/G/c+G} = \rho_{M/G^c/1+G} = \frac{\lambda}{c\mu_B} \tag{5}$$

Waiting probability:

$$P_{M/G/c+G}\left(W_S \leq x\right) \cong \frac{P_{M/M/c+G}\left(W_T = 0\right)}{1 - P_{M/M/c+G}\left(A\right)} + \left[1 - \frac{P_{M/M/c+G}\left(W_T = 0\right)}{1 - P_{M/M/c+G}\left(A\right)}\right] \frac{P_{M/G^c/1+G}\left(0 < W_S \leq x\right)}{1 - \frac{P_0}{1-P_A}} \tag{6}$$

Average waiting time:

$$E_{M/G/c+G}\left[W_S^K\right] \cong \left[1 - \frac{P_{M/M/c+G}\left(W_T = 0\right)}{1 - P_{M/M/c+G}\left(A\right)}\right] \frac{E_{M/G^c/1+G}\left[W_S^K\right]}{1 - \frac{P_0}{1-P_A}} \tag{7}$$

Probability of abandonment:

$$P_{M/G/c+G}\left(A\right) = P_{M/M/c+G}\left(A\right) \tag{8}$$

The remaining notation of Equations (5)-(8) and some details of the mathematical developments to obtain these equations are presented in Appendix B (more details can be obtained in Iravani and Balcioglu (2008), Ferrari (2016) and the references therein). The justification for using the M/G^c/1+G and M/G/c+G models to represent the Call Center of the company is motivated by the best fit of non-exponential distributions to represent the service and patience times, obtained through the statistical analysis applied to the data of the case study considered.

In this study we analyze two situations not yet explored in the literature: the use of queuing models with generic distributions for service times, mixed distributions for patience times, and validation of these models with analysis based on data extracted from a Call Center. This empirical quantitative approach is justified by the studies and analyzes carried out in the company Call Center, observing that the patience distributions are generic, with a strong potential for mixed distribution applications. The obtained parameters and probability distributions were inserted into the analytical queuing models of the literature, and also in the discrete simulation models considered in this study. The same performance measures were calculated and compared using these queuing and simulation models, and these measures were also compared with those extracted from the sample.

3 THE COMPANY CALL CENTER

The Call Center considered in this article serves a company that develops specialized software for the management of pharmacies and drugstores, located in the interior of the State of São Paulo, Brazil. This Call Center offers their customers the following types of services: (i) technical support for customers to solve problems related to the software; (ii) virtual sale for customers who wish to buy company products; (iii) customer service with problems in contract; (iv) services related to monthly payment, bank tickets and other financial services. In this study, only the technical support service was considered. The other types of services have low flow of calls and little regularity, being disregarded. The technical support service is divided into telephone service, when the customer calls the Call Center, and online service, where the customer is served through the internet, through a chat. The Call Center has 15 service positions, 14 of which are for technical support, working during business hours from Monday to Saturday. Customers accessing the company's Call Center connect to the Voice Response Unit (IVR), providing their data for identification and choosing one of the service options.

There are an average of 4,832 monthly calls that choose the option of technical support to solve their problems. These customers join the waiting line to dialogue with the attendants, and an average of 434 (8.98%) of them leave the system after waiting a positive amount of time in queue. The remaining 4,398 (91.01%) solve their problems with the attendants and leave the system, with 0.11% of the customers having their services closed by the attendant, without solving their problems. Those waiting in queue receive information about their position in the queue and the estimated amount of time to be served. These messages are repeated every 1 minute, interspersed with songs and commercials of products sold by the company. Figure 1 summarizes the flows of calls related to the technical support service, whose values represent the minimum, average and maximum amounts of monthly calls. In this study, the processes of arrival, abandonment and service are analyzed from the moment the customer enters the waiting queue, after being screened in the IVR. Therefore, the arrival, abandonment and service completed in the IVR are disregarded.

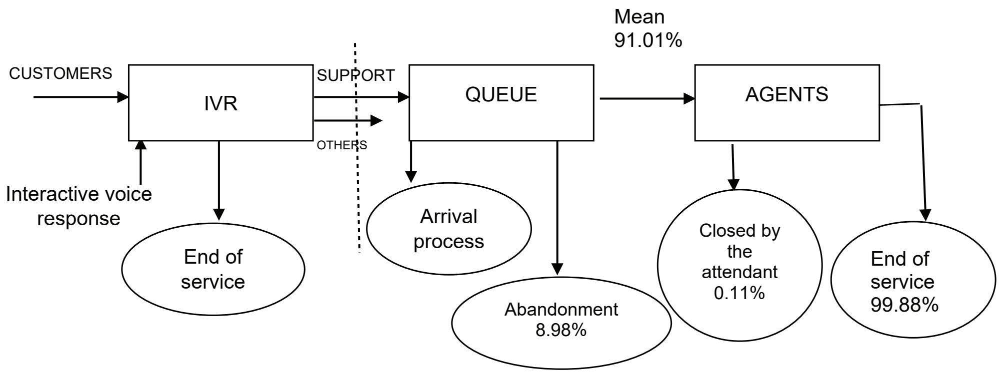


Figure 1 – Call Center Flow in the company.

3.1 Data description

As data source, the records of each call are considered, containing their identification and the time instant when occur each point in the call flow. These data were collected and analyzed during a year to verify typical months, days and times and peak operating. They were extracted from the Call Center management system that records the details of each call and were organized into a worksheet, where each line corresponds to the record of a call. A careful analysis was performed to detect and correct inconsistencies. The intention is to initially reproduce the results of these data with the analytical queuing models M/G/c+G and M/G^c/1+G, and then also generate new results and compare them. This data corresponds to all calls answered, referring to the type of service "Technical support". The month of April was chosen as a typical operation month and, in this month, the information was considered in the analyzes on April days that had the highest number of visits and higher number of abandonments, from 12 noon to 4:00 p.m. This time interval was chosen because it had a large number of consecutive calls (212 on average), corresponding to almost half of the calls of the day. A detailed description of such information can be obtained from Ferrari (2016).

3.2 Arrival process analysis

As the arrival process is random, it is necessary to make an analysis to verify the behavior of the time distribution between arrivals that occur in the company's Call Center. This process records the times when online calls arrive in the waiting line after they have visited the IVR. After leaving the IVR, customers enter the queue and, if they find an unoccupied attendant, the service starts immediately. In this case the queue time is zero. If there are no available attendants, the customer waits a positive amount of time in the queue until starting service, or even leaving the system, due to impatience.

This process is evaluated by means of descriptive statistical information and also by its stochastic variability. The time between the consecutive arrivals of the online calls that waited until starting the service or that left the system was taken to evaluate the process. The statistics of the times between the online arrivals of the company's Call Center during 12-4 pm are summarized in Table 1. The Kolmogorov-Smirnov adherence test also confirmed the approximation of the time between the arrival of online calls with the Exponential distribution of probability, with p-value of 0.43675 (α = 5%), according to results obtained by the EasyFit software, which ordered the most adherent distributions to the data.

3.3 Service process analysis

The Call Center service process of the company is characterized by the time spent by the attendant to make each of the calls online. This process was investigated quantitatively and the service times were obtained from the real data. Considering the same time interval analyzed previously, from 12-4 pm, the mean, median and standard deviation of these service times were then obtained, reported in Table 2. The service times recorded equal to zero were ignored.

Table 1 – Time between arrivals (min).

Sample data	12-4 pm
Mean	1.3665
Median	1.0000
Standard deviation	1.4584
Coef. of variation	1.0673
Arrival rate (call/min)	0.7318

Table 2 – Service time (min).

Sample data	12-4 pm
Mean	30.9238
Median	12.5167
Standard deviation	31.3813
Coef. of variation	1.0148
Service rate (call/min)	0.0323

The Kolmogorov-Smirnov test applied to these service times rejected the Exponential distribution (p-value 0.0016). Then, a Weibull distribution was chosen, with 3 parameters and p-value 0.0034, to represent the service times of the company's Call Center, considering that it was the first adherent distribution in the list ordered by the EasyFit software (the Exponential distribution was only the seventh in this list).

4 ABANDONMENT PROCESS ANALYSIS

The abandonment process occurs when customers of the company's Call Center online calls that are in the queue wait for a positive amount of time and then leave the system. When abandonment occurs, it is not possible to observe the time required for the user to be served. So, this time needed to be served and the patience time were based on censored data. None of these measures was directly observed and therefore must be estimated. For a better understanding, Brown et al.

(2002) define the virtual waiting time as the amount of time that a customer who has infinite patience (virtual customer) has to wait until being attended. They also define the time of patience, or the operational measure of patience (or impatience) of the customer, as the amount of time that a user is willing to wait before leaving the system. These definitions are considered in the analyzes of this work. To estimate those times that are censored and therefore not observable as the customers' patience time, the Kaplan-Meier estimator was used, a Survival Analysis technique (Kalbfleisch and Prentice, 1980). The Kaplan-Meier test was applied to the censored data to infer the average patience time.

In order to analyze the experiences that the customers had in the queue until abandoning the company's Call Center system, the patience times of the customers, calculated with the difference between the values recorded in the variables *hour entry* and *hour end*, were considered if the variable *hour start* displayed no data record. The *hour entry* variable assumes the values of the instant of arrival of the online call in the queue. The values of these differences were used to calculate the mean, median and standard deviation summarized in Table 3. The Exponential distribution can be suggested to represent the patience times because it has a median less than the mean and a coefficient of variation very close to 1 (Table 3). According to the Kolmogorov-Smirnov test applied to the time of abandonment data, the Exponential distribution, the Fatigue Life distribution, and the Normal distribution were not rejected at the 5% probability level by the EasyFit software with p-values 0.8248, 0.5538 and 0.1866, respectively. These three distributions were adopted to represent the patient online call waiting times of the company's Call Center.

Table 3 – Patience and time (min).

Statistics	**12-4 pm**
Mean	55,9089
Median	81.0833
Standard deviation	21.1027
Coef. of Variation	0.3774
Abandonment rate(call/min)	0.0179

As the interest here is to estimate the patience of the company's Call Center customers, i. e., how long the customers are willing to wait before abandoning, then one must consider the time it takes for a user to reach an agent as a censored observation. In fact, if the customer reaches the attendant, it is because the time he/she was willing to wait was longer. Considering the online calls that reached the agent as censored observations, the users' patience was estimated adopting the Kaplan-Meier estimator, using SPSS® software. The mean, median, and standard deviation of the patience time estimates for the company's Call Center online calls are reported in Table 4, and Figure 2 shows the Kaplan-Meier estimates of the patience time survival function of company's Call Center users.

After knowing the descriptive statistics, rates, and probability distributions of the processes of arrival, service and abandonment of online calls from the company's Call Center, the following performance measures were obtained by real data: the average waiting time, the probability of waiting, the probability of abandonment and the traffic intensity. These measures were compared with the results of the same measurements obtained through the analytical queuing models and experimental simulation models that represent the company's Call Center in this study.

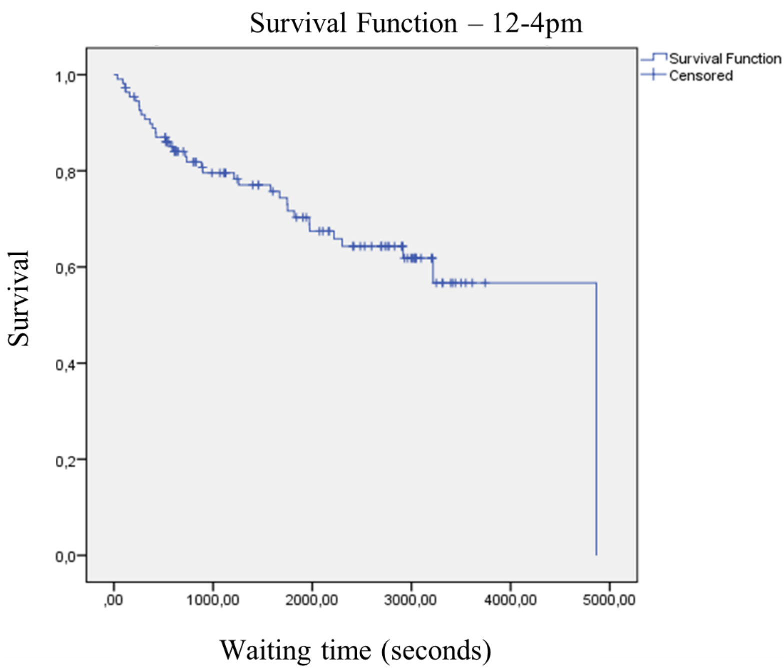


Figure 2 – Survival curve.

Table 4 – Patience time estimated by Kaplan-Meier.

Sample data	12-4 pm
Mean	18.5038
Median	12.0333
Standard deviation	18.0332
Coef. of variation	0.9999
Abandonment rate(call/min)	0.0540

The average waiting time was obtained by calculating the average of the waiting times in the queue of online calls that were served by an attendant or left the system. The probability of waiting was calculated with the quotient between the sum of the number of *online* calls that waited in the queue and were answered, and the sum of the number of *online* calls of the system that were answered or abandoned the system. The probability of abandonment was also obtained through a quotient between the number of *online c*alls that waited in the queue and abandoned the system, and the number of online calls that waited in the queue and were answered or abandoned the system. The traffic intensity was obtained by dividing the average number of attendants providing service by the number of service stations. The results of all these measures of performance obtained through real data are shown in Table 5. These values are quite high, especially the average waiting time (29.0418 min) and the probability of waiting (0.7973). This fact suggests that the probability of abandonment is closely related to the waiting time in the queue.

Table 5 – Performance measures – 12-4 pm.

Performance measures	Valor
Medium waiting time	29.0418 (min)
Waiting probability	0.6818
Abandonment probability	0.3182
Traffic intensity	0.9333

In order to provide an overview of the probability distributions most closely adhering to the arrival, service, and abandonment process data of the company's Call Center (within the analyzed range from 12-4 pm), a summary of these is shown in Table 6. These distributions were used in the analyses of this work.

Table 6 – Summary of the probability distributions.

Process	Adjusted distribution	*p-value* 12 – 4 pm
Arrival	Exponential	0.43675
Service	Weibull	0.00340
	Exponential	0.00160
Abandonment	Exponential	0.82480
	Fatigue Life	0.55380
	Normal	0.18860

5 RESULTS ANALYSIS

After applying the tests of adherence to the company's Call Center data at a 5% probability level (Table 6), it was obtained that the Exponential probability distribution is more adjusted to represent the process of arrival of this Call Center, with p-value 0,43675. Regarding the service process, all distributions analyzed were rejected. However, the Weibull distribution was the best fit among them to represent this process, with $p = 0.00340$. Several probability distributions showed adherence to the Call Center's patience times, including the Exponential distribution ($p = 0.82480$), the Fatigue Life distribution ($p = 0.55380$) and the Normal distribution ($p = 0{,}18860$), as mentioned before. Although the tests of adherence did not reject the Exponential distribution for the patience times, it is shown later in this section that the use of Fatigue Life and Normal distributions, as well as mixed distributions based on these distributions, resulted in more accurate analytical queuing models to represent the company's Call Center than using the Exponential distribution. Therefore, model M/G/c+G, with generic distributions for the service process and abandonment, was considered more appropriate to represent the company's Call Center rather than models M/M/c+M, M/M/c+G and M/G/c+M.

Several possibilities of mixed distributions were also considered in the use of these analytical models, combining the mixture of two Exponential (Ex+Ex), two Normal (N+N) and two Fatigue Life (FL+FL) components. Mixed distributions can be used in situations where the population is suspected to be heterogeneous and composed of different subpopulations. When using a mixed probability distribution in these models, similarly to when using an exponential or another probability distribution, we assume that the times to abandonment (patience) of the users are iid random variables, and also that they are independent of the interarrival times and service times of the users. While the population of users can be heterogeneous, composed by two or more independent subpopulations, it is considered in the analysis that it is not possible to distinguish individuals of different types in the population. The parameters of these mixed distributions, as well as the weights of each one, were obtained by the maximum likelihood estimation method, calculated by the Mathematica® software. Arrival, service and abandonment rates, obtained by the descriptive analysis of the data (Tables 1, 2 and 3), were inserted in these models and the performance measures were calculated. It should be remembered that the users' patience are censored values that were estimated using the Kaplan-Meier estimator, constituting a new abandonment rate that was used to recalculate the parameters of the probability distributions involved in the abandonment process, as already described in section 3.

To obtain the analytical queuing model that adequately represents the company's Call Center the Model Accuracy Index (MAI) was calculated. The MAI statistic is the sum of the weighted squares of the differences between the estimated value $(\widehat{y}_i)$ and the real value of the data (y_{i}) of all terms (n) (Equation (9)). The values of the weights (α_{i}) are assigned according to the level of importance of the performance measure that they represent. In this study, weights were assigned to 1 (α_{i} = 1) for all deviations obtained, considering that the performance measures have the same level of importance. The analytical queuing model that has the lowest MAI was considered the most efficient to represent the Call Center of the company, because the results obtained from the performance measures of this model have smaller deviations in relation to the same measures extracted from the real data.

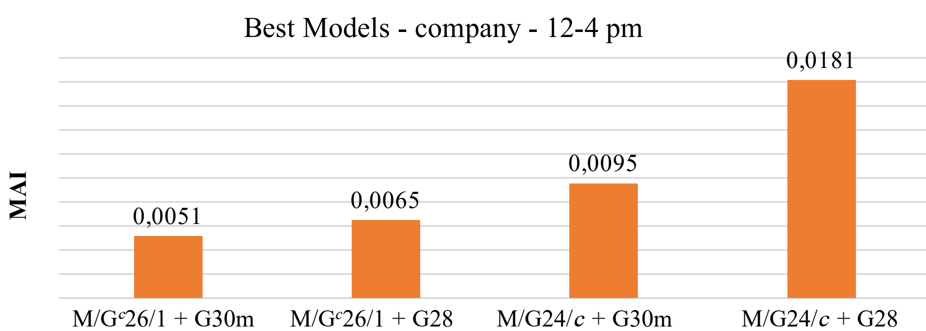
$$MAI = \frac{\sum_{i=1}^{n} \alpha_i \left(y_i - \widehat{y}_i\right)^2}{n} \tag{9}$$

In the performed analyzes, it was observed that there were analytical queuing models that produce better results in some performance measures than in others. We used, then, to find the most appropriate analytical queuing model to represent the company's Call Center, for each performance measure, the "Percentual Difference" (Dif%) metric. This metric is defined by Equation (10), where V is the value of the performance measure obtained by the analytical queuing model and R is the value of the real performance measure.

$$Dif\,\% = \frac{|V - R|}{R} \tag{10}$$

In order to facilitate the understanding of the analyzes, from this section onwards, the queuing models were represented with a specific notation. We identified the arrival and service process distributions and the generic ones in the abandonment process with a numerical index to distinguish each one from the others. When the distribution is of the mixed type, the letter "m" was added next to the numerical index to represent it. For example, G_1 identifies a single parametric distribution (e.g., $G_1 \sim$ LogN(1,2) is a Lognormal distribution with parameters 1 and 2), while G_{1m} represents a mixed distribution (e.g., $G_{1m} \sim$ 0.5LogN(1,2) + 0,5LogN(3,4) is a mixed distribution with two components Lognormal and weight $p = 0.5$). Comparing the performance measures obtained by the analytical queuing models with those extracted from the actual data, we investigated the existence of a single analytical queuing model suitable to represent the company's Call Center, using Equation (9). Figure 3 shows that the model sought is the $M/G^c 26/1+G_{30m}$, with Weibull distribution for the service times and mixed Normal distribution for the patience times.

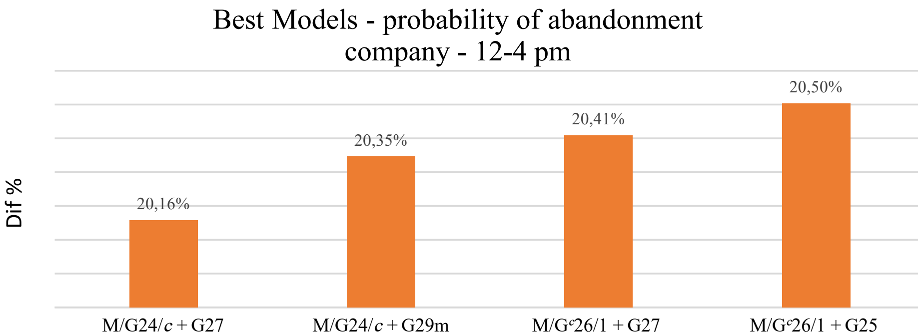
We also sought the analytical queuing model that adequately represents each of the performance measures considered in this study. The metric used in this analysis is the percentage difference of Equation (10). Thus, the performance measure probability of abandoning, as shown in Figure 4, is adequately represented by the analytical queuing model $M/G_{24}/c+G_{27}$, which has Weibull distribution for the service times and Fatigue Life distribution for the patience times. A model that was also adequate to estimate the probability of abandoning, because it has a percentage difference (Dif%) very close to the other model, and has a mixed Fatigue Life distribution for the patience times, is the $M/G_{24}/c+G_{29m}$, with Weibull distribution for the service times.



Distributions:
$G_{24} \sim$ Weibull (0.9276; 26.004; 3.1333) $G_{25} \sim$ Exponential (55.9089) $G_{26} \sim$ Weibull (1.3021; 2.3952; 0)
$G_{27} \sim$ Fatigue Life (0.3726; 52.2792) $G_{28} \sim$ Normal (30.9598; 445.3239) $M_4 \sim$ Exponential (0.0323)
$G_{29m} \sim$ 0.8418×Fatigue Life (0.4055; 43.4875) + 0.1582×Fatigue Life (0.9434; 6.1219)
$G_{30m} \sim$ 0.5221×Normal (29.1926; 232.5241) + 0.4779×Normal (26.7163; 212.7998)
$G_{31m} \sim$ 0.9673×Exponential (54.0799) + 0.0327×Exponential (1.8289)

Figure 3 – Performance of $M/G^c/1+G$ and $M/G/c+G$ models – company.

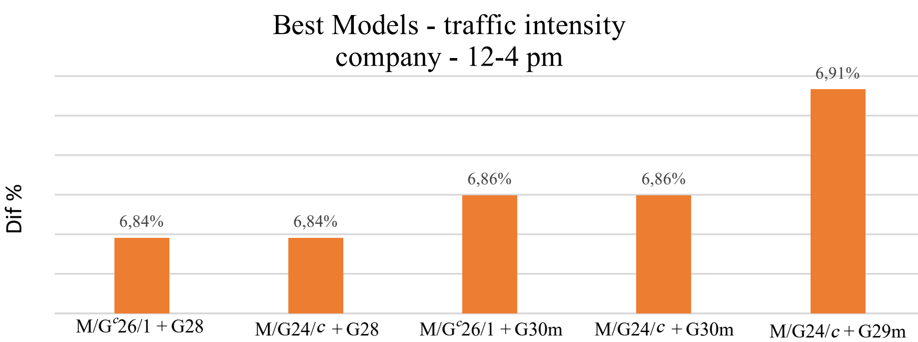
Considering the traffic intensity performance measure, Figure 5 shows that the models $M/G^c{}_{26}/1+G_{28}$ and $M/G_{24}/c+G_{28}$, with Weibull distribution for service times and Normal distribution for the patience times, are the most suitable. The probability of waiting can be adequately represented by the analytical queuing model $M/G_{24}/c+G_{28}$, with Weibull distribution for the service times and a single Normal distribution for the patience, or also by the model $M/G_{24}/c+G_{30m}$,



Distributions:
G_{24} ~ Weibull (0.9276; 26.004; 3.1333) G_{25} ~ Exponential (55.9089) G_{26} ~ Weibull (1.3021; 2.3952; 0)
G_{27} ~ Fatigue Life (0.3726; 52.2792) G_{28} ~ Normal (30.9598; 445.3239) M_4 ~ Exponential (0.0323)
G_{29m} ~ 0.8418×Fatigue Life (0.4055; 43.4875) + 0.1582×Fatigue Life (0.9434; 6.1219)
G_{30m} ~ 0.5221×Normal (29.1926; 232.5241) + 0.4779×Normal (26.7163; 212.7998)
G_{31m} ~ 0.9673×Exponential (54.0799) + 0.0327×Exponential (1.8289)

Figure 4 – Performance of M/G^c/1+G and M/G/c+G – abandonment probability.

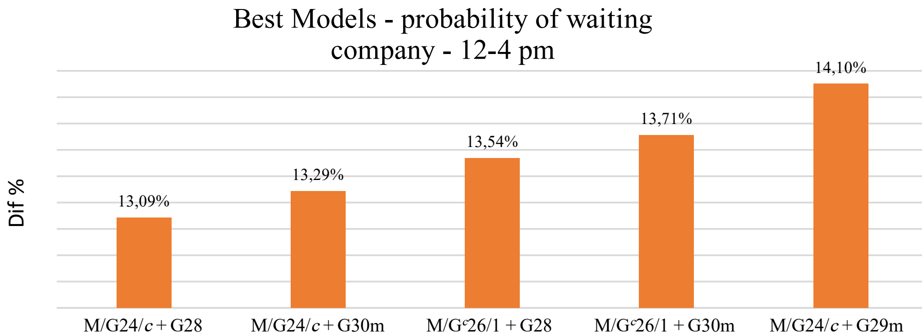
with the same distribution for service times and mixed Normal distribution for patience. These two models have very close percentage differences, as shown in Figure 6. The analytical queuing model M/$G^c{}_{26}$/1+G_{30m}, which has a Weibull distribution for the service times and a mixed Normal distribution for the patience times, is the one that best fits the performance measure average waiting time, as indicated in Figure 7.



Distributions:
G_{24} ~ Weibull (0.9276; 26.004; 3.1333) G_{25} ~ Exponential (55.9089) G_{26} ~ Weibull (1.3021; 2.3952; 0)
G_{27} ~ Fatigue Life (0.3726; 52.2792) G_{28} ~ Normal (30.9598; 445.3239) M_4 ~ Exponential (0.0323)
G_{29m} ~ 0.8418×Fatigue Life (0.4055; 43.4875) + 0.1582×Fatigue Life (0.9434; 6.1219)
G_{30m} ~ 0.5221×Normal (29.1926; 232.5241) + 0.4779×Normal (26.7163; 212.7998)
G_{31m} ~ 0.9673×Exponential (54.0799) + 0.0327×Exponential (1.8289)

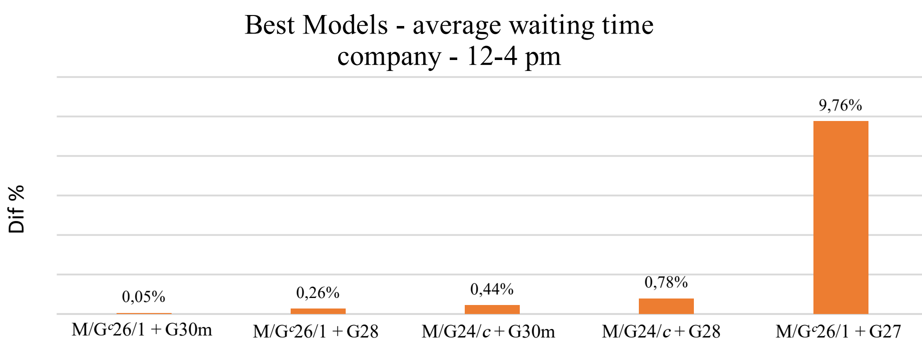
Figure 5 – Performance of M/G^c/1+G and M/G/c+G models – traffic intensity.

With the results obtained, it is observed that the analytical queuing models are adequate to represent the traffic intensity, with percentage differences lower than 7%, and the analytical queuing models are also adequate to represent the average waiting time, with percentage differences less than 1%. These models are also adequate to represent the probability of waiting and the proba-



Distribuições:
G_{24} ~ Weibull (0.9276; 26.004; 3.1333) G_{25} ~ Exponencial (55.9089) G_{26} ~ Weibull (1.3021; 2.3952; 0)
G_{27} ~ Fatigue Life (0.3726; 52.2792) G_{28} ~ Normal (30.9598; 445.3239) M_4 ~ Exponencial (0.0323)
G_{29m} ~ 0.8418×Fatigue Life (0.4055; 43.4875) + 0.1582×Fatigue Life (0.9434; 6.1219)
G_{30m} ~ 0.5221×Normal (29.1926; 232.5241) + 0.4779×Normal (26.7163; 212.7998)
G_{31m} ~ 0.9673×Exponencial (54.0799) + 0.0327×Exponencial (1.8289)

Figure 6 – Performance of M/G^c/1+G and M/G/c+G – probability of waiting.



Distribuições:
G_{24} ~ Weibull (0.9276; 26.004; 3.1333) G_{25} ~ Exponencial (55.9089) G_{26} ~ Weibull (1.3021; 2.3952; 0)
G_{27} ~ Fatigue Life (0.3726; 52.2792) G_{28} ~ Normal (30.9598; 445.3239) M_4 ~ Exponencial (0.0323)
G_{29m} ~ 0.8418×Fatigue Life (0.4055; 43.4875) + 0,1582×Fatigue Life (0.9434; 6.1219)
G_{30m} ~ 0.5221×Normal (29.1926; 232.5241) + 0,4779×Normal (26.7163; 212.7998)
G_{31m} ~ 0.9673×Exponencial (54.0799) + 0.0327×Exponencial (1.8289)

Figure 7 – Performance of M/G^c/1+G and M/G/c+G models – average waiting time.

bility of abandoning, although they present larger percentage differences, 13.09% (Figure 6) and 20.16% (Figure 4), respectively.

We note that the analytical queuing models that have mixed distribution for the patience times were effective to represent the average waiting time of the company's Call Center. In other measures of performance, probability of abandonment and probability of waiting and traffic intensity, mixed distributions for patience also occur in analytical queuing models, but these models were not the most adequate to represent them. According to the type of analytical queuing model adopted to represent the company's Call Center, regardless of any performance measure, the models that use mixed distributions for patience in some cases are better than their corresponding model with a single parametric distribution class to model the time of patience. Observing Table 7, for the analytical model M/G/c+G, the Normal and Fatigue Life mixed distributions for the patience times are better than their non-mixed counterparts (lower MAI). For model M/G^c/1+G, only the patience with mixed Normal distribution is better than its non-mixed correspondent. Another result obtained is that the Call Center of the company, regardless of any performance measure, can be represented by an analytical queuing model with a mixed distribution for the patience times (Figure 3).

Additional analyzes were performed by applying the M/M/c+M and M/M/c+G queuing models in this company data and calculating the same performance measures and the same MAI and Dif% indicators. The results obtained with these analytical queuing models were much worse than those obtained with the M/G^c/1+G and M/G/c+G models. Compared with the analytical model M/G$^c{}_{26}$/1+G$_{30m}$, which best represents the company's Call Center, this model presented an MAI 375 times smaller than the M/M$_4$/c+G$_{27}$ model and 2895 times smaller than the M/M$_4$/c+M$_5$. Therefore, the analytical queuing model M/G$^c{}_{26}$/1+G$_{30m}$ is more effective to represent the company's Call Center than the analytical queuing models M/M$_4$/c+G$_{27}$ and M/M$_4$/c+M$_5$, because the results obtained from its measurements have less deviations from the same measurements taken from the real data. Further details of such comparisons can be found in Appendix C. The mathematical formulations used here in the M/M$_4$/c+G$_{27}$ and M/M$_4$/c+M$_5$ models are the same described in Ferrari (2016).

Table 7 – Comparing mixed and non-mixed patience.

Models M/G/c+G	**MAI**	**Models M/G^c/1+G**	**MAI**
M/G$_{24}$/c + G$_{30m}$	0.0095	M/G$^c{}_{26}$/1 + G$_{30m}$	0.0051
M/G$_{24}$/c + G$_{28}$	0.0181	M/G$^c{}_{26}$/1 + G$_{28}$	0.0065
M/G$_{24}$/c + G$_{29m}$	2.1105	M/G$^c{}_{26}$/1 + G$_{27}$	2.0161
M/G$_{24}$/c + G$_{27}$	2.1153	M/G$^c{}_{26}$/1 + G$_{29m}$	2.2632
M/G$_{24}$/c + G$_{25}$	2.6659	M/G$^c{}_{26}$/1 + G$_{25}$	2.7903
M/G$^c{}_{24}$/c + G$_{31m}$	5.2445	M/G$^c{}_{26}$/1 + G$_{31m}$	5.4244

Arrival process:
Rate λ = 0.7318; mean E(τ) = 1.3665
Distributions:
M $\sim$ Exponencial(1.3665)

Service process:
Rate μ = 0.0323; mean E(S) = 30.9598; c = 14
Distributions:
G$_{24}$ $\sim$ Weibull (0.9276; 26.004; 3.1333)
G$_{26}$ $\sim$ Weibull (1.3021; 2.3952; 0)

Abandonment process:
Rate θ = 0.01789; mean E(R) = 55.9089
Distributions:
G$_{25}$ $\sim$ Exponencial (55.9089)
G$_{27}$ $\sim$ Fatigue Life (0.3726; 52.2792)
G$_{28}$ $\sim$ Normal (30.9598; 445.3239)
G$_{29m}$ $\sim$ 0.8418×Fatigue Life (0,4055; 43.4875) + 0.1582×Fatigue Life (0.9434; 6.1219)
G$_{30m}$ $\sim$ 0.5221×Normal (29.1926; 232.5241) + 0.4779×Normal (26.7163; 212.7998)
G$_{31m}$ $\sim$ 0.9673×Exponencial (54.0799) + 0.0327×Exponencial (1.8289)

5.1 Alternative Scenarios

Several scenarios can be elaborated with the analytical queuing models described in this study and that adequately represent the company's Call Center. Managers can use their results to make predictions about the operations of these systems in these alternative scenarios and take decisions. However, for the analysis of these scenarios, no real data is available from the company's Call Center to verify the accuracy obtained. This accuracy was then verified by comparing the results of the scenarios with the results obtained by discrete simulation models.

In particular, one of the concerns for Call Center administrators is the abandonment of their customers, which can be quantified by the probability of abandoning. The increase in the value of this probability suggests to the manager a low level of service offered by the system. When leaving, customers are indirectly informing managers that it is not worth waiting for the service offered. Maintaining an adequate level of service implies reducing the probability of abandonment, which depends on the number of attendants in operation. It is common practice in some Brazilian systems to maintain the probability of abandonment close to 2% (Bouzada, 2009). Thus, the Call Center manager can size the team by calculating the number of attendants needed to maintain the desired level of service. This scenario, called Scenario 1, was analyzed with the analytical queuing model and the results were verified with the corresponding simulation model.

The probability of abandonment is a subjective measure of the service level offered. The service level can be measured by the percentage of calls served up to an established amount of time between those waiting in queue until they are served by the operators. Thus, the service level is the probability of the user waiting in the queue, for a maximum amount of time to be served, calculated by:

$$NS = P(W \leq x) = \frac{\textit{number of calls answered until time } x}{\textit{number of calls answered by the operator}} \tag{11}$$

In Brazilian Call Center operations, a minimum service level of 75% of the calls that are to be answered within 10 seconds is usually established (BOUZADA, 2009). This scenario, called Scenario 2, was also analyzed with the analytical queuing model and the results verified with the corresponding simulation model. The implementation of the simulation model used for comparison in the analysis of these scenarios was performed in the Arena® software (Rockwell, 2014), using the Arena Contact Center Edition simulation system specially developed for Call Center managers.

In this study, we considered simulations with 300 replications and each replication with duration of one day, generating, at the end of all the rounds, an average of 33 thousand calls in the Call Center of the company. The statistics were collected every 60 minutes and reported in a report produced by the Arena software. The first 100 rounds of the simulation were considered as the warm up period, because after that period the system began to operate in statistical equilibrium, and from then on the statistics were counted. At the end of the replications, the Arena software generated a report containing various information from each of the rounds, as well as a summary report (Siman Summary Report) containing the results of all replications. Among these results are the performance measures, such as the average waiting time, the probability of abandonment, the traffic intensity, and the probability of waiting, which is calculated by dividing the number of customers who were served (excluding those who left) and the number of customers created in the simulation. In addition to the outcome of these measures, the half-interval of the confidence interval was also reported. All simulations were performed on an Intel Core™ I7 processor, 2.4 GHz and 8.00 GB RAM, and took an average of only a few minutes to estimate the performance measures for each queuing problem (Ferrari, 2016).

Before using the simulation model to analyze the scenarios, it was validated by comparing the values obtained by it, for each of the performance measures, with the corresponding values extracted from the actual data from the interval of 12 to 4 pm, collected from the company's Call Center. The purpose of this validation was to verify if the simulation model actually represented the reality of the company's Call Center. In this analysis, we used confidence intervals with $\gamma =$ 95% probability, calculated for each of the performance measures. All confidence intervals of the simulation model contained the results of the performance measures obtained by the corresponding value extracted from the sample of the actual data from the 12-4 pm interval (these results are detailed in Ferrari, 2016). Thus, the simulation model was considered validated and was also used for the scenario analysis, to obtain results with values that extrapolate the sample data. Each of these scenarios is discussed in detail in the following.

It was considered the case of the Call Center of the company that can be represented by the mixed distribution queuing model for the patience, $M/G^c{}_{26}/1+G_{30m}$, as shown in Figure 3. Although a single server is considered, the service rate of this model is multiplied by a quantity c of servers, to be next to the multiple-server system $M/G^c{}_{26}/1+G_{30m}$, with the same distribution for the service times and patience times as the previous one. Thus, the $M/G^c{}_{26}/1+G_{30m}$ analytical queuing model operates approximately as the $M/G^c{}_{26}/1+G_{30m}$ analytical queuing model. Using model $M/G^c{}_{26}/1+G_{30m}$, with Weibull distribution for the service times and mixed Normal distribution for the patience times, the number of attendants required for the probability of abandonment of the Call Center of the company decreased from 31.8% (actual data) to close to 2%. The company's Call Cente*r* operates normally with a team of 14 attendants. To reduce the probability of abandonment, the number of attendants was gradually increased, replacing this quantity in the equations of the analytical model $M/G^c{}_{26}/1+G_{30m}$ and also in the simulation model. The results obtained are shown in Table 8.

First, it is observed that the values of the abandonment probabilities are very close in both analytical and simulated models, evidencing that the $M/G^c{}_{26}/1+G_{30m}$ queuing model is effective to represent the company's Call Center and that it can be used to support the design and operations decisions of this system. For example, if the Call Center manager increases the staff to 24 attendants, the probability of abandonment is expected to be close to 5%, significantly improving the service level. If it rises to 29 attendants, this probability reduces for about 2%. Under the conditions of this scenario with 29 attendants and simulating the $M/G^c{}_{26}/1+G_{30m}$ queuing model

Table 8 – Probability of abandonment versus number of attendants.
Call Center company – 12-4 pm.

Attendants	Queuing model $M/G^{c}{}_{26}/1+G_{30m}$	
	Analytical	Simulated
$c = 16$	31.2%	28.9%
$c = 18$	22.7%	21.7%
$c = 20$	14.1%	14.4%
$c = 22$	5.5%	9.4%
$c = 24$	5.4%	5.3%
$c = 26$	2.7%	3.6%
$c = 28$	2.1%	2.4%
$c = 29$	2.0%	1.3%

Distributions:
$G_{26} \sim$ Weibull (1.3021; 2.3952; 0)
$G_{30m} \sim$ 0.5221 × Normal (29.1926; 232.5241) + 0.4779 × Normal (26.7163; 212.7998)

that represents the company's Call Center, 32,842 users were served by the operators, 10,555 of which were answered in up to 10 seconds. Thus, the service level for this scenario is:

$$NS = \frac{10555}{32842} = 0,3214\,,\ or\ 32,14\%$$

This value for the service level is less than the 75% condition set. Thus, according to this scenario, the company's Call Center with a staff of 29 attendants would be sufficient to guarantee the goal of 2% for the probability of abandonment, but insufficient to reach the service level established in 75% of the calls answered in up to 10 seconds. Analysis of scenarios of this type are useful to show the potential of the use of analytical queuing models with abandon in the sizing of the number of attendants, to adapt the system to the service level desired.

6 CONCLUSIONS

This article analyzed the effectiveness of abandonment queuing models to evaluate performance measures in a Brazilian Call Center. The study investigated the use of generic distributions of probabilities that best fit the Call Center customers' service times, as well as generic (particularly mixed) distributions that best fit their patience times. The major interest was to consider the mixed distributions in the analytical queuing models of abandonment $M/G^{c}/1+G$ and $M/G/c+G$ to be applied in the congestion problems of Call Centers, obtaining performance measures that best express and represent reality. The abandonment phenomenon has been studied in the literature through queuing models that consider a single class of parametric distribution to model customers' patience, often represented by Exponential, Weibull, and Erlang. However, in Call Centers, the population of users in general has different characteristics with respect to the time of patience, being some more patient than others. Thus, it seems more reasonable to consider mixed probability distributions to model the time of patience because they are more appropriate to absorb these characteristics.

The case study conducted in the company showed the effectiveness of these queuing models to estimate performance measures important for the design, planning and operation of Call Center systems. Comparing the obtained measurements with those extracted from the real data, it was verified that some analytical queuing models considered in this study produced results very close to those extracted from the actual data, with deviations of less than 1%, thus capturing well the reality of the company Call Center. However, there are models that produced better results in some performance measures than in others. In some cases, depending on the measure of performance considered, the mixed distribution models for patience presented better results (smaller deviations from actual data) than their counterparts with non-mixed distributions. As for the approximations available for the analytical queuing models $M/G^c/1+G$ and $M/G/c+G$, it is concluded that their values are close to the values observed in practice (obtained by the sample data) and also to the simulated values and within the confidence intervals (95 %) (obtained by the discrete simulation model). We are not aware of other experimental studies in the literature exploring this line of research.

Some limitations of this study occurred in the conception of the mixed probability distributions to model the abandonment phenomenon in the $M/G^c/1+G$ and M/G/c+G models. To compose these mixed distributions, for simplicity, only two components of the same parametric family were used, i.e., mixtures of two Exponential (Ex+Ex), two Normal (N+N) and two Fatigue Life (FL+FL) components. An interesting topic for future research is to use a more elaborate method to determine the number of components of these mixed distributions. It is also observed, as a limitation of this study, that some factors that affect customers' impatience, such as gender, age, type of service, levels of patience, among others, were not incorporated into the composition of these mixed probability distributions, which characterize the subpopulations of Call Center users. This motivates another interesting future research. Further promising research would be to explore the use of analytical queuing models that involve mixed distributions in the abandonment process in other Call Center architectures; the use of other methods to estimate the mixed distributions in the modeling of users' patience times; the development of optimization procedures based on the proposed analytical queuing models to support better system configuration decisions.

Acknowledgements

We would like to thank the three anonymous reviewers for their useful comments and suggestions of revision, and the company for kindly providing the data for the case study. We also thank Capes, CNPq and FAPESP for their partial financial support for this research.

References

AALEN O & GJESSIN H. 2001. Understanding the shape of the hazard rate: A process point of view. *Statistical Science*, **16**: 1–22.

BACELLI F & HEBUTERNE G. 1981. On queues with impatient customers. In: *Proceedings of the International Symposium on Computer Performance*, 159–179.

BOUZADA MAC. 2009. Análise de cenários em um call center usando simulação. In: *Proceedings of SIMPOI*, 1-16, Brazil.

BRILL PH. 1979. An embedded level crossing technique for dams and queues. *Journal of Applied Probability*, **16**(1): 174-186.

BRILL, PRECY H. 2008. *Level Crossing Methods in Stochastic Models*. New York: Springer.

BROWN L, GANS N, MANDELBAUM A, SAKOV A, SHEN H, ZELTYN S & ZHAO L. 2002. *Statistical analysis of a telephone call center: A queuing-science perspective*. Technical Report. Technion, Israel Institute of Technology, 1-61. http://iew3.technion.ac.il/serveng/References/references.html.

COHEN JM. 1977. On up and down crossings. *Journal of Applied Probability*, **14**(2): 405-410.

COHEN JM & RUBINOVITCH M. 1977. On level crossing and cycles in dam processes. *Mathematics of Operations Research*, **2**(4): 297-310.

COX DR & OAKES D. 1990. *Analysis of Survival Data*. London: Chapman and Hall.

FERRARI SC. 2016. *Abordagens de modelos de fila com abandono para análise de congestão em Call Centers*. PhD Dissertation. Universidade Federal de São Carlos, São Carlos-SP, Brazil.

FERRARI SC & MORABITO R. 2020. Application of queueing models with abandonment for Call Center congestion analysis. *Gestão & Produção*, **27**(1): e3765. doi.org/10.1590/0104-530X3765-20.

FORBES C, EVANS M, HASTINGS N & PEACOCK B. 2011. *Statistical Distributions*. New Jersey: John Wiley and Sons.

IRAVANI F & BALCIOGLU B. 2008. Approximations for the M/G/N+G type call center. *Queuing Systems*, **58**: 137-153.

KALBFLEISCH JD & PRENTICE RL. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

LIMA H. 2012. *Call Center em números: métricas e como medir o desempenho*. Available in: http://blog.teclan.com.br/call-center-em-numeros-metricas-e-como-medir-o-desempenho/ (06.07.12). Accessed in: 18.12.12.

MANDELBAUM A & ZELTYN S. 2012. Data-stories about (im)patient customers in tele-queues. *Queuing Systems*, **75**(2): 115-146.

MANDELBAUM A & ZELTYN S. 2005. *The Palm/Erlang-A queue with applications to call centers*. Technical Report. Technion, Israel Institute of Technology. http://iew3.technion.ac.il/serveng/References/references.html.

MANDELBAUM A, SAKOV, A & ZELTYN S. 2001. *Empirical analysis of a call center*. Technical Report. Technion, Israel Institute of Technology. http://iew3.technion.ac.il/serveng/References/references.html.

OLIVEIRA CC. 2009. *Espera e abandono na fila M/M/n+G e variantes*. Dissertation. Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, Brazil.

PALM C. 1953. Methods of judging the annoyance caused by congestion. *Tele*, **4**: 189–208.

STANFORD RE. Reneging phenomenon in single channel queues (1979). *Mathematics of Operations Research*, **4**(2): 162-178.

ZOHAR E, MANDELBAUM A & SHIMKIN N. 2002. Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science*, **48**(4): 556–583.

WHITT W. 1993. Approximations for the GI/G/m queue. *Production and Operations Management*, **2**: 114–161.

How to cite

FERRARI SC & MORABITO R. 2023. Performance analysis of a Brazilian Call Center with impatient customers using M/G^c/1+G and M/G/*c*+G queuing models. *Pesquisa Operacional*, **43**: e271290. doi: 10.1590/0101-7438.2023.043.00271290.

APPENDIX A QUEUING MODEL EQUATIONS OF THE M/G/1+G

Let the random variables W_S describing the waiting time of the users until they are attended by an agent, W_R the waiting time of the users until leaving the system (time of patience), and W_T the waiting time for all users (those who abandoned and those who were served). Using theorems 4, 5 and 6 of Stanford (1979) and defining the random variable V as the virtual hold time with $F(x)$ and $f(x)$, i.e., its respective distribution function and density function (assuming that they exist for $x>0$), respectively, Iravani and Balcioglu (2008) defined for the G/G/1+G queue:

Probability of the system being empty:

$$P_0 = F(0) \tag{A.1}$$

Probability of abandonment:

$$P_A = 1 - \frac{1-P_0}{\rho} \tag{A.2}$$

Distribution function of the waiting time of the customers served:

$$W_S(x) = \int_0^x \frac{dF(t)\bar{R}dt}{1-P_A} \tag{A.3}$$

Probability of the waiting time of the customers served being equal to zero:

$$P(W_S = 0) = \frac{P_0}{1-P_A} \tag{A.4}$$

Distribution function of the waiting time to abandonment:

$$W_R(x) = \int_0^x \frac{\bar{F}(t)R(t)dt}{P_A} \tag{A.5}$$

Distribution function of the waiting time of all customers:

$$W_T(x)=(1-P_A)W_S(x)+P_A W_R(x) \tag{A.6}$$

The virtual waiting time density function of the M/G/1+G system was established by Iravani and Balcioglu (2008) using the level crossing method, originally developed by Brill (1979), Cohen (1977), Cohen and Rubinovitch (1977) and Brill (2008). This method ensures that, for a certain level $x > 0$ and the system being in equilibrium, the rate of passage above the level is equal to the rate of passage below the level, that is, the downcrossing rate of x is equal to the upcrossing rate of x. Using this, Iravani and Balcioglu (2008) deduced the integral equation for the virtual waiting time:

$$f(x) = \lambda P_0 \overline{B}(x) + \lambda \int_0^x \overline{B}(x-t)\overline{R}(t)f(t)\,dt \tag{A.7}$$

with normalization equation equal to:

$$P_0 + \int_0^{\infty} f(t)\,dt = 1 \tag{A.8}$$

Equation (A.7) is an integral second-order Volterra equation, which requires a numerical solution in the determination of approximate results of these equations. Iravani and Balcioglu (2008) used the trapeze rule to obtain the approximations for $f(x)$. This successive application of the trapeze rule (recursion) ends to a *MAX* value, such that for $n \geq MAX$ it is assumed that $f(n\Delta) = 0$. Applying the trapeze rule in (A.8), Iravani and Balcioglu (2008) obtained an equation to determine the value of the probability of the system being empty:

$$P_0 \left\{ 1 + \Delta \left[\frac{f_1(0) + f_1(MAX\Delta)}{2} + \sum_{n=1}^{MAX-1} f_1(n\Delta) \right] \right\} = 1 \tag{A.9}$$

The probability of abandonment P_A can be calculated using equations (A.2) and (A.9). The distribution function of the average waiting time of the users that are served can be determined by employing the trapeze rule in equation (A.3), considering the integer number n_X satisfying the equality $n_X\Delta = x$:

$$P(0 < W_s \leq x) = \frac{\Delta P_0}{1 - P_A} \left\{ \frac{f_1(\Delta)\overline{R}(\Delta) + f_1(n_x\Delta)\overline{R}(n_x\Delta)}{2} + \sum_{n=2}^{n_x-1} f_1(n\Delta)\overline{R}(n\Delta) \right\} \tag{A.10}$$

Another performance measure presented in Iravani and Balcioglu (2008) is the *k-th* moment of the waiting time of the customers served. Therefore, they first derived equation (A.7) and then applied the rule of trapeze to obtain:

$$E\left[W_S^k\right] = \frac{\Delta P_0}{1 - P_A} \left\{ \frac{\Delta^k f_1(\Delta)\overline{R}(\Delta) + (MAX\Delta)^k f_1(MAX\Delta)\overline{R}(MAX\Delta)}{2} + \sum_{n=2}^{MAX-1} (n\Delta)^k f_1(n\Delta)\overline{R}(n\Delta) \right\} \tag{A.11}$$

More details of the mathematical developments to obtain equations (A.1)-(A.11) can be found in Iravani and Balcioglu (2008), Ferrari (2016) and the references therein.

APPENDIX B QUEUING MODEL EQUATIONS OF THE M/G/*C*+G

The M/G/*c*+G model does not allow a treatable analytical solution. So, this study used the same formulations of Iravani and Balcioglu (2008) by using the M/G/1+G system to obtain approximations for the M/G/*c*+G system. Consider the system with a single server with the same Poisson arrival process and the same distribution of the abandon time as the multiple-server system. Consider also that the same service time distribution of the single-server system is preserved in the multiple-server system, but its rate is increased by c times (the number of servers), so it remains $c\mu_B$. Denoting the system by M/G^c/1+G, the traffic intensity for both systems is equal to:

$$\rho_{M/G/c+G} = \rho_{M/G^c/1+G} = \frac{\lambda}{c\mu_B} \tag{B.1}$$

Iravani and Balcioglu (2008) argued that the possible parallelism between the analytically treatable models M/M^c/1+G and M/M/c+G would be indicative of a possible parallelism between the queuing models M/G^c/1+G and M/G/c+G. Consequently, we have the equality of the traffic intensities of the last two systems, as described in equation (B.1). One reason for understanding this parallelism is that when all servers are busy on both systems, customers waiting in the queue will be served with the same service rate $c\mu$ by leaving the equivalent systems. In this way, considering that these users wait to be attended or abandon the system, Iravani and Balcioglu (2008) proposed the following approximation for the probability of waiting:

$$P_{M/G/c+G}(W > x \mid W > 0) \cong P_{M/G^c/1+G}(W > x \mid W > 0), \quad x > 0 \tag{B.2}$$

By writing these probabilities separately for customers who are served (before to abandon) and for all customers, we have:

$$P_{M/G/c+G}(W_S > x \mid W_S > 0) \cong P_{M/G^c/1+G}(W_S > x \mid W_S > 0) \tag{B.3}$$

$$P_{M/G/c+G}(W_T > x \mid W_T > 0) \cong P_{M/G^c/1+G}(W_T > x \mid W_T > 0) \tag{B.4}$$

Using the definition of conditional probability in these equalities, it follows that:

$$\frac{P_{M/G/c+G}(0 < W_S \leq x)}{P_{M/G/c+G}(W_S > 0)} \cong \frac{P_{M/G^c/1+G}(0 < W_S \leq x)}{P_{M/G^c/1+G}(W_S > 0)} \tag{B.5}$$

and still:

$$\frac{P_{M/G/c+G}(0 < W_T \leq x)}{P_{M/G/c+G}(W_T > 0)} \cong \frac{P_{M/G^c/1+G}(0 < W_T \leq x)}{P_{M/G^c/1+G}(W_T > 0)} \tag{B.6}$$

After some algebraic manipulations, Iravani and Balcioglu (2008) proposed the first approximation, called primary scaling approximation (PSA), for the M/G/c+G queuing model, presented by equations (B.7)-(B.10):

$$P_{M/G/c+G}(W_S \leq x) \cong \frac{P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M/c+G}(A)} + \left[1 - \frac{P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M/c+G}(A)}\right] \frac{P_{M/G^c/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \tag{B.7}$$

$$P_{M/G/c+G}(W_T \leq x) \cong P_{M/M/c+G}(W_T = 0) + \left[1 - P_{M/M/c+G}(W_T = 0)\right] \frac{P_{M/G^c/1+G}(0 < W_T \leq x)}{1 - P_0} \tag{B.8}$$

$$E_{M/G/c+G}\left[W_S^K\right] \cong \left[1 - \frac{P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M/c+G}(A)}\right] \frac{E_{M/G^c/1+G}\left[W_S^K\right]}{1 - \frac{P_0}{1-P_A}} \tag{B.9}$$

$$E_{M/G/c+G}\left[W_T^K\right] \cong \left[1 - P_{M/M/c+G}(W_T = 0)\right] \frac{E_{M/G^c/1+G}\left[W_T^K\right]}{1 - P_0} \tag{B.10}$$

More details of the mathematical developments to obtain equations (B.1)-(B.10) can be obtained in Iravani and Balcioglu (2008), Ferrari (2016) and the references therein.

APPENDIX C COMPARISON BETWEEN THE ANALYTICAL QUEUING MODELS M/G/*C*+G, M/M/*C*+M AND M/M/*C*+G

Models	**MAI**
M/G$^c{}_{26}$/1 + G_{30m}	0.0051
M/G^{c}26/1 + G_{28}	0.0065
M/G_{24}/c + G_{30m}	0.0095
M/G_{24}/c + G_{28}	0.0181
M/M_4/c + G_{27}	1.9123
M/G$^c{}_{26}$/1 + G_{27}	2.0161
M/G_{24}/c + G_{29m}	2.1105
M/G_{24}/c + G_{27}	2.1153
M/G$^c{}_{26}$/1 + G_{29m}	2.2632
M/G_{24}/c + G_{25}	2.6659
M/G$^c{}_{26}$/1 + G_{25}	2.7903
M/G$^c{}_{24}$/c + G_{31m}	5.2445
M/G$^c{}_{26}$/1 + G_{31m}	5.4244
M/M_4/c+M_5	14.7646
M/M_4/c + G_{29m}	45.7600
M/M_4/c + G_{25}	91.3964
M/M_4/c + G_{31m}	101.9488
M/M_4/c + G_{28}	137.0806
M/M_4/c + G_{30m}	151.9219

Remark: $M_5 \sim$ Exp (55.9089).

Comparison ratio of the models $M/G^{c}{}_{26}/1{+}G_{30m}$ and $M/M_4/c{+}G_{27}$:

$$R = \frac{MAI\ de M/M_4/c+G_{27}}{MAI\ de M/G^c_{26}/1+G_{30m}} = \frac{1.9123}{0.0051} = 374.96$$

Remark: the MAI of model $M/G^{c}{}_{26}/1{+}G_{30m}$ is 375 times less than model $M/M_4/c{+}G_{27}$.

Comparison ratio of the models $M/G^{c}{}_{26}/1 + G_{30m}$ and $M/M_4/c{+}M_5$:

$$R = \frac{MAI\ de M/M_4/c+M_5}{MAI\ de M/G^c_{26}/1+G_{30m}} = \frac{14.7646}{0.0051} = 2895.02$$

Remark: the MAI of model $M/G^{c}{}_{26}/1{+}G_{30m}$ é 2895 times less than model $M/M_4/c{+}M_5$.