**Pak. J. Statist.**
**2026 Vol. 42(3), 185-203**

# IMPROVING CALL CENTER OPERATIONS THROUGH SIMULATION: INSIGHTS FROM IRAQ'S TELECOM INDUSTRY

**Raed Hameed Mahdi**[1] and **Noor Saifurina Nana Khurizan**[2§]
[1] Iraqi Ministry of Communications,
Baghdad-Al-Nisoor Square, Iraq
[2] School of Mathematical Science,
University Sains Malaysia, Penang, Malaysia
[§] Corresponding author Email: saifurina@usm.my

## ABSTRACT

This study analyses operational inefficiencies among the telecommunications call centers of the Republic of Iraq through discrete event simulation. Information in 14 Governorates on 220 working days was then analyzed to determine queue accumulation and capacities constraints. The distributions of arrival and service time were statistically fitted and incorporated in an ARENA simulation model. Winsorization was employed to remove outliers in SPSS to make the model more stable and maintain the variation in the data. Simulation findings indicate that single server systems result into congestion at all times. The waiting time in high demand Governorates was over 50 minutes on average. The capacity was topped to two and three servers and waiting times and queue lengths were cut by nearly 60 percent. Most Governorates had error rates which were lower than 7 percent as was produced staffing as opposed to equal allocation. The model allows managers to determine minimum requirements in Governorate in terms of servers. The concept of multi-server deployment is proposed as a suitable solution for high-load areas. Staffing schedules and adjustments on the hour of service can also be safely tested and then implemented using the model. These lessons will help make evidence-based policy decisions and effective utilization of the scarce resources in the public service call centers.

## KEYWORDS

Queuing theory; simulation call center; Call center service; discrete event simulation; waiting time

## 1. INTRODUCTION

Call centers constitute a core operational unit in contemporary service organizations. They function as the primary interaction point between firms and customers, directly shaping service quality, customer satisfaction, and operational efficiency. With the advancement of information and communication technologies, call center operations have evolved toward more integrated and responsive systems, improving coordination and service delivery outcomes (Ferrari and Morabito, 2020; Bolandifar et al., 2023). Modern call centers are no longer limited to voice-based communication. They operate as multi-channel service platforms integrating telephone calls, emails, online chats, and customer relationship management systems. CRM technologies enable agents to access customer

© 2026 Pakistan Journal of Statistics 

histories and real-time information, supporting faster resolution and service consistency (Batarseh and Wang, 2018). This integration has increased system complexity and amplified the importance of effective operational management. In telecommunications, call center performance is particularly critical due to high demand volatility and time-sensitive service expectations. In Iraq, rapid growth in subscriber numbers has exceeded the pace of capacity expansion. As a result, call centers experience persistent congestion, long queues, and excessive waiting times. Empirical evidence suggests that prolonged delays undermine customer trust and generate negative service perceptions, which may ultimately lead to customer churn and revenue loss. Managing such systems requires accurate estimation of customer waiting behavior. Key performance indicators such as waiting time, queue length, and service congestion are influenced by multiple interacting factors, including agent availability, arrival rate variability, call complexity, and system configuration (Batarseh and Wang, 2018). Behavioral studies further indicate that perceived unfairness and uncertainty during waiting significantly deteriorate customer attitudes toward service providers (Johnson and Ganiyu, 2019). These findings highlight the need for analytical tools capable of capturing both operational and behavioral dimensions of call center performance. Traditional analytical queueing models have been widely applied to study call center systems. However, these models often rely on simplifying assumptions such as exponential arrivals, stationary demand, and homogeneous service processes. Such assumptions limit their applicability in large-scale and highly stochastic environments. Discrete-event simulation addresses these limitations by modeling system dynamics under realistic arrival and service time distributions. Simulation-based studies demonstrate that multi-server queueing systems with general distributions provide more accurate performance estimates and stronger support for optimization decisions (Pankratova et al., 2022). Among simulation tools, ARENA has been extensively used to analyze customer–server systems. Its modular structure allows detailed representation of complex service processes, scenario testing, and sensitivity analysis. Recent statistical studies have relied on quality of fit tests to verify the suitability of probability distributions to real-world data. A prime example is the modified chi-squared test proposed by Bagdonavicius and Nikulin, which was later adapted for application to extended Weibull distributions and controlled data (Aidi et al., 2021). This test has demonstrated high efficiency in assessing the suitability of probability models through simulation studies and real-world applications (Khaoula Aidi1, 2021). Recent statistical studies have relied on the development of goodness-of-fit tests to verify the suitability of probability distributions, particularly in the context of dependency models and waiting systems. A notable example is the work of Abu-Youssef and Gerges (2022), who proposed a novel non-parametric test based on the goodness-of-fit methodology to test the exponential hypothesis against a broader range of life distributions, demonstrating high efficiency using Monte Carlo simulations and real-world applications. These findings support the use of goodness-of-fit tests to validate statistical distributions before their inclusion in simulation models.

Previous research confirms its effectiveness in evaluating system performance and supporting resource allocation decisions across various service domains, including healthcare, logistics, and customer service operations (Guseva et al., 2018). Despite the breadth of international literature on call center simulation, empirical studies focusing on developing economies remain limited. Most existing research examines call centers in

developed countries characterized by stable infrastructure, mature management practices, and advanced technologies. Regional demand heterogeneity, staffing constraints, and operational limitations in contexts such as Iraq are rarely addressed. Despite the extensive international literature on call center simulation, empirical studies in developing economies remain limited. Most existing research focuses on call centers in developed countries with stable infrastructure, mature management practices, and advanced technologies. Consequently, these findings are not directly transferable to contexts characterized by demand volatility, staffing constraints, and limited operational flexibility, such as Iraq. In particular, there is a lack of empirical evidence quantifying the extent of queuing congestion across regions and evaluating how alternative service capacities affect waiting times and system stability. Furthermore, data quality challenges, including the presence of extreme observations, are often overlooked, leading to biased parameter estimation and unreliable simulation outputs (Fitrianto et al., 2022). This gap creates uncertainty for managers and policymakers when designing staffing and capacity decisions under resource constraints. Furthermore, data quality challenges, including the presence of extreme observations, are often overlooked, leading to biased parameter estimation and unreliable simulation outputs (Fitrianto et al., 2022). This gap creates uncertainty for managers and policymakers when designing staffing and capacity decisions under resource constraints.

For clarity and to ensure a consistent understanding of the terminology used throughout this paper, Table 1 summarizes the key terms and their definition.

**Table 1**
**Definitions of Key Terms Used**

| Term | Definition |
|---|---|
| Call Center | A service unit that handles customer requests and inquiries through telephone communication. |
| Queueing System | A system that describes customer arrivals, waiting in a queue, and receiving service from one or more servers. |
| Discrete Event Simulation (DES) | A simulation approach in which system behavior is represented as a sequence of discrete events such as arrivals and service completions. |
| G/G/1 Queue | A single-server queueing model where both arrival times and service times follow general (data-driven) statistical distributions. |
| G/G/c Queue | A multi-server queueing model with *c* servers and general distributions for arrival and service times. |
| Waiting Time | The time a customer spends waiting in the queue before service begins. |
| Queue Length | The number of customers waiting in the system at a given time. |
| Utilization Rate | The proportion of time a server is busy providing service. |
| Winsorization | A statistical technique used to limit the effect of extreme values by replacing them with boundary values. |
| Model Validation | The process of comparing simulation results with real data to assess the accuracy of the mode |

**Scope and Limitations**

This paper aims at assessing the operational performance of the telecommunications call centers in Iraq on the basis of discrete-event simulation using empirical data on 220 working days (14 governorates in 2018-2022). The analysis is confined to call center operation during the working hours (9:00 a.m. - 2:00 p.m.) only and the simulation is based on First-In, First-Out (FIFO) service discipline without priority services. Besides, there is no explicit modeling of customer abandonment behaviour and agent-level variance. Consequently, the findings are to be viewed in the context of the existing operational data and assumptions followed in the simulation framework.

## 2. METHODOLOGY

To analyze the dynamics of call center queuing, the study is based on a structured analytical framework that involves the analysis of empirical data and discrete-event simulation (DES). The methodological framework involves data preprocessing, statistical and simulation modeling in a sequential and logical way, as shown in Figure 1. The methodological steps will consist of data collection (220 days, 2018-2022), outlier detection by means of boxplots, statistical distribution, and modeling of single- and multi-channel services by means of ARENA. The rigorous verification and validation procedures such as error rate analysis and goodness-of-fit tests are used to achieve the reliability of the results. The methodology strikes a balance between theoretical queuing models (e.g., M/M/c) and empirical validation in coming up with practical insight into how to maximize service efficacy.

Figure 1 presents the research steps:

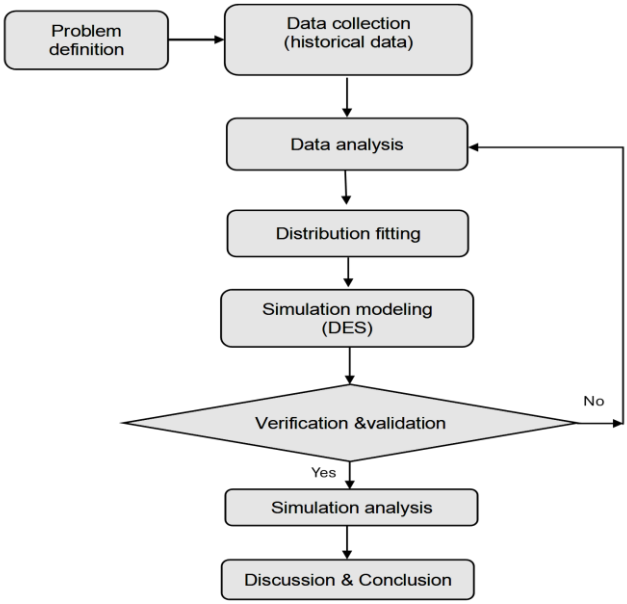


**Figure 1: Research Steps**

### 2.1 Data Collection

The analysis focused on customer service centers' operations over 220 days spanning 2018–2022, ensuring temporal consistency for comparative evaluations. Data included customer service volumes and service timestamps from a telecommunications customer service center in Iraq, operating Sunday to Thursday (9:00 a.m. - 2:00 p.m.). This timeframe minimizes seasonal variability, ensuring uniform conditions for assessing queuing dynamics.

### 2.2 Data Preprocessing

Outliers were identified using SPSS boxplots based on the interquartile range (IQR). Instead of mean substitution, which reduces data variability, a Winsorizing technique was applied. Values below the 5th percentile and above the 95th percentile were replaced with the nearest boundary values. Service and arrival times were Winsorized using SPSS to limit the influence of extreme observations while preserving the overall data structure. The Winsorized variables were then used as inputs to the simulation model to improve robustness (Fitrianto et al. 2022).

### 2.3 Statistical Distribution Fitting

Key metrics arrival intervals, service durations, and wait times were analyzed to identify underlying patterns.

The process involved (Oladimeji and Ibidoja, 2020):

i) **Data Aggregation**: Historical records of call arrivals and service times were compiled.
ii) **Exploratory Analysis**: Histograms and boxplots assessed data structure and candidate distributions (e.g., exponential, Poisson).
iii) **Goodness-of-Fit Testing**: Kolmogorov-Smirnov (K-S) and chi-square tests evaluated alignment between empirical data and theoretical distributions (Delgado and Vainora, 2022). Recent studies indicate that extended or blended distributions may provide a more accurate representation of time data compared to the traditional exponential distribution, especially when there is high variance or nonlinear behavior in risk rates (Gerges, 2022).
iv) **Distribution Selection**: The best-fitting distribution was selected based on statistical significance and relevance to queuing theory.

### 2.4 Discrete-Event Simulation (DES) Framework

A DES model was developed using ARENA software to simulate call center operations under stochastic demand. Events such as arrivals, queueing, service completion, and departures were modeled. Scenarios tested varying server counts (1-3) to quantify impacts on waiting times and queue lengths. The DES framework enabled granular analysis of agent schedules, resource utilization, and service-level compliance. The choice of ARENA in the current study is due to the fact that it is a popular discrete-event simulation (DES) application that could be used to describe the customer-server interactions in a complex way by capturing arrivals, queue formation, service completion, and departures. In contrast to the classic deterministic models of queueing, based on simplifying assumptions (e.g. Poisson arrivals and exponential wait times), ARENA can take advantage of empirically determined arrival and service-time distributions, thereby enhancing realism in the modeling of a highly variable call center demand. Furthermore, ARENA is useful in

conducting scenario testing by changing the quantity of servers (1-3) and the performance parameters of waiting time, queue length and utilization that can offer evidence-based information to the decision-maker regarding staffing and capacity planning.

**2.5 Queuing System Modeling**

The system was initially modeled as a single-server queue under the FIFO discipline and later extended to multi-server configurations to evaluate the impact of additional resources. While classical queueing models often assume Poisson arrivals and exponential service times, empirical analysis showed that both arrival and service processes follow general distributions in this system (Piao, Zheng, and Wang, 2023). Accordingly, Kendall's notation was used to classify the queueing structure, and the model was represented as G/G/1 and G/G/c for single- and multi-server cases, respectively. This formulation better captures real-world variability and provides a more realistic basis for simulation and performance evaluation. Table 2 shows the basic mathematical formulas used to analyze a single-server queuing system when access and service times have general distributions. These formulas help estimate performance indicators such as wait time, queue length, and utilization rate.

**Table 2**
**Analytical Performance Measures and Approximations for the G/G/1 Queueing Model (Amouzegar and Moshirvaziri 2006)**

| Parameter | Equation | Description |
|---|---|---|
| Traffic Intensity (ρ) | $\rho = \frac{\lambda}{\mu} \quad \rho = \frac{\lambda}{\mu}$ | Ratio of arrival rate ($\lambda$) to service rate ($\mu$). Must be < 1 for stability. |
| Squared Coefficient of Variation (Arrivals) | $C_a^2 = \frac{\sigma_a^2}{(1/\lambda)^2}$ | Measures variability in arrival times ( $\sigma_a^2$ = variance). |
| Squared Coefficient of Variation (Service) | $C_s^2 = \frac{\sigma_s^2}{(1/\mu)^2}$ | Measures variability in service times ( $\sigma_s^2$ = variance). |
| Average Waiting Time in Queue ($Wq$) | $W_q \approx \frac{\rho(C_a^2 + C_s^2)}{2\mu(1-\rho)}$ <br> $C_a^2 = \frac{\sigma_a^2}{(1/\lambda)^2}, C_s^2 = \frac{\sigma_s^2}{(1/\mu)^2}$ <br> $Wq \approx \left(\frac{C_a^2 + C_s^2}{2}\right)\left(\frac{\rho_0 (c\rho)^c}{c!c\mu(1-\rho)^2}\right)$ <br> $\frac{\lambda}{c\mu}$ | Kingman's approximation for general distributions. |
| Average Queue Length ($Lq$) | $L_q \approx \frac{\rho(C_a^2 + C_s^2)}{2(1-\rho)}$ | Derived from Little's Law ($Lq=\lambda Wq$). |
| Average System Time ($W$) | $W = W_q + \frac{1}{\mu}$ | Total time in system (queue + service). |
| Probability of Idle Server ($P0$) | $\rho \approx 1-\rho$ | Probability the server is idle. |

Table 3 presents the equations for analyzing a multi-server queuing system, showing how to calculate wait time, queue length, and server utilization rate when there are more than one service providers in the system.

Performance metrics and approximation formulas for the G/G/c multi-server queueing model (Amouzegar and Moshirvaziri, 2006)

**Table 3**
**Equations for Analyzing a Multi-Server Queuing System**

| Parameter | Equation | Description |
|---|---|---|
| Traffic Intensity (ρ) | $\rho = \frac{\lambda}{c\mu}$ | Ratio of arrival rate to total service capacity (c servers). Must be < 1. |
| Squared Coefficients ($C^2$) | $C_a^2 = \frac{\sigma_a^2}{(1/\lambda)^2}$ , $C_s^2 = \frac{\sigma_s^2}{(1/\mu)^2}$ | Measures variability in arrivals/service times. |
| Probability of Zero Customers ($P_0$) | $\left[\sum_{n-0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)}\right]^{-1}$ | Complex term requiring iterative calculation. |
| Avg. Waiting Time ($W_p$) | $Wq \approx \left(\frac{C_a^2 + C_s^2}{2}\right)\left(\frac{\rho_0 (c\rho)^c}{c!c\mu(1-\rho)^2}\right)$ | Allen-Cunneen approximation for multi-server queues. |
| Avg. Queue Length ($L_p$) | $Lq = \lambda W_q$ | Derived from Little's Law. |
| Avg. System Time ($W$) | $W = W_q + \frac{1}{\mu}$ | Total time in system (queue + service). |
| Server Utilization | $\frac{\lambda}{c\mu}$ | Fraction of time servers are busy. |

- **Arrival rate ($\lambda$):** Calculated from observed interatrial times.
- **Service rate ($\mu$):** Determined as the inverse of the average service duration.
- **Utilization ($\rho$):** Defined as $\lambda/(c \times \mu)$, where $c$ is the number of servers.

Defined as:

$$p = \frac{\lambda}{c\mu} \tag{1}$$

where $cu$ is the number of servers. The ARENA model simulated six-hour operational windows, iteratively adjusting server counts to optimize service levels (e.g., target wait times).

The system works in different scenarios. Figure 2 shows the work inside the Arena simulation program, where the scenario is represented by working for six hours in real-time, providing the service with one service provider, and then increasing the number of service providers until obtaining the best waiting time and queue length.

**Figure 2: General Queuing System by Arena model**

### 2.6 Utilization Rate Optimization

Utilization rates were evaluated to balance operational efficiency and service quality. A target range of 75–85% was adopted to minimize agent burnout while maintaining acceptable queue lengths. High utilization (>90%) correlated with prolonged wait times, whereas rates <70% indicated underused resources (Katta et al. 2023) (Limlawan and Anussornnitisarn, 2021) (Lim et al., 2023).

$$E = t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \quad n = \left( \frac{Z.\sigma}{E} \right)^2 \tag{2}$$

### 2.7 Simulation Replications and Statistical Rigor

To ensure statistical robustness in simulation studies, a minimum of 1000 replications per scenario is often recommended. The required number of simulation replications can be calculated using a formula that considers the desired precision of the estimates, the variability of the outcomes, and the acceptable level of Monte Carlo error. The precision of the estimates is directly related to the number of replications; more replications lead to smaller Monte Carlo errors (Chipman and Bingham, 2022). Higher replication numbers enhance the power of statistical tests, making it easier to detect significant effects (Lohmann et al., 2022). The required number of simulation replications was calculated using the following formula:

$$n = \left( \frac{Z.\sigma}{E} \right)^2 \tag{3}$$

where:

- $n$: required number of replications
- $Z$: z-value for the desired confidence level (1.96 for 95%)
- $\sigma$: standard deviation from preliminary runs
- $E$: acceptable margin of error

The margin of error $E$ was defined as:

$$E = t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \tag{4}$$

where:

- $t_{\alpha/2,n-1}$ : $t$-value for a 95% confidence level with $n - 1$ degrees of freedom
- $s$: sample standard deviation
- $n$: number of replications

This approach ensures that the reported simulation outputs are statistically robust with a confidence interval of 95%.

### 2.8 Model Verification and Validation (V&V)

Verification and validation (V&V) ensure that simulation models accurately represent real call center operations and support decision-making. In discrete-event simulation, V&V commonly relies on statistical comparison between simulated and observed data to assess model reliability (Shanmugasundaram and Banumathi, 2017). In this study, arrival and service time distributions were validated using chi-square and Kolmogorov–Smirnov tests (Haviv and Ravner, 2021). The suitability of probability distributions is assessed using quality of fit tests, such as the chi-squared test, which is one of the most commonly used methods for verifying the validity of statistical models, especially when dealing with non-perfect or controlled data, as demonstrated by the study by Aidi et al., (2021). Recent non-parametric studies indicate that relying on flexible tests without assuming prior distributions improves the reliability of time and queuing models, especially in environments characterized by uncertainty or incomplete data (Saleh and Luhayb, 2023).

Model parameters, including server numbers, input distributions, and run duration, were iteratively adjusted to identify inconsistencies. Model accuracy was evaluated using percentage error between simulated and empirical outputs, a standard measure in service system validation (Afolabi et al., 2020a) (Kuncová et al., 2020). Error rates below 10% indicate acceptable model validity and operational realism (Fonseca and Casas, 2023).

**(a)** Percentage Error Calculation

Used to compare real-world metrics and simulated metrics.

$$E_i = \left| \frac{A_i - S_i}{A_i} \right| \times 100 \tag{5}$$

where:

- $E_i$ : percentage error for metric $i$
- $A_i$ : actual (observed) value
- $S_i$ : simulated value

A small $E_i$ (e.g., below 5%) indicates good model verification.

### 2.9 Model Assumptions

i) Arrival and Service Distributions: Customer arrivals and service times follow empirical distributions (e.g., lognormal, gamma, Weibull) fitted from real data. Stationarity is not strictly assumed; averages are applied over fixed time blocks.
ii) Independence: Inter-arrival times and service durations are independent. One customer's process does not affect another.
iii) Service Discipline: First-In, First-Out (FIFO) queueing is used; no prioritization.
iv) Governorate Uniformity: All governorates share the same model structure. Input parameters differ to reflect local load and efficiency.
v) Operating Hours: Centers operate 9:00 a.m. – 2:00 p.m., Sunday to Thursday. No new arrivals after 2:00 p.m.; existing customers continue until served.

## 3. RESULTS & DISCUSSION

The data for this study were collected from fourteen call centers distributed across 14 center governorates over 220 days. The analysis includes the number of people who received service and the service waiting time analysis at six-hour intervals.

### 3.1 Dataset Overview

The data were gathered in 14 call centers located in the Governorates of Iraq during 220 working days (2018-2022). The data set consisted of volumes of services and performance measures in terms of time. There were 510,514 customers that were served within the observation timeframe. The average numbers served per day and the overall time of serving were 166 and 520 minutes (≈ 8.7 hours), respectively. Figure 3 illustrates daily service volumes, highlighting variability in customer interactions.

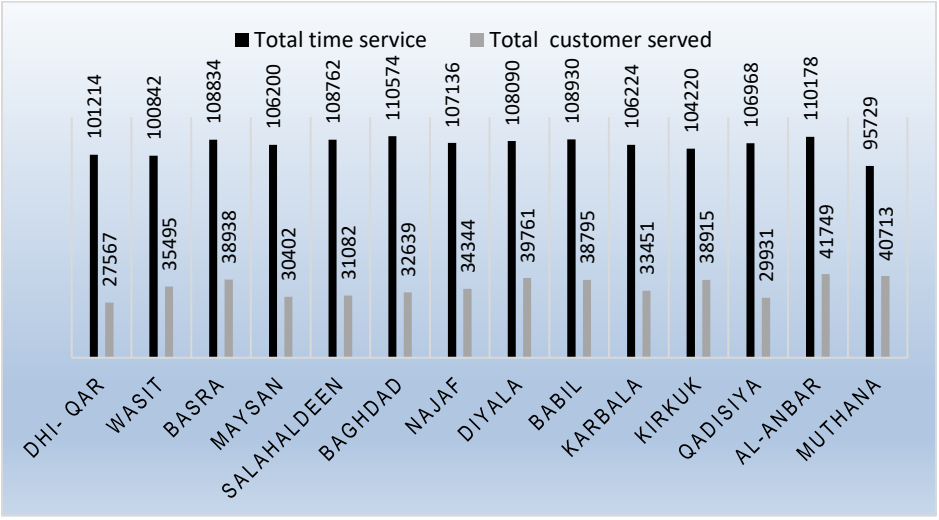


**Figure 3: Total Number of Customers and Service Time within 220 Days**

### 3.2 Outlier Detection and Data Robustness

The SPSS boxplots on the basis of interquartile range (IQR) were used to identify outliers in data on arrival and service times. To ensure stability in data while maintaining its variability, there was a Winsorization method where we replaced the values that were below the 5th percentile and those that were above the 95th percentile by the boundary values.

### 3.3 Queuing System Parameter Analysis

Goodness of fit (e.g. Kolmogorov-Smirnov and chi-square) tests were used to statistically fit arrival intervals and service times to each governorate. The empirical distributions used by different governorates were different (e.g., Lognormal, Gamma, Weibull, Erlang and Normal). Table 4 shows the statistical distributions representing

customer access and service times in each region. These results are used as primary inputs in the simulation model to ensure a realistic representation of system behavior

**Table 4**
**Best-Fit Statistical Distributions and Estimated Parameters for Inter-Arrival and Service Times by Governorate**

| State | Distribution of Interval Time | Value of Parameters | Distribution of Service Time | Parameters |
|---|---|---|---|---|
| Thi-Qar | Lognormal | $\mu = 1.53, \sigma^2 = 0.584$ | Normal | $\mu = 2.74, \sigma^2$ =0.446 |
| Wasit | Lognormal | $\mu$ 1.3, $\sigma = 0.576^2$ | Normal | $\mu = 2.4, \sigma^2 = 0.656$ |
| Basra | Weibull | $\beta = 1.12, \alpha = 2.62$ | Normal | $\mu = 0.26, \sigma^2 = 8$ |
| Maysan | Gamma | $\beta = 0.131, \alpha = 9.46$ | Erlang | $\mu = 0263, k = 8$ |
| Salah al Deen | Gamma | $\beta = 0.173, \alpha = 6.46$ | Erlang | $\mu = 2.42, k = 0.443$ |
| Baghdad | Erlang | $\mu = 0.15, k = 7$ | Normal | $\mu = 0.286, \sigma^2$=7.85 |
| Najaf | Weibull | $\beta = 1.59, \alpha = 2.21$ | Erlang | $\mu = 0.32, k = 7$ |
| Diyala | Lognormal | $\mu = 1.54, \sigma^2 = 0.57$ | Erlang | $\mu = 2.7, k = 0.431$ |
| Babil | Gamma | $\beta 0.176, \alpha = 6.1$ | Normal | $\mu = 2.37, \sigma^2 = 3.06$ |
| Karbala | Gamma | $\beta = 0.206, \alpha = 6$ | Weibull | $\beta = 2.44$, α=0.646 |
| Kirkuk | Gamma | $\beta = 0.134, \alpha = 9$ | Normal | $\mu = 0.303, \sigma^2$ =7 |
| Qadisiya | Gamma | $\beta = 0.283, \alpha = 4.54$ | Erlang | $\mu = 2.71, k = 0.436$ |
| Al-Anbar | Erlang | $\mu = 0.186, k = 8$ | Normal | $\mu = 2.76, \sigma^2 = 0.453$ |
| Muthana | Lognormal | $\mu = 1.51, \sigma^2 = 0.559$ | Normal | $\mu = 2.70, \sigma^2 = 0.443$ |

### 3.4 Governorate Performance Disparities

The single-server setting simulation revealed that waiting times and queue lengths in governorates were large. Najaf had the lowest average waiting time (30 minutes) and Qadisiya had the highest waiting time (59 minutes). Salah Al-Deen had the highest average length of the queues (53 customers). In almost all the governorates the utilization rates were near 100 years with the single channel scenario.

High $L_q$ in Salah Al-Deen and Babil correlates with prolonged wait times, suggesting inefficiencies in single-channel systems during peak demand. Utilization ($\rho$) remained near 100% for all governorates, indicating system saturation and unstable queuing dynamics under single-server conditions.

Table 5 presents the simulation results for a single-server call center, showing average wait times, queue lengths, and utilization rates, highlighting the congestion problem in this scenario.

**Table 5**
**Simulation Results for Single-Server (One-Channel) Call Center Operations Across Governorates**

| State | $W_q$ | $L_q$ | $\rho$ | Number of Customers |
|---|---|---|---|---|
| Thi-Qar | 40 | 26 | 100% | 129 |
| Wasit | 45 | 35 | 100% | 147 |
| Basra | 49 | 50 | 99% | 172 |
| Maysan | 34 | 27 | 100% | 167 |
| Salah Al-Deen | 58 | 53 | 100% | 151 |
| Baghdad | 44 | 42 | 100% | 179 |
| Najaf | 30 | 20 | 100% | 161 |
| Diyala | 36 | 23 | 100% | 133 |
| Babil | 54 | 48 | 99% | 172 |
| Karbala | 49 | 40 | 100% | 150 |
| Kirkuk | 42 | 34 | 100% | 162 |
| Qadisiya | 59 | 46 | 100% | 134 |
| Al-Anbar | 43 | 29 | 100% | 127 |
| Muthanna | 43 | 28 | 100% | 132 |

### 3.5 Validation of Simulation Accuracy

Percentage error between simulated and empirical data (Equation 5) quantified model fidelity:

$$Error\ percentage = \frac{observed\ value\ -\ true\ value}{true\ value} \times 100 \tag{5}$$

where:

- Lowest error: Basra (2.8%).
- Highest error: Salah-Al Deen (7.9%).

Most governorates exhibited errors < 8%, confirming the model's reliability (Table 6).

The validation of the model was done through a comparison of the results of the simulation and the observed data of the operation using percentage error. The error was minimal in Basra (2.8%), and the biggest error was in Salah-Al Deen (7.9%). The validation errors were less than 8 in most governorates, which is a pointer of close similarities in the performance of the simulated system and the real system.

**Table 6**
**Comparison between Simulated Outputs and Observed Call Center Data for Model Validation**

| State | Simulation Data | Actual Data | Percentage of Error |
|---|---|---|---|
| Thi-Qar | 129 | 125 | 3.2 |
| Wasit | 147 | 155 | 5.2 |
| Basra | 172 | 177 | 2.8 |
| Maysan | 167 | 180 | 7.2 |
| Salah-Aldeen | 151 | 140 | 7.9 |
| Baghdad | 179 | 168 | 6.5 |
| Najaf | 161 | 155 | 3.9 |
| Diyala | 133 | 126 | 5.6 |
| Babil | 172 | 185 | 7.0 |
| Karbala | 150 | 157 | 4.5 |
| Kirkuk | 162 | 168 | 3.6 |
| Qadisiya | 134 | 141 | 5.0 |
| Al-Anbar | 128 | 133 | 3.8 |
| Muthanna | 132 | 139 | 5.0 |

### 3.6 Implications

Multi-server configurations were tested in the simulation where the service capacity of one channel was extended to two and three channels. The waiting and queues times and length were shorter in the dual- and triple-channel layouts than in the single-server one within governorates.

- Operational Scalability: Multi-channel systems (M/M/c) resolve congestion in high-demand Governorates (e.g., Salah Al-deen). Figure 2. ARENA simulation workflow for single- and multi-channel queuing systems
- Resource Allocation: Utilization rates guide staffing decisions to balance cost and service levels.
- Simulation Utility: Arena's DES framework enables rapid scenario testing, supporting data-driven policy design.
- While triple-channel systems improved service quality, they increased operational costs without proportionate performance gains.". Dual-channel models offered the best balance between cost and efficiency, reducing wait times by 60%.

Table 6 present higher error in Maysan Salah-Aldeen, Baghdad and Babil was attributed to inconsistent data recording and agent behavior not captured by the base model. Future versions will incorporate Governorateal calibration for higher fidelity.

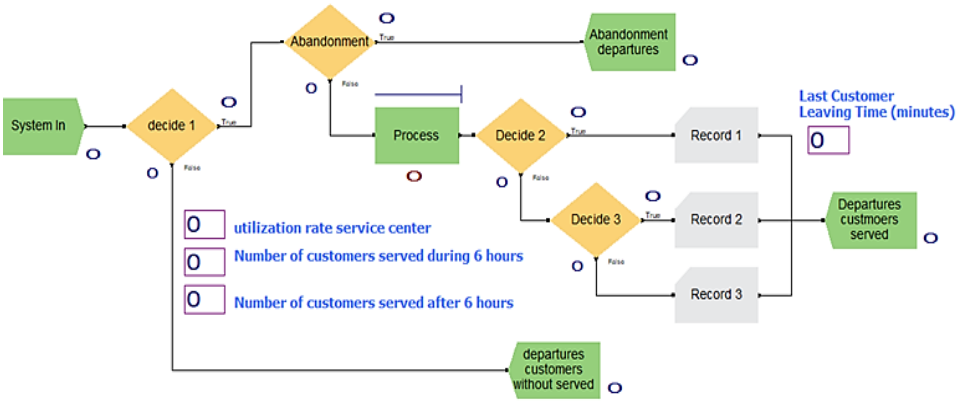


**Figure 4: Final Model Created in Arena Simulation**

### 3.7 Multi-Channel Optimization

Introducing dual- and triple-channel systems (Table 7) significantly improved performance:

- Dual channel:
  - Shortest $W_q$: Diyala (0.17 min), $L_q$: 0.11 customers.
  - Longest $W_q$: Karbala (1.3 min), $L_q$: 1.1 customers.
- Triple-channel: Further reduced $W_q$ and $L_q$ by 40–60%, aligning with M/M/c queuing theory.

Enhanced service capacity (2–3 channels) balanced arrival-service rate ratios, stabilizing queues and reducing wait times. Figure 4 (Arena workflow) details decision points (e.g., customer flow, agent utilization) critical to optimizing operational efficiency.

A target range of 75–85% minimizes agent burnout while maintaining service quality, as exceeding 90% correlates with customer dissatisfaction and abandonment.

Table 7 examines the impact of increasing the number of servers to two and three on call center performance, showing a significant reduction in wait times and queue lengths compared to a single-server system.

**Table 7**
**Queue Performance Metrics for Dual and Triple Server Call Center Configurations by Governorate**

| State | 2 Channels | | 3 Channels | |
|---|---|---|---|---|
| | $Wq$ | $Lq$ | $Wq$ | $Lq$ |
| Thi-Qar | 0.22 | 0.29 | 0.005 | 0.003 |
| Wasit | 0.34 | 0.25 | 0.015 | 0.012 |
| Basra | 0.82 | 0.85 | 0.046 | 0.048 |
| Maysan | 0.24 | 0.2 | 0.003 | 0.002 |
| Salah-Aldeen | 1.20 | 1.11 | 0.028 | 0.025 |
| Baghdad | 0.27 | 0.25 | 0.020 | 0.023 |
| Najaf | 0.28 | 0.2 | 0.020 | 0.014 |
| Diyala | 0.17 | 0.22 | 0.003 | 0.002 |
| Babil | 0.55 | 0.52 | 0.035 | 0.032 |
| Karbala | 1.70 | 1.51 | 0.026 | 0.021 |
| Kirkuk | 0.22 | 0.18 | 0.008 | 0.006 |
| Qadisiya | 1.30 | 1.07 | 0.041 | 0.033 |
| Al-Anbar | 0.17 | 0.11 | 0.007 | 0.004 |
| Muthanna | 0.20 | 0.13 | 0.004 | 0.003 |

### 3.8 Utilization Rate Analysis

Table 8 evaluates resource efficiency across channel configurations each governorate was reported to be utilized with 1-channel, 2-channel and 3-channel configurations. The utilization in the single channel system was around 99-100%. With the dual-channel system utilization declined to about 60-76, with the triple-channel system declining to about 43-51.

Table 8 presents staff utilization rates under different server scenarios, helping to determine the appropriate balance between operational efficiency and service quality.

**Table 8**
**Server Utilization Rates under Single, Dual and Triple-Channel Call Center Configurations**

| State | 1 Channel | 2 Channels | 3 Channels |
|---|---|---|---|
| Dhi-Qar | 100% | 65% | 44% |
| Wasit | 100% | 65% | 44% |
| Basra | 99% | 70% | 47% |
| Maysan | 100% | 71% | 46% |
| Salah-Aldeen | 100% | 76% | 51% |
| Baghdad | 100% | 71% | 48% |
| Najaf | 100% | 66% | 43% |
| Diyala | 100% | 60% | 51% |
| Babil | 99% | 76% | 50% |
| Karbala | 100% | 69% | 47% |
| Kirkuk | 100% | 64% | 44% |
| Qadisiya | 100% | 76% | 50% |
| Al-Anbar | 100% | 68% | 45% |
| Muthanna | 100% | 65% | 43% |

## 4. DISCUSSION

- The single server case led to long queues and huge wait lines since the system was running at almost full capacity. At utilization levels of approximately 100, the system is very sensitive to variations in arrival rates and service time and results in chronic congestion and unstable queue dynamics.
- High Waiting Times Implication.
  Waiting time of over 50 minutes in some of the governorates is a sign that there is a significant inefficiency in service provision, which could have adverse impacts on customer satisfaction and perceived quality of service. Most severe congestion was observed in Governorates having a high demand (e.g. Qadisiya and Salah Al-Deen) and this indicates that the same staffing level in the regions is not adequate when there is heterogeneous demand.
- Reason why Multi-Server Arrays enhanced performance of the System.
  The findings indicate that more servers enhanced the performance in terms of service capacity and imbalance of arrival rates and service rates was reduced. This, consequently, lessened the number of queues and shortened waiting time by a significant margin. This illustrates that the multi-server deployment is an efficient operation strategy of high-load governorates.
- Dual vs. Triple Servers: Cost-Performance Trade-Off.
  Even though triple-server setups caused delays to more significantly decrease, the difference between triple-server and dual-server situations might not be significant in relation to the incremental staffing cost. It means that the benefit of the third server decreases, and the best option is the duo server since it is the most feasible in the context of both quality of services and the cost of their operation being balanced in most governorates.
- Reliable and Practical Applications of the Simulation Model.
  The reliability of the simulation model at modeling the actual call center operations is supported by the validation results (errors are less than 8% in the majority of the cases). This degree of precision implies that the model can be adopted as a decision-support mechanism to test the staffing policies and capacity changes without interfering with the actual processes.
- Implications to Managers and Policies.
  The results endorse the change towards the need-based staffing among the governorates as opposed to equal distribution of resources. A two-server base-line and three-server capacity during peak usage can offer a valuable approach to mitigate the congestion and enhance the service level and ensure the sustainable rate of utilization.

## 5. CONCLUSION

This research offers several novel contributions compared to previous studies. First, it relies on real operational data from 14 Iraqi governorates, a resource rarely explored in contact center studies in developing countries. Second, it employs general distributions

(G/G/1 and G/G/c) derived from actual data instead of relying on traditional exponential distribution assumptions, thus enhancing the model's realism. Third, the study incorporates Winsorization to address outliers, improving simulation accuracy and reducing the error rate to less than 7%. Furthermore, the study provides a clear quantitative comparison between single-channel and multi-channel systems, practically identifying the optimal configuration in terms of efficiency and cost. These aspects make the proposed model a valuable tool for supporting managerial decisions.

Based on the simulation results, the study recommends adopting a dual-server system as the primary option for most contact centers, as it achieves a good balance between wait times and utilization rates. Maintaining staff utilization rates within the 75–85% range is advised to avoid operational strain and reduced service quality. Regular simulation testing of increased demand scenarios before actual implementation is also recommended. Finally, the quality of operational data recording should be improved to ensure greater accuracy in future models and to support planning decisions efficiently. The model can be improved by using time-dependent arrival rates, customer impatience (abandonment), and more flexible schedules of the staff in the future. The generalizability of the findings would also be enhanced with the help of further validation with the help of other datasets and extended operation periods.

## ACKNOWLEDGMENT

The authors acknowledge the editors and reviewers of Pakistan Journal of Statistics, as well as Universiti Sains Malaysia.

**Conflicts of Interest**

The authors have no relevant financial or non-financial interests to disclose.

## REFERENCES

1. Afolabi, A.O., Ganiyu, R.A., Arulogun, O.T. and Ganiyu, A.O. (2020). Development of a model for an inbound call centre using hierarchical timed coloured Petri nets. *Asian Journal of Research in Computer Science*, 5(3), 39-51. https://doi.org/10.9734/ajrcos/2020/v5i330137
2. Amouzegar, M. and Moshirvaziri, K. (2006). A simulation framework for networked queue models: Analysis of queue bounds in a G/G/c supply chain. *Journal of Applied Mathematics and Decision Sciences*, 2006, 1-13. https://doi.org/10.1155/JAMDS/2006/87514
3. Batarseh, O.G. and Wang, Y. (2018). An interval-based approach to model input uncertainty in M/M/1 simulation. *International Journal of Approximate Reasoning*, 95, 46-61. https://doi.org/10.1016/j.ijar.2017.12.007
4. Bolandifar, E., DeHoratius, N. and Olsen, T. (2023). Modeling abandonment behavior among patients. *European Journal of Operational Research*, 306(1), 243-254. https://doi.org/10.1016/j.ejor.2022.08.023
5. Chipman, H. and Bingham, D. (2022). Let's practice what we preach: Planning and interpreting simulation studies with design and analysis of experiments. *Canadian Journal of Statistics*, 50(4), 1228-1249. https://doi.org/10.1002/cjs.11719

6. Delgado, M.A. and Vainora, J. (2022). Conditional distribution model specification testing using chi-square goodness-of-fit tests. *Journal of Econometrics*, 229(1), 1-15.
7. Ferrari, S.C. and Morabito, R. (2020). Application of queueing models with abandonment for call center congestion analysis. *Gestão & Produção*, 27(1), 1-15. https://doi.org/10.1590/0104-530X3765-20
8. Fitrianto, A., Muhamad, W.Z.A.W., Kriswan, S. and Susetyo, B. (2022). Comparing outlier detection methods using boxplot generalized extreme studentized deviate and sequential fences. *Aceh International Journal of Science and Technology*, 11(1), 38-45. https://doi.org/10.13170/aijst.11.1.23809
9. Fonseca i Casas, P. (2023). A continuous process for validation, verification, and accreditation of simulation models. *Mathematics*, 11(4), 845. https://doi.org/10.3390/math11040845
10. Guseva, E., Varfolomeyeva, T., Efimova, I. and Movchan, I. (2018). Discrete event simulation modelling of patient service management with Arena. *Journal of Physics: Conference Series*, 1015(3), 032095. https://doi.org/10.1088/1742-6596/1015/3/032095
11. Haviv, M. and Ravner, L. (2021). A survey of queueing systems with strategic timing of arrivals. *Queueing Systems*, 99(1-2), 163-198. https://doi.org/10.1007/s11134-021-09717-8
12. Johnson, O.V. and Ganiyu, M. (2019). Simulation approach to life-data queue event modeling. *Control Theory and Informatics*, 8, 32-47. https://doi.org/10.7176/CTI/8-04
13. Katta, R., Strouphauer, E., Ibraheim, M.K., Li-Wang, J. and Dao, H. (2023). Practice efficiency in dermatology: Enhancing quality of care and physician well-being. *Cureus*, 15(5), e39195. https://doi.org/10.7759/cureus.39195
14. Kumar Rajuwar, M. and Kalita, D. (2018). Simulation of queuing system for car service center using Arena simulation software. *International Journal of Production Engineering*, 4(2), 1-12.
15. Kuncová, M., Fábry, J. and Klímová, A.M. (2020). Discrete event simulation: Model of a call center in SIMUL8 software. *Proceedings of the European Council for Modelling and Simulation*, 34(1), 48-55. https://doi.org/10.7148/2020-0048
16. Lim, G., Lim, A.J., Quinn, B., Carvalho, B., Zakowski, M. and Lynde, G.C. (2023). Obstetric operating room staffing and operating efficiency using queueing theory. *BMC Health Services Research*, 23(1), 1-17. https://doi.org/10.1186/s12913-023-10143-0
17. Limlawan, V. and Anussornnitisarn, P. (2021). Enhance system utilization and business revenue with AI-based queue reservation system. *International Journal of Machine Learning and Computing*, 11(3), 236-241. https://doi.org/10.18178/ijmlc.2021.11.3.1041
18. Lohmann, A., Astivia, O.L.O., Morris, T.P. and Groenwold, R.H.H. (2022). It's time! Ten reasons to start replicating simulation studies. *Frontiers in Epidemiology*, 2, 973470. https://doi.org/10.3389/fepid.2022.973470
19. Oladimeji, A.S. and Ibidoja, O.J. (2020). The distribution of service time of patients. *Journal of Reliability and Statistical Studies*, 13(1), 61-72. https://doi.org/10.13052/jrss0974-8024.1313
20. Pankratova, E., Moiseeva, S. and Farkhadov, M. (2022). Infinite-server resource queueing systems with different types of Markov-modulated Poisson process and renewal arrivals. *Mathematics*, 10(16), 2962. https://doi.org/10.3390/math10162962

21. Piao, L., Zheng, Y. and Wang, L. (2023). Numerical comparison of the three Poisson arrival queuing system models. *Frontiers in Artificial Intelligence and Applications*, 378, 524-532. https://doi.org/10.3233/FAIA231060
22. Shanmugasundaram, S. and Banumathi, P. (2017). A comparative study on M/M/1 and M/M/C queueing models using Monte Carlo simulation. *Global Journal of Pure and Applied Mathematics*, 13(11), 7843-7853.
23. Yelvita, F.S. (2022). Queueing and simulation strategies for enhancing call center efficiency and customer satisfaction. *International Journal of Service Science*, 9(2), 101-115.