ANFIS computing and cost optimization of an $M/M/c/M$ queue with feedback and balking customers under a hybrid hiatus policy

Aimen Dehimi [1], **Mohamed Boualem**[2,*], **Sami Kahla**[3] **and Louiza Berdjoudj**[4]

[1] *University of Bejaia, Faculty of Exact Sciences, Applied Mathematics Laboratory, 06000 Bejaia, Algeria*
E-mail: ⟨aimen.dehimi@univ-bejaia.dz⟩

[2] *University of Bejaia, Faculty of Technology, Research Unit LaMOS, 06000 Bejaia, Algeria*
E-mail: ⟨mohammed.boualem@univ-bejaia.dz⟩

[3] *Research Center in Industrial Technologies, P.O. Box 64, 16014 Cheraga, Algeria*
E-mail: ⟨samikahla40@yahoo.com⟩

[4] *University of Bejaia, Faculty of Exact Sciences, Research Unit LaMOS, 06000 Bejaia, Algeria*
E-mail: ⟨louiza.berdjoudj@univ-bejaia.dz⟩

Abstract. The present investigation studies a hybrid hiatus policy for a finite-space Markovian queue, incorporating realistic features such as Bernoulli feedback, multiple servers, and balking customers. A hybrid hiatus policy combines both a working hiatus and a complete hiatus. As soon as the system becomes empty, the servers switch to a working hiatus. During a working hiatus, the servers operate at a reduced service rate. Upon completion of the working hiatus and in the absence of waiting customers, the servers enter a complete hiatus. Once the complete hiatus period concludes, the servers resume normal operations and begin serving waiting customers. In the context of Bernoulli feedback, the dissatisfied customer can re-enter the system to receive another service. By utilizing the Markov recursive approach, we examined the steady-state probabilities of the system and queue sizes and other queueing indices, viz. Average queue length, average waiting time, throughput, etc. Using the Quasi-Newton method, a cost function is developed to determine the optimal values of the system's decision variables. Furthermore, a soft computing approach based on an adaptive neuro-fuzzy inference system (ANFIS) is employed to validate the accuracy of the obtained results.

Keywords: ANFIS computing, feedback multi-server queue, hybrid hiatus policy, optimization, recursive approach

Received: July 9, 2024; accepted: August 14, 2024; available online: October 7, 2024

DOI: 10.17535/crorr.2024.0013

1. Introduction

Queueing models with server vacations find extensive applications across various real-life systems such as telecommunications, data and voice transmission networks, and production systems. Over the past several decades, significant research efforts have been dedicated to these models, resulting in comprehensive surveys and seminal works [4, 10, 16, 22] and references therein.

One notable advancement in this area is the introduction of working vacation policies, where servers continue to operate at reduced rates during vacation periods. This concept was first proposed by [20], marking a pivotal development in queueing theory. Extensive literature has

*Corresponding author.

since explored various queueing models incorporating working vacations across diverse contexts [1, 2, 8, 11]. Another significant scenario is queueing models with vacations under balking, prevalent in manufacturing systems, call centers, and transportation networks. Recent research has focused on multi-server systems with impatient customers under both multiple and single vacation policies [16], bulk arrival queueing models with variant working vacation and impatience [6], a finite-capacity discrete-time multi-server queue with synchronous single and multiple working vacations, Bernoulli feedback, and impatient customers [25], and differentiated working vacation policies with impatient customers in single-server queues [7].

In the realm of multi-server queueing models with vacation, two primary types exist: synchronous vacations where all servers take vacations simultaneously [3, 5, 18], and asynchronous vacations where servers take vacations independently [14, 17]. However, despite their practical relevance, these systems are complex, and there remains a gap in their detailed analysis. In this study, we introduce a novel operational policy known as the hybrid hiatus, as proposed by Vadivukarasi and Kalidass [23]. This policy involves servers alternating between two operational states: a working hiatus period, where they operate at reduced capacity, and a complete hiatus period, where no services are provided. The decision to switch between these states depends on real-time queue dynamics and system conditions. For instance, in a hospital emergency department, during a working hiatus, the department operates with reduced staff. If patient demand is low, the department may temporarily close until demand increases, or continue operating at a reduced capacity if immediate care is needed. This policy aims to enhance resource efficiency and responsiveness in dynamic service environments.

To tackle the complexities of these systems, intelligent systems like the Adaptive Neuro-Fuzzy Inference System (ANFIS) have been employed. ANFIS integrates fuzzy logic with neural networks to model nonlinear systems effectively. It has been widely applied for categorization, prediction, control, and optimization tasks, including transient analysis in queueing systems. Initially proposed by [15], ANFIS has made significant contributions to queueing theory [9, 12, 13, 21, 24], enabling researchers to compare analytical formulas with ANFIS-generated numerical outcomes for enhanced system understanding.

In this investigation, we study a finite-capacity Markovian multi-server queue with balking and feedback, governed by a hybrid hiatus policy consisting of a working hiatus and a complete hiatus. When the system becomes empty, the servers transition to a working hiatus, where they serve customers at a reduced rate. Upon completing the working hiatus and with no waiting customers, the servers opt for a complete hiatus. Once the complete hiatus concludes, the servers return to their normal operational state to serve waiting customers.Using the Markov recursive approach, we analyze the steady-state probabilities of the system and queue sizes, along with various queueing metrics such as the expected number of customers in the system and queue, expected waiting times, expected balking rates, and probabilities associated with different server states. We develop a cost function to optimize the system's decision variables using the Quasi-Newton method. Furthermore, we employ a soft computing approach based on an adaptive neuro-fuzzy inference system (ANFIS) to validate the accuracy of our findings.

The structure of this paper is outlined as follows: Section 2 provides a model description and associated mathematical assumptions. In Section 3, we establish the steady-state solution of the model using the recursive method. Section 4 presents explicit formulas for queueing metrics and discusses the ANFIS approach. In Section 5, we introduce the cost model formulation. Section 6 includes numerical illustrations and discusses cost optimization. Finally, Section 7 presents general conclusions and perspectives.

2. Mathematical formulation of the model

We consider a finite capacity multi-server $M/M/c/M$ queueing system with balking customers, hybrid hiatus, and feedback. Key assumptions underlying this model include:

- Customers enter the system following a Poisson process with rate λ.
- During normal busy periods, service times are exponentially distributed with rate β.
- Service times slow down during working hiatus periods, modeled by an exponential distribution with rate α ($\alpha < \beta$).
- Customers are served based on the FCFS (First-Come-First-Served) discipline, and the system has a finite capacity, denoted as M, with c servers.
- Upon arrival, a customer finds the system in one of several states: on hiatus (no servers available), during a normal busy period, or during a working hiatus. The customer decides to either join the queue with probability κ or balk with probability $\kappa' = 1 - \kappa$.
- A hybrid hiatus involves both a working hiatus (WH) and a complete hiatus (CH). When the system becomes empty, servers transition to WH where they operate at a reduced service rate. After completing the working hiatus and if there are waiting customers, servers return to normal busy mode to serve them. If no customers are waiting, servers move to a complete hiatus. Once the CH concludes, servers return to normal operation to attend to any waiting customers.
- If a customer is dissatisfied with the service provided, they have two options: they can leave the system with a probability q, or return later with a probability $q' = 1 - q$. Feedback customers returning later are treated as new arrivals in the system.

The introduced variables are independent of each other.

2.1. Practical motivations

Several key operational dynamics are explored in this study of an $M/M/c/M$ queueing system applied to a hospital emergency department scenario. Patients arrive according to a Poisson process, seeking medical attention serviced with exponentially distributed times under normal conditions (β), and slower times during working hiatuses ($\alpha < \beta$). The department has a finite capacity M, and patients may balk upon arrival if all treatment rooms are occupied, governed by a probability κ. Hybrid hiatuses are implemented, where during working hiatus periods, the department operates at reduced capacity and may transition to complete hiatus if no patients are waiting. Otherwise, it will return to its usual busy state and start serving patients. Patients dissatisfied with wait times or the quality of care can choose to leave (with a probability q) or return later (with a probability $q' = 1 - q$), treated as new arrivals upon their return. This model provides insights into optimizing emergency department operations by managing patient flow, resource utilization, and service quality in dynamic healthcare environments.

3. Steady-state Solution

Let us consider the bivariate process $\{(A(t), N(t)), t \geq 0\}$, where $A(t)$ denotes the number of customers in the system at time t, and $N(t)$ represents the state of the servers at time t, taking one of three values: $N(t) = 0$ when the servers are in normal busy period at time t, $N(t) = 1$ when the servers are in working hiatus period at time t, and $N(t) = 2$ when the servers are in complete hiatus period at time t.

The joint probability $P_{m,j} = \lim_{t\to\infty} P\{A(t) = m, N(t) = j, (m, j) \in \Omega\}$ denotes the steady-state probabilities of the system. Figure 1 depicts the transition diagram of the considered model.

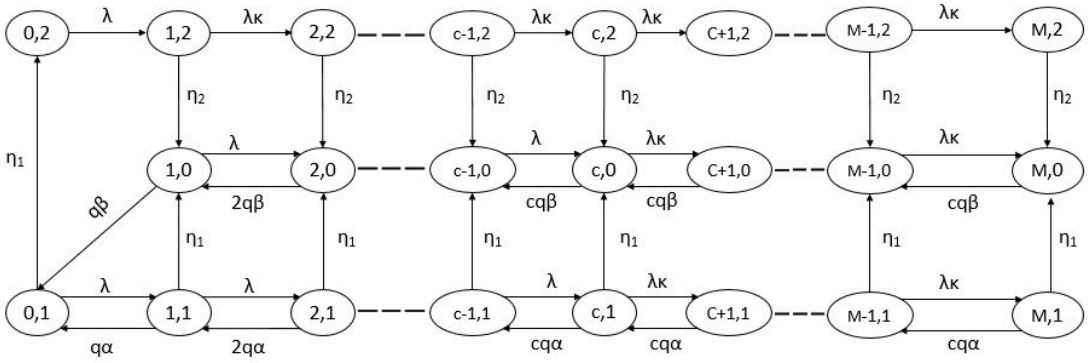

Figure 1: *State transition rate diagram.*

Using the principle of balance equations

$$\lambda P_{0,2} = \eta_1 P_{0,1}, \qquad m = 0, \tag{1}$$

$$(\lambda\kappa + \eta_2)P_{1,2} = \lambda P_{0,2}, \qquad m = 1, \tag{2}$$

$$(\lambda\kappa + \eta_2)P_{m,2} = \lambda\kappa P_{m-1,2}, \qquad 2 \le m \le M-1, \tag{3}$$

$$\lambda\kappa P_{M-1,2} = \eta_2 P_{M,2}, \qquad m = M, \tag{4}$$

$$(\lambda + q\beta)P_{1,0} = 2q\beta P_{2,0} + \eta_1 P_{1,1} + \eta_2 P_{1,2}, \ \ m = 1, \tag{5}$$

$$(\lambda + mq\beta)P_{m,0} = \lambda P_{m-1,0} + (m+1)q\beta P_{m+1,0} + \eta_1 P_{m,1} + \eta_2 P_{m,2}, \ \ 2 \le m \le c-1, \tag{6}$$

$$(\lambda\kappa + cq\beta)P_{c,0} = \lambda P_{c-1,0} + cq\beta P_{c+1,0} + \eta_1 P_{c,1} + \eta_2 P_{c,2}, \tag{7}$$

$$(\lambda\kappa + cq\beta)P_{m,0} = \lambda\kappa P_{m-1,0} + cq\beta P_{m+1,0} + \eta_1 P_{m,1} + \eta_2 P_{m,2}, \ \ c+1 \le m \le M-1, \tag{8}$$

$$cq\beta P_{M,0} = \lambda\kappa P_{M-1,0} + \eta_1 P_{M,1} + \eta_2 P_{M,2}, \tag{9}$$

$$(\lambda + \eta_1)P_{0,1} = \alpha q P_{1,1} + q\beta P_{1,0}, \ \ m = 0, \tag{10}$$

$$(m\alpha q + \lambda + \eta_1)P_{m,1} = \lambda P_{m-1,1} + (m+1)q\alpha P_{m+1,1}, \ \ 1 \le m \le c-1, \tag{11}$$

$$(\lambda\kappa + cq\alpha + \eta_1)P_{c,1} = \lambda P_{c-1,1} + cq\alpha P_{c+1,1}, \tag{12}$$

$$(\lambda\kappa + cq\alpha + \eta_1)P_{m,1} = \lambda\kappa P_{m-1,1} + c\alpha q P_{m+1,1}, \ \ c+1 \le m \le M-1, \tag{13}$$

$$(cq\alpha + \eta_1)P_{M,1} = \lambda\kappa P_{M-1,1}, \tag{14}$$

The normalizing condition is

$$\sum_{m=0}^{M} (P_{m,0} + P_{m,1} + P_{m,2}) = 1. \tag{15}$$

Now, we present the solution of the equations above in the following theorem.

Theorem 1. *The probabilities describing the system size in different operational periods, namely the hiatus period ($P_{m,2}$), working hiatus period ($P_{m,1}$), and normal busy period ($P_{m,0}$), in the steady-state are respectively expressed as follows:*

$$P_{m,2} = \Lambda_m P_{M,2} = \Lambda_m \left(\sum_{m=0}^{M} (\Lambda_m + \theta_1 \chi_m) + \sum_{m=1}^{M} (\theta_2 \Upsilon_m - \delta_m) \right)^{-1}, \ m = 0, 1, 2, ..., M, \tag{16}$$

$$P_{m,1} = \theta_1 \chi_m P_{M,2}, \tag{17}$$

$$P_{m,0} = (\theta_2 \Upsilon_m - \delta_m) P_{M,2}, \tag{18}$$

where

$$\Lambda_m = \begin{cases} 1, & m = M, \\ \frac{\eta_2}{\lambda\kappa}, & m = M-1, \\ \frac{\lambda\kappa+\eta_2}{\lambda\kappa}\Lambda_{m+1}, & 0 \le m \le M-2, \end{cases} \tag{19}$$

$$\chi_m = \begin{cases} 1, & m = M, \\ \frac{cq\alpha+\eta_1}{\lambda\kappa}, & m = M-1, \\ \frac{\lambda\kappa+cq\alpha+\eta_1}{\lambda\kappa}\chi_{m+1} - \frac{cq\alpha}{\lambda\kappa}\chi_{m+2}, & c \le m \le M-1, \\ \frac{\lambda\kappa+(m+1)q\alpha+\eta_1}{\lambda}\chi_{m+1} - \frac{(m+1)q\alpha}{\lambda}\chi_{m+2}, & m = c-1, \\ \frac{\lambda+(m+1)q\alpha+\eta_1}{\lambda}\chi_{m+1} - \frac{(m+2)q\alpha}{\lambda}\chi_{m+2}, & 0 \le m \le c-2, \end{cases} \tag{20}$$

$$\theta_1 = \frac{\lambda\Lambda_0}{\eta_1\chi_0}, \tag{21}$$

$$\Upsilon_m = \begin{cases} 1, & m = M, \\ \frac{cq\beta}{\lambda\kappa}, & m = M-1, \\ \frac{\lambda\kappa+cq\beta}{\lambda\kappa}\Upsilon_{m+1} - \frac{cq\beta}{\lambda\kappa}\Upsilon_{m+2}, & c \le m \le M-1, \\ \frac{\lambda\kappa+(m+1)q\beta}{\lambda}\Upsilon_{m+1} - \frac{(m+1)q\beta}{\lambda}\Upsilon_{m+2}, & m = c-1, \\ \frac{\lambda+(m+1)q\beta}{\lambda}\Upsilon_{m+1} - \frac{(m+2)q\beta}{\lambda}\Upsilon_{m+2}, & 1 \le m \le c-2, \end{cases}$$

$$\delta_m = \begin{cases} 0, & m = M, \\ \frac{\eta_1\theta_1+\eta_2}{\lambda\kappa}, & m = M-1, \\ \frac{\theta_1\eta_1\chi_{m+1}+\eta_2\Lambda_{m+1}}{\lambda\kappa}, & c \le m < M-1, \\ \frac{\theta_1\eta_1\chi_{m+1}+\eta_2\Lambda_{m+1}}{\lambda}, & m = c-1, \\ \frac{\theta_1\eta_1\chi_{m+1}+\eta_2\Lambda_{m+1}}{\lambda}, & 1 \le m \le c-2, \end{cases}$$

$$\theta_2 = \frac{\theta_1(\lambda+\eta_1)\chi_0 - \theta_1 q\alpha\chi_1 + q\beta\delta_1}{q\beta\Upsilon_1}, \tag{22}$$

and

$$P_{M,2} = \left(\sum_{m=0}^{M} (\Lambda_m + \theta_1\chi_m) + \sum_{m=1}^{M} (\theta_2\Upsilon_m - \delta_m) \right)^{-1}. \tag{23}$$

4. Metrics of system performance

▷ The probabilities associated with different server states–normal busy period, working hiatus, and hiatus–are defined as follows:

$$P_{rb} = P_{M,2} \sum_{m=1}^{M} (\theta_2 \Upsilon_m - \delta_m), \tag{24}$$

$$P_{wh} = \theta_1 P_{M,2} \sum_{m=0}^{M} \chi_m, \tag{25}$$

$$P_h = P_{M,2} \sum_{m=0}^{M} \Lambda_m. \tag{26}$$

▷ The expressions for the expected number of customers in the system (L_s) and in the queue (L_q) are defined as follows:

$$L_s = P_{M,2} \left[\theta_2 \sum_{m=1}^{M} m\Upsilon_m - \sum_{m=1}^{M} m\delta_m + \theta_1 \sum_{m=1}^{M} m\chi_m + \sum_{m=1}^{M} m\Lambda_m \right], \tag{27}$$

$$L_q = P_{M,2} \left[\theta_2 \sum_{m=c}^{M} (m-c)\Upsilon_m - \sum_{m=c}^{M} (m-c)\delta_k + \theta_1 \sum_{m=c}^{M} (m-c)\chi_m + \sum_{m=1}^{M} m\Lambda_k \right]. \tag{28}$$

▷ The expected balking rate:

$$B_r = \lambda P_{M,2} \left[\theta_2 \sum_{m=c}^{M} \kappa^{'} \Upsilon_m - \sum_{m=c}^{M} \kappa^{'} \delta_m + \theta_1 \sum_{m=c}^{M} \kappa^{'} \chi_m + \sum_{m=c}^{M} \kappa^{'} \Lambda_m \right]. \tag{29}$$

▷ The expressions for the expected waiting time of customers in the system (W_s) and in the queue (W_q) are given by:

$$W_s = \frac{L_s}{\lambda'}, \quad \text{where} \quad \lambda^{'} = \lambda - B_r, \tag{30}$$

$$W_q = \frac{L_q}{\lambda'}. \tag{31}$$

4.1. Adaptive neuro-fuzzy inference system

The Adaptive Neuro-Fuzzy Inference System (ANFIS), as proposed in [15], combines the principles of fuzzy logic and neural networks to create a powerful tool capable of modeling complex systems. ANFIS operates on a multilayer architecture using Takagi-Sugeno fuzzy inference rules, allowing it to handle multiple inputs and outputs simultaneously with the aid of fuzzy parameters. This approach enables ANFIS to dynamically learn and interpret intricate patterns in both linear and nonlinear relationships. By employing Gaussian functions for membership and utilizing Sugeno-type systems, ANFIS constructs fuzzy if-then rules that are trained using paired input-output data. This training process ensures ANFIS can swiftly adapt and optimize its performance across diverse applications, including telecommunications, atmospheric research, and traffic management.

5. Cost optimization

To construct the cost model, we consider the following cost elements associated with various events:

- $C_{rb}-$ cost per unit time when the servers are in normal busy period,
- C_h- cost per unit time when the servers are in working hiatus period or on hiatus period,
- C_r- cost per unit time when a customer balks,
- C_β (resp. C_α)$-$ cost per service per unit time during normal busy period (resp. during working hiatus period),
- $C_{s-f}-$ cost per service per unit time for a feedback customer,
- C_b- fixed purchase cost of the server per unit.

Our primary objective is to define the total expected cost per unit time for the system in this context:

$$G(\beta,\alpha) = C_{rb}P_{rb} + C_h(P_{wh} + P_h) + C_rB_r + c\beta C_\beta + c\alpha C_\alpha + cq^{'}(\beta+\alpha)C_{s-f} + cC_b.$$

6. Numerical simulation

This section centers on the numerical evaluation of diverse performance metrics within the proposed queueing model, accomplished through parameter variation. It further illustrates how practitioners can effectively utilize and interpret the resultant findings.

6.1. Performance metrics analysis

In this part, we obtain some various performance measures of interest that are computed under different scenarios by using a MATLAB program.

(η_1,η_2)	L_s	P_{rb}	P_{wh}	P_h
(0.5,0.6)	3.1363	0.6283	0.3586	0.0131
(0.6,0.7)	3.0172	0.6695	0.3161	0.0144
(0.7,0.8)	2.9379	0.7022	0.2824	0.0155
(0.8,0.9)	2.8832	0.7286	0.2550	0.0164
(0.9,1.0)	2.8445	0.7505	0.2323	0.0172

Table 1: *Impact of working hiatus and hiatus rates* (η_1,η_2) *when* $\lambda = 6$, $\kappa = 0.4$, $\beta = 2.5$, $\alpha = 1$, $c = 3$, $M = 12$, $q = 0.7$.

κ	W_s	L_q	B_r
0.1	0.7328	0.0488	1.2949
0.3	0.7711	0.2189	1.1412
0.5	0.8288	0.5123	0.9164
0.7	0.8983	0.9291	0.6146
0.9	0.9787	1.4841	0.2304

Table 2: *Impact of non-balking probability* κ *when* $\lambda = 6$, $\eta_1 = 0.5$, $\eta_2 = 0.8$, $\beta = 2.5$, $\alpha = 1$, $c = 3$, $M = 12$, $q = 0.7$.

Figure 2 presents the Gaussian function used to select fuzzy input parameters, like λ, β, α.

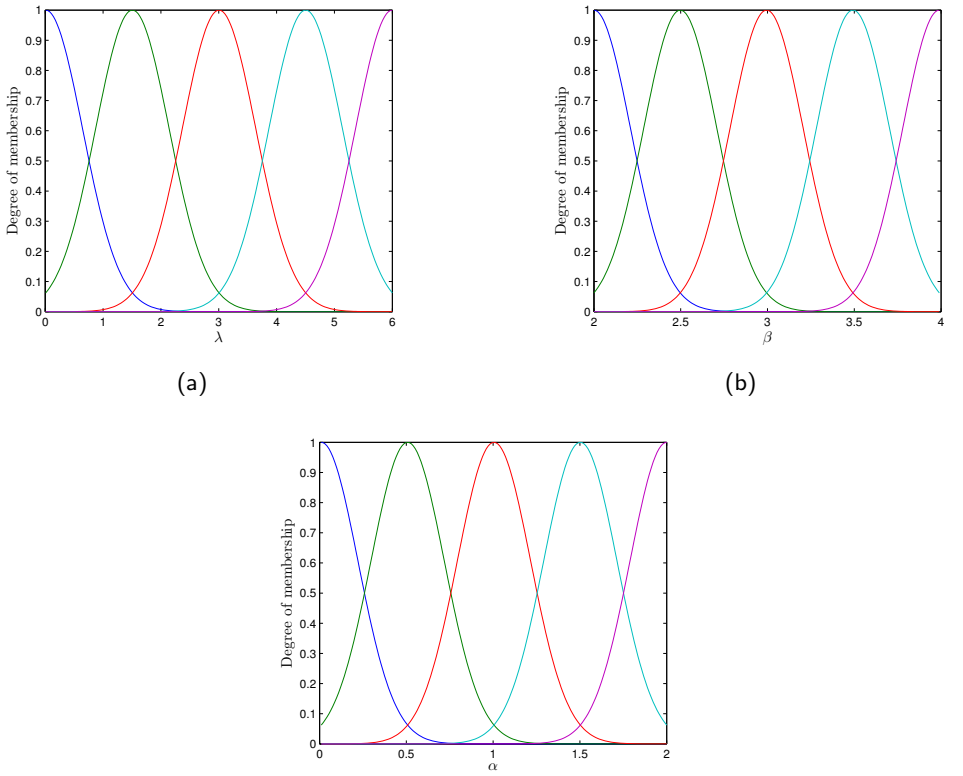

Figure 2: *ANFIS membership function for input variables* λ, β *and* α.

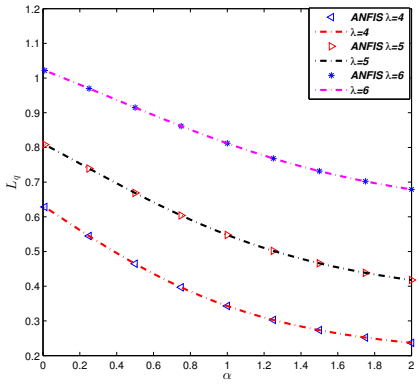

Figure 3: Impact on L_q of α by varying λ.

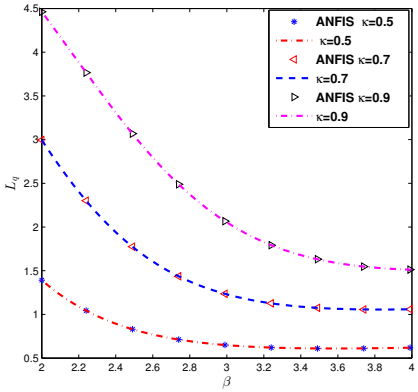

Figure 4: *Impact on* L_q *of* β *by varying* κ.

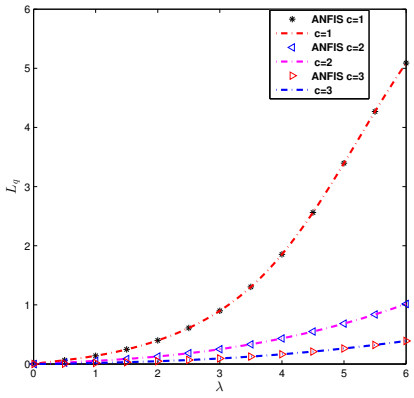

Figure 5: *Impact on L_q of λ by varying c.*

Table 1 illustrates that by increasing the working hiatus rate η_1 and hiatus rate η_2, the system tends to transition more quickly to the normal busy period, increasing the probability of the system being in a normal busy state and decreasing the probability of a working hiatus. Consequently, this results in a decrease in the mean number of customers in the system. Simultaneously, the probability of entering a complete hiatus state increases. This trend is observed as η_1 and η_2 are jointly increased, with η_2 being smaller than η_1.

Table 2 investigates the influence of non-balking probability κ on key performance metrics in a queueing model under fixed parameter settings. As κ increases, the expected waiting time W_s in the system tends to increase, indicating longer waits for customers. Concurrently, the expected number of customers in the queue L_q also rises, reflecting an accumulation of customers waiting for service. Interestingly, the expected balking rate B_r decreases as κ increases, suggesting that higher probabilities of customers entering the system lead to fewer customers choosing to leave without service. These insights underscore the critical role of non-balking probability in shaping customer experience and operational dynamics within the queueing system.

From Figures 3 to 5, we explore the nuanced impacts of key parameters on the average queue length L_q and compare these insights with results derived from the ANFIS model. Figure 3 demonstrates a reduction in L_q as the service rate during working hiatus periods increases, whereas increasing the arrival rates leads to an increase in L_q. Additionally, Figure 4 reveals a decrease in L_q with increasing service rates during normal busy periods, while a decrease in balking probability results in higher L_q. In Figure 5, we observe a rise in L_q with increased arrival rates, whereas increasing the number of servers leads to a decrease in L_q. Comparing these findings with ANFIS model predictions shows a close alignment between both approaches, indicating the reliability and accuracy of the ANFIS model in simulating queue dynamics and validating analytical results.

6.2. Optimal numerical cost

This section aims to determine the optimal service rates β and α that minimize the expected cost function. Due to the complex and non-linear nature of this optimization problem, analytical solutions are impractical. Therefore, we employ advanced nonlinear optimization techniques, specifically the Quasi-Newton method (QNM), to find the optimal values (β^*, α^*) under fixed parameter conditions within the cost model framework.

The optimization problem is formulated as follows:

$$\min_{\beta,\alpha} G(\beta,\alpha)$$
$$\text{s. t. } \begin{cases} \beta > \alpha \\ \alpha > 0. \end{cases}$$

In what follows, the optimal solutions are given by applying the QNM for various system parameters. To do this, we fix the parameters as: $C_{rb} = 80$, $C_h = 60$, $C_r = 50$, $C_\beta = 1$, $C_\alpha = 1$, $C_{s-f} = 1$, $C_b = 1$.

λ	β^*	α^*	$G^*(\beta^*, \alpha^*)$
7	5.2308	4.6689	149.6463
7.5	5.5459	5.0649	155.2061
8	5.8278	5.4562	160.7644
8.5	6.1411	5.8804	166.3211
9	6.4199	6.2789	171.8748

Table 3: *The optimal* (β^*, α^*) *and* $G^*(\beta^*, \alpha^*)$ *for various values of* λ *when* $\kappa = 0.4$, $\eta_1 = 0.6$, $\eta_2 = 1.1$, $c = 4$, $M = 10$, $q = 0.7$.

c	β^*	α^*	$G^*(\beta^*, \alpha^*)$
2	9.1250	8.0762	165.4715
3	6.2465	6.1956	151.7746
4	4.9263	4.2588	144.0853
5	4.0254	3.0062	139.3775
6	3.3214	2.1391	136.3725
7	2.7058	1,3942	133.9343

Table 4: *The optimal* (β^*, α^*) *and* $G^*(\beta^*, \alpha^*)$ *for various values of* c *when* $\kappa = 0.4$, $\eta_1 = 0.6$, $\eta_2 = 1.1$, $\lambda = 6.5$, $M = 10$, $q = 0.7$.

Table 3 illustrates the optimal values (β^*, α^*) and corresponding objective function $G^*(\beta^*, \alpha^*)$ for different arrival rates λ in a specific queueing model. With fixed parameters $\kappa = 0.4$, $\eta_1 = 0.6$, $\eta_2 = 1.1$, $c = 4$, $M = 10$, and $q = 0.7$, the results show that as λ increases from 7 to 9, both β^* and α^* also increase. This indicates that higher arrival rates necessitate larger values of β and α for optimal system performance. Furthermore, $G^*(\beta^*, \alpha^*)$ increases with λ, reflecting improved system performance metrics associated with higher arrival rates. The findings provide critical insights into optimizing queueing systems under varying demand conditions.

From Table 4, increasing the number of servers results in lower minimum costs and service rates, indicating that reducing the server count would be costly. The study optimizes (β, α) and evaluates G across various c values in a finite-capacity Markovian multi-server queue with balking and feedback. Optimal (β^*, α^*) values decrease as c increases, suggesting adjustments to maintain efficiency and customer satisfaction. G^* decreases with higher c, showing enhanced system performance under these conditions, offering strategic insights for managing queueing systems effectively.

7. Conclusion

In this investigation, we explored a finite-capacity Markovian multi-server queue with balking and feedback mechanisms, governed by a hybrid hiatus policy integrating both working and complete hiatus periods. We examined the effectiveness of this policy as servers transitioned from normal operations to a reduced service rate during working hiatus periods, followed by a complete hiatus when no customers were waiting. Once these hiatus concluded, servers resumed normal operations to attend to waiting customers. Using the Markov recursive approach, we analyzed the system's steady-state probabilities and queue metrics, including key measures such as the expected number of customers in the system and queue, expected waiting times, expected balking rate, and probabilities associated with different server states. To optimize decision variables within the system, we developed a cost function implemented through the Quasi-Newton method. Additionally, we validated our results using a soft computing technique, specifically an adaptive neuro-fuzzy inference system (ANFIS), to ensure the robustness and accuracy of our findings. These comprehensive analyses and methodologies provide valuable insights into enhancing operational efficiency and customer satisfaction in complex queuing environments.

The model discussed in the paper can be extended to address more complex scenarios, such as an unreliable multi-server queue with heterogeneous customers, which introduces additional layers of complexity to the problem. Furthermore, the exponential assumptions can be relaxed by incorporating phase-type distributions for service times. These extensions would broaden the applicability of the model, allowing for more realistic simulations and analyses in diverse queueing environments.

Acknowledgements

The authors would like to thank the annoymous reviewers for their valuable comments and suggestions to improve the quality of this paper.

References

[1] Baba, Y. (2005). Analysis of a GI/M/1 queue with multiple working vacations. Operations Research Letters, 33(2), 201-209. doi: 10.1016/j.orl.2004.05.006

[2] Banik, A. D., Gupta, U. C. and Pathak, S. S. (2007). On the $GI/M/1/N$ queue with multiple working vacations. analytic analysis and computation. Applied Mathematical Modelling, 31(9), 1701–1710. doi: 10.1016/j.apm.2006.05.010

[3] Bouchentouf, A. A., Boualem, M., Yahiaoui, L. and Ahmad, H. (2022). A multi-station unreliable machine model with working vacation policy and customers impatience. Quality Technology and Quantitative Management, 19(6), 766–796. doi: 10.1080/16843703.2022.2054088

[4] Bouchentouf, A. A., Cherfaoui, M. and Boualem, M. (2019). Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. Opsearch, 56(1), 300–323. doi: 10.1007/s12597-019-00357-4

[5] Bouchentouf, A. A., Cherfaoui, M. and Boualem, M. (2021). Analysis and performance evaluation of Markovian feedback multi-server queueing model with vacation and impatience. American Journal of Mathematical and Management Sciences, 40(3), 261–282. doi: 10.1080/01966324.2020.1842271

[6] Bouchentouf, A. A. and Guendouzi, A. (2020). The $M^X/M/c$ Bernoulli feedback queue with variant multiple working vacations and impatient customers: Performance and economic analysis. Arabian Journal of Mathematics, 9(2), 309–327. doi: 10.1007/s40065-019-0260-x

[7] Bouchentouf, A. A., Guendouzi, A. and Majid, S. (2020). On impatience in Markovian $M/M/1/N/DWV$ queue with vacation interruption. Croatian Operational Research Review, 11, 21–37. doi: 10.17535/crorr.2020.0003

[8] Bouchentouf, A. A. and Yahiaoui, L. (2017). On feedback queueing system with reneging and retention of reneged customers, multiple working vacations and Bernoulli schedule vacation interruption. Arabian Journal of Mathematics, 6(1), 1–11. doi: 10.1007/s40065-016-0161-1

[9] Divya, K. and Indhira, K. (2024). Performance analysis and ANFIS computing of an unreliable Markovian feedback queueing model under a hybrid vacation policy. Mathematics and Computers in Simulation, 218, 403-419. doi: 10.1016/j.matcom.2023.12.004

[10] Doshi, B. T. (1986). Single server queues with vacation-A survey. Queueing Systems, 1, 29–66. doi: 10.1007/BF01149327

[11] Ganie, S. M. and Manoharan, P. (2018). Impatient customers in an $M/M/c$ queue with single and multiple synchronous working vacations. Pakistan Journal of Statistics and Operation Research, 571-594.

[12] Indumathi, P. and Karthikeyan, K. (2024). ANFIS-Enhanced $M/M/2$ Queueing Model Investigation in Heterogeneous Server Systems with Catastrophe and Restoration. Contemporary Mathematics, 2482-2502. doi: 10.37256/cm.5220243977

[13] Jain, M. and Meena, R. K. (2018). Vacation model for Markov machine repair problem with two heterogeneous unreliable servers and threshold recovery. Journal of Industrial Engineering International, 14, 143-152. doi: 10.1007/s40092-017-0214-x

[14] Jain, M. and Upadhyaya, S. (2011). Synchronous working vacation policy for finite-buffer multiserver queueing system. Applied Mathematics and Computation, 217(24), 9916–9932. doi: 10.1016/j.amc.2011.04.008

[15] Jang, J. S. (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern, 23(3), 665–685. doi: 10.1109/21.256541

[16] Kadi, M., Bouchentouf, A. A. and Yahiaoui, L. (2020). On a multiserver queueing system with customers' impatience until the end of service under single and multiple vacation policies. Applications and Applied Mathematics: An International Journal (AAM), 15(2), 4. Retrieved from: digitalcommons.pvamu.edu

[17] Lin, C. H. and Ke, J. C. (2009). Multi-server system with single working vacation. Applied Mathematical Modelling, 33(7), 2967-2977. doi: 10.1016/j.apm.2008.10.006

[18] Prakati, P. (2024). $M/M/C$ queue with multiple working vacations and single working vacation under encouraged arrival with impatient customers. Reliability: Theory and Applications, 19(1 (77), 650-662.

[19] Selvaraju, N. and Goswami, C. (2013). Impatient customers in an $M/M/1$ queue with single and multiple working vacations. Computers and Industrial Engineering, 65(2), 207–215. doi: 10.1016/j.cie.2013.02.016

[20] Servi, L. D. and Finn, S. G. (2002). $M/M/1$ queues with working vacations ($M/M/1/WV$). Performance Evaluation, 50(1), 41–52. doi: 10.1016/S0166-5316(02)00057-3

[21] Sethi, R., Jain, M., Meena, R. K. and Garg, D. (2020). Cost optimization and ANFIS computing of an unreliable $M/M/1$ queueing system with customers' impatience under N-policy. International Journal of Applied and Computational Mathematics, 6, 1-14. doi: 10.1007/s40819-020-0802-0

[22] Tian, N. and Zhang, Z. G. (2006). Vacation queueing models: theory and applications(Vol 93). Springer Science and Business Media.

[23] Vadivukarasi, M. and Kalidass, K. (2022). A discussion on the optimality of bulk entry queue with differentiated hiatuses. Operations Research and Decisions, 32(2), 137-150. doi: 10.37190/ord220209

[24] Upadhyaya, S. and Kushwaha, C. (2020). Performance prediction and ANFIS computing for unreliable retrial queue with delayed repair under modified vacation policy. International Journal of Mathematics in Operational Research, 17(4), 437-466. doi: 10.1504/IJMOR.2020.110843

[25] Yahiaoui, L., Bouchentouf, A. A. and Kadi, M. (2019). Optimum cost analysis for an $Geo/Geo/c/N$ feedback queue under synchronous working vacations and impatient customers. Croatian Operational Research Review, 10, 211-226. doi: 10.17535/crorr.2019.0019