revista GEPROS

Editor Responsável: Prof. Dr. Hermes Moretti Ribeiro da Silva

Recebido: 14/02/2022
Aprovado: 14/02/2022

# M(t)/M(t)/m(t)/C(t) and non-stationary hypercube models

This article was invited for publication.

Regiane Máximo Siqueira[1], State University of São Paulo Julio de Mesquita Filho, Bauru, São Paulo, Brazil.

## RESUMO

**Objetivo -** Verificar o efeito da variação da taxa de chegada e atendimento durante o dia em detrimento da análise de equilíbrio do sistema.
**Desenho / metodologia / abordagem -** Os modelos M(t)/M(t)/m(t)/C(t) e o hipercubo não estacionário foram abordados considerando a variação do número de servidores ao longo do tempo.
**Resultados -** O estudo mostrou como a abordagem dinâmica é mais realista do que a abordagem de equilíbrio em sistemas onde a variação de parâmetros é um fator importante a ser considerado.
**Originalidade / valor -** A maioria dos estudos envolvendo sistemas de filas é baseada na análise em regime permanente da operação desses sistemas. No entanto, para determinados sistemas de filas, variações nas taxas de chegada de clientes, tempos de atendimento e outras condições de operação ocorrem em intervalos de tempo muito curtos, o que dificulta a análise efetiva do desempenho desses sistemas.
**Palavras-chave -** Teoria das filas, Modelo M(t)/M(t)/m(t)/C(t), Modelos não-estacionários.

## *ABSTRACT*

***Purpose*** *– Verify the effect of the variation of the arrival and service rate during the day in detriment of the equilibrium analysis of the system.*
***Design/methodology/approach*** *– The models M(t)/M(t)/m(t)/C(t) and the non-stationary hypercube were approached considering the change in the number of servers over time.*
***Findings*** *– The study showed how the dynamic approach is more realistic than the equilibrium approach in systems where the variation of parameters is an important factor to be considered.*
***Originality/value*** *– Most studies involving queue systems are based on steady-state analysis of the operation of these systems. However, for certain queuing systems, variations in customer arrival rates, service times and other operating conditions occur within very short time intervals, which makes it difficult to effectively analyze the performance of these systems.*
***Keywords*** *- Queuing theory, M(t)/M(t)/m(t)/C(t) Model, Non-stationary models.*

1. Av. Engenheiro Luiz Edmundo Carrijo Coube, 14-01. PO BOX 17033-360, Bauru, São Paulo, Brazil, regiane@feb.unesp.br, https://orcid.org/0000-0002-4695-2678;
SIQUEIRA, R.M M(t)/M(t)/m(t)/C(t) and non-stationary hypercube models. **GEPROS. Gestão da Produção, Operações e Sistemas**, v.17, nº 1, p. 112 - 146, 2022.

**DOI**: http://dx.doi.org/10.15675/gepros.v17i1.2864

## 1. INTRODUCTION

Systems such as call centers, hospitals, Emergency Attendance Systems (SAE's), computer servers, among others, have the demand heavily dependent on time (KIM; WHITT, 2014). In addition, these systems work with limited resources making their management non-trivial, they generally allocate resources in times of high demand and in periods of low demand they work with fewer resources in an attempt to adapt resources to demand variation over time. They are systems essentially characterized by uncertainties, mainly regarding demand, service time and location of servers, as the case may be. To design and analyze the configuration of these systems, managers and decision makers need to balance the level of service offered to users with the investments required to be able to offer a good level of service. They also need to define and analyze performance measures in order to be able to properly assess the system configuration, this can be done with the help of queuing theory analytical models and/or experimental simulation models.

In the equilibrium analysis, the arrival rate (λ) and the service rate (μ) are assumed to be constant and may in some cases vary with the state of the system (λn and μn, respectively). With this assumption, after some time of operation, the system passes the transient phase and goes into regime. Stationary analysis is a common approach to dealing with time-varying rates, the idea is to divide the time interval of interest (eg one day) into t small periods i = 1, . . . , T with constant arrival rates and numbers of servers within each period i (GREEN *et al.*, 2001; SOUZA *et. al.*, 2012). Queue models are solved for each period i by deriving stationary performance measures, this approach works accurately if consecutive intervals are statistically independent of each other.

Some works that address the issue of demand variation over time were found in the literature, such as Ingolfsson *et al*. (2002) who used a model to optimize the number of patrol cars in supermarkets for call centers considering demand variation throughout the day, λ(t). Kim and Whitt (2014) studied the behavior of the arrival rate of a call center and a hospital, and verified whether this could be represented by a heterogeneous poisson. Still others deal with the variation in the number of servers over time, as in Chen et. al. (2013a, 2013b) address a non-stationary queuing model, the M(t)/Ek/c(t) model was used to analyze a large maritime container terminal, whose objective was to reduce the number of trucks in the queue. The analysis proved to be very useful and can significantly increase the system's flexibility. The M(t)/M/s(t) model used in Ingolfsson *et. al*. (2005). Alanis (2013) presented a Markov Chain-based model for repositioning ambulances in emergency medical services systems. Stolletz (2008) analyzed the M(t)/M(t)/c(t) queue and evaluated the effect of dividing the day into several periods with different arrival rates comparing it with the stationary approach. Ingolfsson *et al.* (2014) compare and discuss different methods to approximate the M(t)/M/c(t) queuing system. In another approach, Schmid (2012) and Maxwell *et al.* (2010) presented a dynamic programming model for repositioning servers throughout the day, Maxwell *et al*. (2010) still uses in their dynamic programming model a Markov chain to consider the problem of repositioning ambulances. vary with the state of the system (λn and μn, respectively). With this assumption, after some time of operation, the system passes the transient phase and goes into regime. Stationary analysis is a common approach to dealing with time-varying rates, the idea is to divide the time interval of interest (eg one day) into t small periods i = 1, . . . , T with constant arrival rates and numbers of servers within each period i (GREEN *et al*., 2001; SOUZA *et. al*., 2012). Queue models are solved for each period i by deriving stationary performance measures, this approach works accurately if consecutive intervals are statistically independent of each other.

In this sense, it is important to verify the effectiveness and applicability of the various queuing models in the analysis of systems in which the average arrival and service rates depend on time and/or systems that change their dynamics over time (eg with an increase or decrease number of servers). In these cases, the equilibrium hypothesis does not allow a more effective analysis of the performance of these systems, which can lead to oversizing or undersizing the number of servers needed for a certain level of service, measured during a chosen time interval.

Applications in this sense can be seen, for example, in Broyles *et. al*. (2010) who study the problem of predicting hospital inputs as a function of the rate of admission and service that are not stationary in a Markov chain. Chen *et. al*. (2013) use the M(t)/Ek/c(t) model, considering λ(t) and μ(t), in a marine terminal gate integrated with a Genetic Algorithm in order to optimize the queue of trucks in the terminal. Examples in other areas can also be found in the literature, such as the application in Call Centers (JOUINI *et. al*., 2009).

The hypercube model is a spatially distributed queuing model, originally proposed by Larson (1974) and extended by several authors (SWERSEY, 1994; GALVÃO; MORABITO, 2008; BOFFEY *et al*., 2007; IANNONI *et al*. (2007, 2009) ; MORABITO *et al*. (2008); SIMPSON; HANCOCK (2009); SOUZA *et al*. (2013, 2015)), which has been used to analyze and plan various emergency care systems. In emergency systems it is common to change the number of servers throughout the day, in addition the demand can change during the day depending on location and time. It is important to emphasize that all studies found in the literature were carried out considering the equilibrium hypothesis.

This work aims to compare the equilibrium analysis where the arrival rate, service rate and number of servers is constant with the out-of-balance approach, in which the arrival rate, service and number of servers varies with time in the fundamental models of Markovian queues (user-to-server approach) and the hypercube model (server-to-user approach). The idea is to check the effect of these changes on some performance measures. The article is organized as follows, Section 2 presents the fundamental non-stationary Markovian models and the non-stationary hypercube model. In Section 3 we present the Runge-Kutta method, in Section 4 we present the results and discussions of the illustrative examples and Section 5 presents the Conclusion.

## 2. NON-STATIONARY MODEL

### 2.1 The queue *M(t)/M/m*

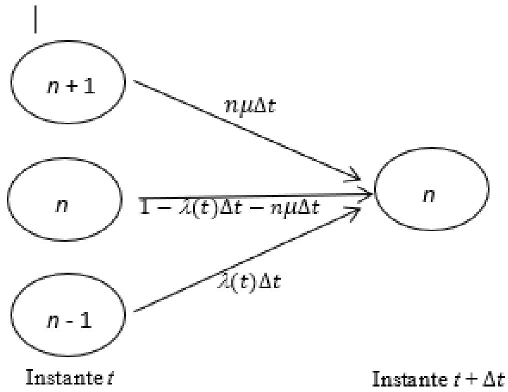
The queue *M(t)/M/m* has the arrival rate varying over time according to a heterogeneous Poisson and the service times exponentially distributed with averages $E(t, X) = \lambda(t)$ and $E(S) = \frac{1}{\mu}$, respectively. Users wait in a single queue with unlimited capacity, being served by the FCFS (First Come First Served) discipline and the utilization factor $\rho(t) = \frac{\lambda(t)}{m\mu}$ corresponds to the average usage of the system at the time *t* ($\rho(t) < 1$). This model considers that the user entry rate does not vary with the state of the system, while the service rate changes depending on the state in which the system is, as follows:

$$\lambda(t)_n = \lambda(t) \text{ , for } n = 0, 1, 2\dots, \qquad t \geq 0 \qquad (1)$$

$$\mu_n = \begin{cases} n\mu, & for\ n = 1,2, \dots m-1 \\ m\mu, & for\ n = m, m+1, \dots \end{cases} \qquad (2)$$

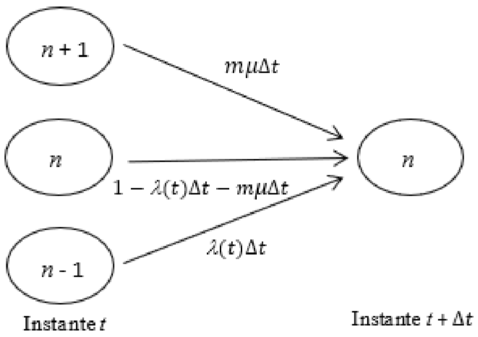
To simplify the analysis, in Figures 1 and 2 we do not consider the arrival of more than one user in a small period of time (with the probability of arrival $o(\Delta t)$). Figure 1 illustrates the possible state transitions for the model *M(t)/M/m*, when the number of users is less than the number of servers ($n < m$). In these states, any user who arrives at the system is promptly attended to. The state $n + 1$ (instant $t$) switches to state n with a termination of service during $\Delta t$, with probability $n\mu\Delta t$. Similarly, the state $n - 1$ (instant $t$) switches to state n upon the arrival of a call during $\Delta t$, with probability $\lambda(t)\Delta t$ (this transition does not exist if $n = 0$). Also, if at time t the system is in state n and there is no arrival or termination of service during $\Delta t$, the system remains in state n, with probability $1 - \lambda(t)\Delta t - n\mu\Delta t$.

**Figure 1 -** State transitions from the M(t)/M/m model from the n state, when n < m.




Still considering the same analysis, when $n \geq m$ (Figure 2), all m servers are busy and if a user arrives at the system, he waits in a simple queue, operating with FCFS discipline, as mentioned at the beginning of this Section. Thus, state n + 1 (at time t) changes to state n with a termination of service during $\Delta t$, with probability $m\mu\Delta t$. If, at time t, the system is in state n and there is no arrival or termination of service during $\Delta t$, the system remains in state n, with probability $1 - \lambda(t)\Delta t - n\mu\Delta t$. State n - 1 (instant t) changes to state n upon the arrival of a call in the system with probability $\lambda(t)\Delta t$ (the same way when $n < m$).

**Figure 2 -** State transitions from the M(t)/M/m model from the n state, when n ≥ m.




Thus, the probability that the system is in state n at the moment $t + \Delta t$ is defined by:

$$\begin{aligned}
&\boldsymbol{P_0(t+\Delta t) = P_1(t)\mu\Delta t + P_0(t)[1-\lambda(t)\Delta t] + o(t)} && \boldsymbol{n=0} \quad 3\\
&P_n(t+\Delta t) = P_{n+1}(t)(n+1)\mu\Delta t + P_n(t)[1-(\lambda(t)+n\mu)\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) && n = 1,2,\dots,m-1\\
&P_n(t+\Delta t) = P_{n+1(t)} m\mu\Delta t + P_n(t)[1-(\lambda(t)+m\mu)\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) && n = m, m+1, \dots
\end{aligned}$$

The previous equations can be rewritten as follows:

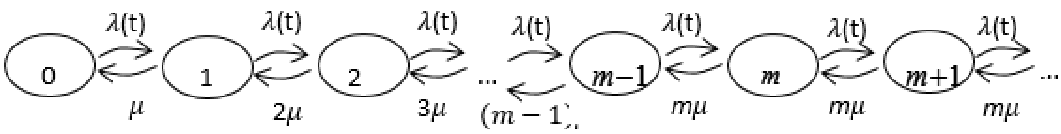
$$\begin{aligned}
&\frac{P_0(t+\Delta t)-P_0(t)}{\Delta t} = P_1(t)\mu + \lambda(t)P_0(t) + o(t) && n=0\\
&\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t} = -(\lambda(t)+n\mu)P_n(t) + (n+1)\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) + o(t) && n = 1,2,\dots,m-1\\
&\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t} = -(\lambda(t)+m\mu)P_n(t) + m\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) + o(t) && n = m, m+1, \dots
\end{aligned}$$

When $\Delta t \to 0$, we have to:

$$\begin{aligned}
&\boldsymbol{\frac{dP_0(t)}{dt} = P_1(t)\mu + \lambda(t)P_0(t)} && \boldsymbol{n=0} \quad 4\\
&\frac{dP_n(t)}{dt} = -(\lambda(t)+n\mu)P_n(t) + (n+1)\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) && n = 1,2,\dots,m-1\\
&\frac{dP_n(t)}{dt} = -(\lambda(t)+m\mu)P_n(t) + m\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) && n = m, m+1, \dots
\end{aligned}$$

Differential equation system 4 describes the system over time, allowing the analysis of the transient behavior of the queuing system. Figure 3 shows the transition of states in an M(t)/M/m system to $n = 0,1,2, \dots$.

**Figure 3 -** Model State Transition *M*(*t*)/*M*/*m*.




From the solution of the system of differential equations, it is possible to calculate the performance measures that describe this system over time ($t \geq 0$). The number of users in the queue ($L_q(t)$) at time t is obtained from the probabilities of the states $P_n(t)$. The other performance measures, the number of users on the system ($L(t)$), the length of stay in the system ($W(t)$) and the time spent in line $W_q(t)$) all performance measures describing the system over time ($t \geq 0$) are obtained using Little's formula by Equations 5, 6, 7 e 8.

$$\boldsymbol{L_q(t) = \sum_{n=m}^{\infty}(n-m)P_n(t)}, \qquad t \geq 0 \qquad 5$$
$$\boldsymbol{L(t) = L_s(t) + L_q(t) = \frac{\lambda(t)}{\mu} + L_q(t)}, \quad t \geq 0 \qquad 6$$
$$\boldsymbol{W(t) = W_s + W_q(t) = \frac{1}{\mu} + W_q(t)}, \quad t \geq 0 \qquad 7$$
$$\boldsymbol{W_q(t) = \frac{L_q(t)}{\lambda(t)}}, \qquad t \geq 0 \qquad 8$$

On what:

$L_s(t) = \frac{\lambda(t)}{\mu}$ is the average number of users in service in $t \geq 0$.
$\overline{W}_s = \frac{1}{\mu}$ is the average service time of the system.

It is usual to calculate the average performance measures of a queuing system. Therefore, the average number of users in the queue over a period of time $(t_1, t_2)$,$\overline{L}_q(t_1,t_2)$, is obtained from Equation 9. The other performance measures, $\overline{L}(t_1, t_2)$ , $\overline{W}(t_1, t_2)$ e $\overline{W}_q(t_1, t_2)$, are mean values obtained in the range $(t_1, t_2)$, from Equations 10, 11, 12 and 13, respectively.

$$\boldsymbol{\overline{L}_q(t_1, t_2) = \frac{\int_{t_1}^{t_2}\sum_{n=m}^{\infty}(n-m)P_n(t)}{t_2 - t_1}} \qquad 9$$
$$\boldsymbol{\overline{L}(t_1, t_2) = \overline{L}_s + \overline{L}_q = \frac{\lambda}{\mu} + \overline{L}_q}, \qquad 10$$
$$\boldsymbol{\overline{W}(t_1, t_2) = \overline{W}_s + \overline{W}_q = \frac{1}{\mu} + \overline{W}_q} \qquad 11$$
$$\boldsymbol{\overline{W}_q = \frac{\overline{L}_q(t_1, t_2)}{\lambda}} \qquad 12$$

## 2.2 The queue *M(t)/M(t)/m*

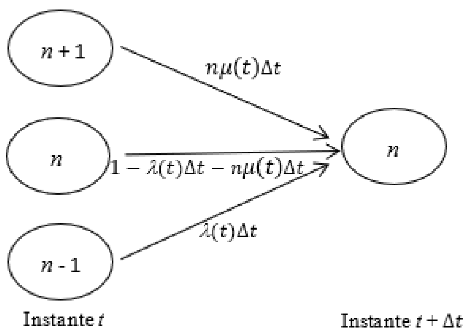
In the M(t)/M(t)/m queue, the arrival rate and service times vary over time according to a heterogeneous Poisson and have averages $E(t,X) = \lambda(t)$ and $E(t,S) = \frac{1}{\mu(t)}$, respectively. Users wait in a single queue with unlimited capacity, being served by the FCFS (First Come First Served) discipline and the utilization factor $\rho(t) = \frac{\lambda(t)}{m\mu(t)}$ corresponds to the use of the system at the moment *t* ($t \geq 0,\ \rho(t) < 1$). This model considers that the user entry rate does not vary with the state of the system but rather as a function of *t* ($t \geq 0$), while service fees change depending on the state the system is in and depending on *t* ($t \geq 0$), this way:

$$\lambda_n(t) = \lambda(t)\text{, for n} = 0, 1, 2..., \qquad t \geq 0 \qquad (13)$$

$$\mu_n(t) = \begin{cases} n u(t), & \text{for } n = 1,2,\dots,m-1 \\ m\mu(t), & \text{for } n = m, m+1, \dots \end{cases}, \qquad t \geq 0 \qquad (14)$$

To simplify the analysis of Figures 4 and 5, the arrival of more than one user in a small period of time will not be considered, which happens with probability $o(\Delta t)$. Figure 4 illustrates the possible state transitions of the M(t)/M(t)/m model, when the number of users is smaller than the number of servers (n < m). In these states, any user who arrives at the system is promptly attended to. State n + 1 (at time t) changes to state n with a termination of service during $\Delta t$, with probability $n\mu(t)\Delta t$. Similarly, state n – 1 (at time t) changes to state n upon the arrival of a call during $\Delta t$, with probability $\lambda(t)\Delta t$ (this transition does not exist if $n = 0$). Also, if at time t the system is in state n and there is no arrival or termination of service during $\Delta t$, the system remains in state n, with probability $1 - \lambda(t)\Delta t - n\mu(t)\Delta t$.

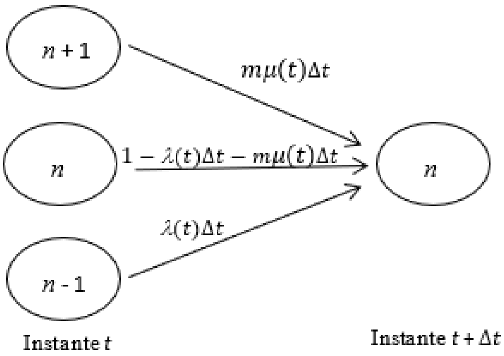
**Figure 4 -** State transitions from the M(t)/M(t)/m model from the n state, when *n* < *m*.




When $n \geq m$ (Figure 5), all m servers are busy and if a user arrives at the system, he waits in a simple queue, operating with FCFS discipline, as mentioned at the beginning of this

Section. Thus, state n + 1 (at time t) changes to state n with a termination of service during $\Delta t$, with probability $m\mu(t)\Delta t$. If, at time t, the system is in state n and there is no arrival or termination of service during $\Delta t$, the system remains in state n, with probability $1 - \lambda(t)\Delta t - n\mu(t)\Delta t$. State n - 1 (at time t) changes to state n upon the arrival of a call in the system with probability $\lambda(t)\Delta t$ (in same way when n < m).

**Figure 5 -** State transitions from the M(t)/M(t)/m model from the n state, when $n \geq m$.




Thus, the probability that the system is in state n at the moment $t + \Delta t$ is defined by:

$$
\begin{aligned}
&P_0(t+\Delta t) = P_1(t)\mu(t)\Delta t + P_0(t)[1-\lambda(t)\Delta t] + o(t) \qquad n = 0 \\
&P_n(t+\Delta t) = P_{n+1}(t)(n+1)\mu(t)\Delta t + P_n(t)[1-(\lambda(t)+n\mu(t))\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) \qquad n = 1,2,\ldots,m-1 \\
&P_n(t+\Delta t) = P_{n+1(t)}m\mu(t)\Delta t + P_n(t)[1-(\lambda(t)+m\mu(t))\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) \qquad n = m, m+1, \ldots
\end{aligned}
\tag{15}
$$

The previous equations can be rewritten as follows:

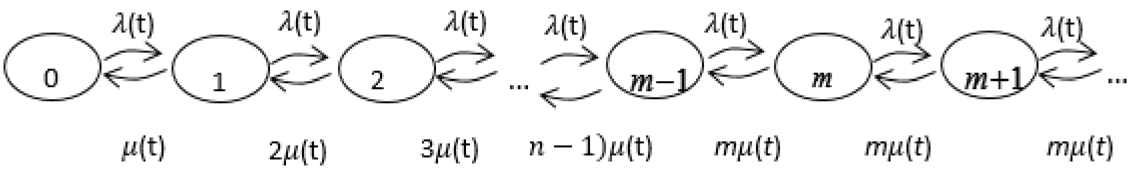
$$
\begin{aligned}
&P_0(t+\Delta t) = P_1(t)\mu(t)\Delta t + P_0(t)[1-\lambda(t)\Delta t] + o(t) \qquad n = 0 \\
&P_n(t+\Delta t) = P_{n+1}(t)(n+1)\mu(t)\Delta t + P_n(t)[1-(\lambda(t)+n\mu(t))\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) \qquad n = 1,2,\ldots,m-1 \\
&P_n(t+\Delta t) = P_{n+1(t)}m\mu(t)\Delta t + P_n(t)[1-(\lambda(t)+m\mu(t))\Delta t] + P_{n-1}(t)\lambda(t)\Delta t + o(t) \qquad n = m, m+1, \ldots
\end{aligned}
\tag{15}
$$

Doing $\Delta t \to 0$, we have to:

$$\frac{dP_0(t)}{dt} = P_1(t)\mu(t) + \lambda(t)P_0(t) \qquad n = 0$$
$$\frac{dP_n(t)}{dt} = -(\lambda(t) + n\mu(t))P_n(t) + (n+1)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) \quad n = 1,2,\dots,m-1$$
$$\frac{dP_n(t)}{dt} = -(\lambda(t) + m\mu(t))P_n(t) + m\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) \qquad n = m, m+1,\dots \qquad 16$$

The differential equations above describe the system over time, allowing the analysis of the transient behavior of the queuing system. Figure 6 shows the transition of states in a non-stationary M(t)/M(t)/m system, for $n = 0,1,2,\dots$.

**Figure 6 -** Model State Transition *M*(*t*)/*M*(*t*)/*m*.




The number of users in the queue (Lq(t)) at time t and the queue time Wq(t) are obtained from Equations 5 and 8, respectively. The other performance measures, the number of users in the system (L(t)) and the time spent in the system (W(t)) and o), the equations are now:

$$L(t) = L_s(t) + L_q(t) = \frac{\lambda(t)}{\mu(t)} + L_q(t), \;\; t \geq 0 \qquad 17$$
$$W(t) = S + W_q(t) = \frac{1}{\mu(t)} + W_q(t), \;\; t \geq 0 \qquad 18$$

The average number of users in the queue over a period of time $(t_1, t_2)$,$\bar{L}_q(t_1, t_2)$ and the average time spent in line are obtained by equations 9 and 12. The other performance measures, $\bar{L}(t_1, t_2)$ e $\bar{W}(t_1, t_2)$ are obtained from Equations 19 and 20.

$$\bar{L}(t_1, t_2) = L_s + L_q = \frac{\lambda}{\mu(t)} + L_q, \qquad 19$$
$$\bar{W}(t_1, t_2) = S + W_q = \frac{1}{\mu(t)} + W_q \qquad 20$$

### 2.3 The queue *M*(*t*)/*M*(*t*)/*m*(*t*)

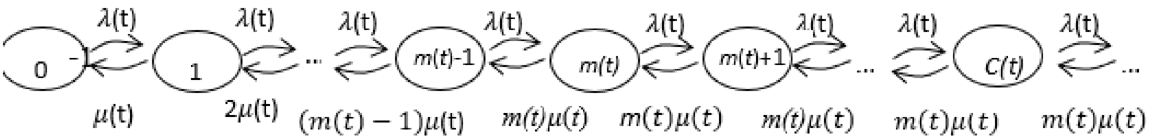
We can observe that, in all the models presented, the rate of change of the probability of a state n, in relation to time, is the sum of the probabilities of reaching state n minus the probability of the system leaving state n in an interval of time $\Delta t \rightarrow 0$.

The M(t)/M(t)/m(t) model considers that the arrival rate of users in the system arrives according to a heterogeneous Poisson distribution $(\lambda(t))$, the distribution of service time is exponential with service rate dependent on time $(\mu(t))$, the system has m(t) servers in the range $t_i < t < t_s$ (the number of servers may vary over time). The state transition diagram that represents the system can be seen in Figure 7. The transition rates that represent this system are:

**Figure 7** - Model State Transition *M*(*t*)/*M*(*t*)/*m*(*t*).

$$\lambda_n(t) = \begin{cases} \lambda(t), & \text{para } n = 0,1,2,\dots,m(t)-1 \\ 0, & \text{para } n = m(t), m(t)+1,\dots,C(t) \end{cases} \quad t \geq 0 \qquad 21$$

$$\mu_n(t) = \begin{cases} n\mu(t), & \text{para } n = 1,2,\dots,m-1 \\ m(t)\mu(t), & \text{para } n = m(t), m(t)+1,\dots,C(t) \end{cases} \quad t \geq 0 \qquad 22$$




The system of equations that represents this system is:

$$\begin{aligned} \frac{dP_0(t)}{dt} &= P_1(t)\mu(t) + \lambda(t)P_0(t) & n &= 0 \\ \frac{dP_n(t)}{dt} &= -(\lambda(t) + n\mu(t))P_n(t) + (n+1)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) & n &= 1,2,\dots,m(t)-1 \\ \frac{dP_n(t)}{dt} &= -(\lambda(t) + m(t)\mu(t))P_n(t) + m(t)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) & n &= m(t), m(t)+1,\dots \end{aligned} \qquad 23$$

Note that if the number of servers changes over time, the number of equations describing the system also changes. Furthermore, we consider that the number of users in the queue is always the same regardless of changing the number of servers. The average number of users in the queue at each instant of time t (t≥0) is given by Equation 24. The other performance measures calculated at each instant of time t (t≥0) can be calculated by equations 17 and 18. The average waiting times in the queue over a period of time, $(t_1, t_2)$, can be obtained by Equation 25, The other average performance measures over a time interval $(t_1, t_2)$ can be calculated by equations 19 and 20.

$$L_q(t) = \sum_{n=m}^{\infty}(n - m(t))P_n(t), \qquad t \geq 0 \qquad 24$$

$$\bar{L}_q(t_1, t_2) = \frac{\int_{t_1}^{t_2} \sum_{n=m(t)}^{\infty}(n-m(t))P_n(t)}{t_2 - t_1} \qquad 25$$

## 2.4 Models with limited capacity

The M(t)/M/m/C and M(t)/M(t)/m/C (m≤C) models differ from the M(t)/M/m and M(t)/M(t) models )/m, respectively, only by limiting C on the number of users present in the system (in queue and in service), resulting in a maximum queue size (C - m). The M(t)/M(t)/m(t)/C(t) (m(t) ≤ C(t)) model considers that the arrival rate of users in the system arrives according to a Poisson distribution heterogeneous $(\lambda(t))$, the distribution of service time is exponential with service rate dependent on time $(\mu(t))$, the system has m(t) servers in the range $t_i < t < t_s$ (the number of servers may vary over time), the system has limited capacity $(C(t) - m(t))$ which can also vary in time.

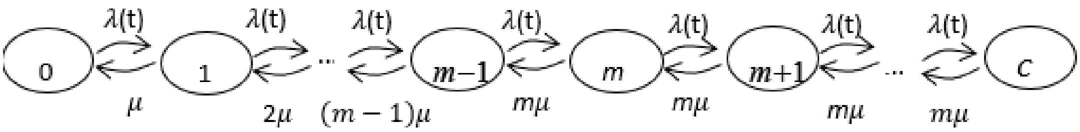
### 2.4.1 The queue *M(t)/M/m/C*

The state transition diagram that represents the M(t)/M/m/C queue system can be seen in Figure 8. The transition rates that represent this system are:

**Figure 8** - Model State Transition *M(t)/M/m/C*.

$$\lambda_n(t) = \begin{cases} \lambda(t), & \text{para} \quad n = 0,1,2,\ldots,m-1 \\ 0, & \text{para} \quad n = m, m+1, \ldots, C \end{cases} \qquad t \geq 0 \qquad 26$$

$$\mu_n = \begin{cases} n\mu, & \text{para } n = 0,1,2,\ldots,m-1 \\ m\mu, & \text{para } n = m, m+1,\ldots,C \end{cases} \qquad 27$$



The system of equations for this system is:

$$\frac{dP_0(t)}{dt} = P_1(t)\mu + \lambda(t)P_0(t) \qquad n = 0 \qquad 28$$
$$\frac{dP_n(t)}{dt} = -(\lambda(t) + n\mu)P_n(t) + (n+1)\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) \qquad n = 1,2,\ldots,m-1$$
$$\frac{dP_n(t)}{dt} = -(\lambda(t) + m\mu)P_n(t) + m\mu P_{n+1}(t) + \lambda(t)P_{n-1}(t) \qquad n = m, m+1, \ldots, C-1$$
$$\frac{dP_C(t)}{dt} = -m\mu P_C(t) + \lambda(t)P_{C-1}(t) \qquad n = C$$

For models that consider limited capacity, the probability of loss is an important performance measure to be considered, and represents the probability that the system has C users $(P_{perda}(t) = P_C(t))$. The user input fee $(\bar{\lambda}(t) < \lambda(t))$ used in Little's formula is defined by:

$$\bar{\lambda}(t) = \lambda(t)\sum_{n=0}^{C-1} P_n(t) + 0P_C(t) = \lambda(t)(1 - P_C(t)) \qquad 29$$

The utilization factor, $\rho(t)$, corresponds to the utilization of the system at time t, which is given by $\rho(t) = \frac{\bar{\lambda}(t)}{m\mu}$. The number of users in the queue ($L_q(t)$) at time t is obtained from Equation 30. The other performance measures, the number of users in the system ($L(t)$), the time spent in the system (W(t)) and the time spent in queue $W_q(t)$) all performance measures describing the system over time (t ≥ 0) are obtained using Equations 31, 32 and 33, respectively.

$$L_q(t) = \sum_{n=m}^{C}(n-m)P_n(t), \qquad t \geq 0 \qquad 30$$
$$L(t) = L_s(t) + L_q(t) = \frac{\bar{\lambda}(t)}{\mu} + L_q(t), \quad t \geq 0 \qquad 31$$
$$W(t) = W_s + W_q(t) = \frac{1}{\mu} + W_q(t), \quad t \geq 0 \qquad 32$$
$$W_q(t) = \frac{L_q(t)}{\bar{\lambda}(t)} \qquad t \geq 0 \qquad 33$$

The average number of users in the queue over a period of time $(t_1, t_2)$,$\bar{L}_q(t_1, t_2)$, is obtained from Equation 34. The other performance measures, $\bar{L}(t_1, t_2)$ , $\bar{W}(t_1, t_2)$ e $\bar{L}_q(t_1, t_2)$, are mean values obtained in the range $(t_1, t_2)$, obtained from Equations 35, 36 and 37, respectively.

$$\bar{L}_q(t_1,t_2) = \frac{\int_{t_1}^{t_2} \sum_{n=m}^{C} (n-m) P_n(t)}{t_2 - t_1} \qquad 34$$

$$\bar{L}(t_1,t_2) = \bar{L}_s + \bar{L}_q = \frac{\bar{\lambda}}{\mu} + \bar{L}_q \qquad 35$$

$$\bar{W}(t_1,t_2) = \bar{W}_s + \bar{W}_q = \frac{1}{\mu} + \bar{W}_q \qquad 36$$

$$\bar{W}_q = \frac{\bar{L}_q(t_1,t_2)}{\bar{\lambda}} \qquad 37$$

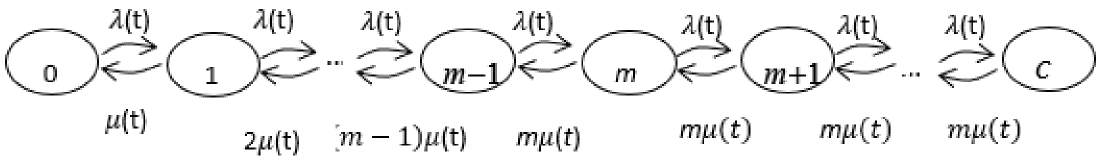
### 2.4.2 The queue *M(t)/M(t)/m/C*

The state transition diagram that represents the system in equilibrium can be seen in Figure 9. The transition rates that represent this system are:

**Figure 9** - Model State Transition *M(t)/M(t)/m/C*.

$$\lambda_n(t) = \begin{cases} \lambda(t), & \text{para } \; n = 0,1,2,\ldots,m-1 \\ 0, & \text{para } \; n = m, m+1, \ldots, C \end{cases} \quad t \geq 0 \qquad 38$$

$$\mu_n(t) = \begin{cases} n\mu(t), & \text{para } n = 1,2,\ldots,m-1 \\ m\mu(t), & \text{para } n = m, m+1, \ldots, C \end{cases} \quad t \geq 0 \qquad 39$$




The system of equations that represents this system is

$$\frac{dP_0(t)}{dt} = P_1(t)\mu(t) + \lambda(t)P_0(t) \qquad n = 0 \qquad 40$$

$$\frac{dP_n(t)}{dt} = -\big(\lambda(t) + n\mu(t)\big)P_n(t) + (n+1)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) \qquad n = 1,2,\ldots,m-1$$

$$\frac{dP_n(t)}{dt} = -\big(\lambda(t) + m\mu(t)\big)P_n(t) + m\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) \qquad n = m, m+1, \ldots, C-1$$

$$\frac{dP_C(t)}{dt} = -m\mu(t)P_C(t) + \lambda(t)P_{C-1}(t) \qquad n = C$$

The number of users in the queue (Lq(t)) and the time spent in queue Wq(t)) at time t are obtained from Equations 30 and 33, respectively. The number of users in the system (L(t)), the time spent in the system (W(t)), over time ($t \geq 0$), are obtained using Equations 41 and 42, respectively.

$$L(t) = L_s(t) + L_q(t) = \frac{\bar{\lambda}(t)}{\mu(t)} + L_q(t), \quad t \geq 0 \qquad 41$$

$$W(t) = W_s + W_q(t) = \frac{1}{\mu(t)} + W_q(t), \quad t \geq 0 \qquad 42$$

The average number of users in the queue over a period of time $(t_1, t_2)$,$\bar{L}_q(t_1,t_2)$ and the average time spent in the queue over a period of time. $(t_1, t_2)$, $\bar{W}_q(t_1,t_2)$, are obtained from Equations 34 and 37. The other performance measures, $\bar{L}(t_1,t_2)$ , $\bar{W}(t_1,t_2)$ and, they are mean values obtained in the interval (t_1,t_2 ), obtained from Equations 43 and 44, respectively.

$$\bar{L}(t_1,t_2) = \bar{L}_s + \bar{L}_q = \frac{\bar{\lambda}}{\mu(t)} + \bar{L}_q \qquad 43$$

$$\bar{W}(t_1,t_2) = \bar{W}_s + \bar{W}_q = \frac{1}{\mu(t)} + \bar{W}_q \qquad 44$$

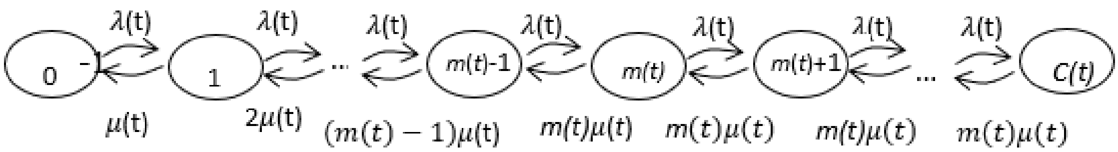
### 2.4.3 The queue *M*(*t*)/*M*(*t*)/*m*(*t*)/*C*(*t*)

The state transition diagram that represents the system can be seen in Figure 10. The transition rates that represent this system are:

**Figure 10** - Model State Transition *M*(*t*)/*M*(*t*)/*m*(*t*)/*C*(*t*).

$$\lambda_n(t) = \begin{cases} \lambda(t), & \text{para } n = 0,1,2,\ldots,m(t)-1 \\ 0, & \text{para } n = m(t), m(t)+1,\ldots,C(t) \end{cases} \qquad 45$$

$$\mu_n(t) = \begin{cases} nu(t), & \text{para } n = 1,2,\ldots,m-1 \\ m(t)\mu(t), & \text{para } n = m(t), m(t)+1,\ldots,C(t) \end{cases} \qquad 46$$




The system of equations that represents this system is:

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= P_1(t)\mu(t) + \lambda(t)P_0(t) && n = 0 \qquad 47 \\
\frac{dP_n(t)}{dt} &= -(\lambda(t) + n\mu(t))P_n(t) + (n+1)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) && n = 1,2,\ldots,m(t)-1 \\
\frac{dP_n(t)}{dt} &= -(\lambda(t) + m(t)\mu(t))P_n(t) + m(t)\mu(t)P_{n+1}(t) + \lambda(t)P_{n-1}(t) && n = m(t), m(t)+1,\ldots,C(t)-1 \\
\frac{dP_n(t)}{dt} &= -m(t)\mu(t)P_n(t) + \lambda(t)P_{n-1}(t) && n = C(t)
\end{aligned}$$

The number of users in the queue (Lq(t)) at time t is obtained from Equation 30. The other performance measures, the number of users in the system (L(t)), the time spent in the system (W (t)) and the queuing time (Wq(t)) all performance measures describing the system over time (t ≥ 0) are obtained using Equations 41, 42 and 33, respectively.

$$L_q(t) = \sum_{n=m(t)}^{C(t)} (n - m(t))P_n(t), \qquad t \geq 0 \qquad 48$$

The average number of users in the queue over a period of time $(t_1, t_2)$,$\bar{L}_q(t_1, t_2)$, is obtained from Equation 49. The other performance measures, $\bar{L}(t_1, t_2)$ , $\bar{W}(t_1, t_2)$ e $\bar{W}_q(t_1, t_2)$, are mean values obtained in the interval (t_1,t_2 ), obtained from Equations 43, 44 and 37, respectively.

$$\bar{L}_q(t_1, t_2) = \frac{\int_{t_1}^{t_2} \sum_{n=m(t)}^{C(t)} (n - m(t))P_n(t)}{t_2 - t_1} \qquad 49$$

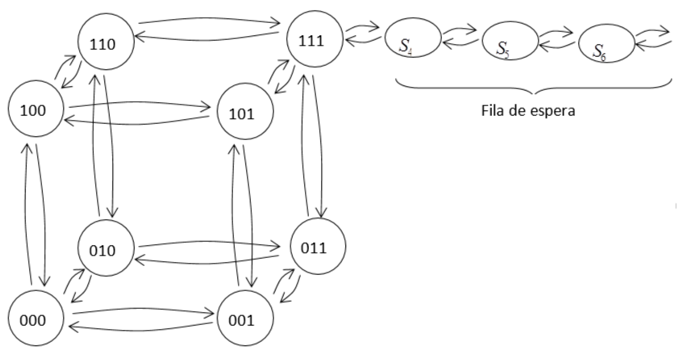
## 2.5 Non-stationary hypercube model

The hypercube model (LARSON, 2007) is a descriptive model used as a tool for the analysis and planning of urban emergency systems. In addition to considering uncertainties regarding the origin of calls, service times and availability of servers, the model addresses geographic and temporal complexities in the region. It can analyze both coordinated and centralized systems, when the user calls a central requesting some type of service and a server moves to the client. Originally, the solution of the model is given by the construction of the system's equilibrium equations, which are defined assuming that the system reaches equilibrium and that the arrival and service rates do not vary with time.

Basically, the idea is to expand the state space of an M/M/m queuing system to represent each server individually, which may include more complicated dispatch policies. The solution of the model is given starting from the construction of the set of equilibrium equations for the system. The results are based on the values of system state probabilities, enabling the calculation of performance measures, such as: server workloads, average response time of the system or each server, frequency of service of each server in each region, among others. Some of these hypotheses can be changed, such as, for example, multiple dispatch and partial backup, as in Chelst and Barlach (1981), Mendonça and Morabito (2000), Iannoni (2005) and Iannoni *et. al.* (2008a, 2008b).

The hypercube model is based on dividing the region served by the system into geographic atoms (demand regions). Each atom is considered a punctual source of calls and independent from the others and the service to each atom is performed by servers that are distributed in the region. The location of servers must be known or estimated using geometric probability. If one server is busy, other servers can answer the call, even if it is outside their preferred area, with cooperation between the servers prevailing.

**Figure 11** – System states with three servers.




Server availability is represented through the server state space. A particular state of the queueless system is given by the list of servers that are free or busy. Let {000}, {001}, {010},{111} be the $2^3$ = 8 possible states of the system, in which the 0's and 1's indicate whether servers are free or busy, respectively. For example, state {011} represents the state where server 1 is free and servers 2 and 3 are busy ({011} describes the state of servers from left to right). Figure 11 shows the state space for systems with m = 3 servers.

The model addresses both systems in which queuing is not allowed, and those in which, when all servers are busy, incoming calls wait in a queue, through which users are serviced as the servers become free according to the FCFS discipline.

The hypercube model is a descriptive model that does not allow, if applied in isolation, the direct solution of location problems in order, for example, to reduce the average response time to the user or reduce the workload of ambulances. However, it is a model that allows the estimation of important performance measures, enabling the analysis of alternative scenarios.

According to Larson and Odoni (2007), there are nine critical hypotheses that must be verified for the application of the classic hypercube model:

1) The area must be divided into $N_A$ Travel times atoms.

2) Requests for service on each atom $j$ ($j$ = 1, ..., $N_A$), arrive independently according to a Poisson distribution.

3) Travel times $(\tau_{i,j})$ from atom $i$ to atom $j$ ($i$, $j$ = 1, ..., $N_A$) must be known or estimated.

4) The system operates with m servers (different or not) spatially distributed, which can move and serve any of the atoms.

5) The location of the servers must be known, at least probabilistically.

6) Only one server is dispatched to answer a call.

7) There is a server dispatch preference list for each atom.

8) The total time to answer a call is composed of the sum of the following times: server preparation time (setup time), server travel time to the place of occurrence, service execution time with the user (time in scene) and the time to return to base.

9) Variations in total service time due to variations in travel time are considered second order when compared to variations in stage times and/or team preparation time.

To present the non-stationary hypercube model, we used, also for comparison purposes with the equilibrium approach, an example of a system with three servers found in Chiyoshi *et. al* (2001). Consider an emergency system operating in a region represented by three atoms, using a fixed preference dispatch policy, shown in Table 1.

**Table1–** Dispatch Preferences Matrix.

| | Dispatch Matrix | | |
|---|---|---|---|
| | Preferences | | |
| Atom | **1º** | **2º** | **3º** |
| **1** | 1 | 2 | 3 |
| **2** | 2 | 3 | 1 |
| **3** | 3 | 1 | 2 |

Font – Chiyoshi *et. al*. (2001).

In a similar way to the previous models, the solution of the model is given by the construction of the transition equations of the system. The equations defining the non-stationary hypercube with m = 3 servers are defined by:

$$\{000\} \rightarrow \frac{dP_{000}}{d_t} = -\lambda(t)P_{000}(t) + \mu_1(t)P_{001}(t) + \mu_2(t)P_{010}(t) + \mu_3(t)P_{100}(t)$$ 50
$$\{001\} \rightarrow \frac{dP_{001}}{d_t} = -(\lambda(t) + \mu_1(t))P_{001}(t) + \lambda_1(t)P_{000}(t) + \mu_2(t)P_{110}(t) + \mu_3(t)P_{101}(t)$$
$$\{010\} \rightarrow \frac{dP_{010}}{d_t} = -(\lambda(t) + \mu_2(t))P_{010}(t) + \lambda_2(t)P_{000}(t) + \mu_1(t)P_{011}(t) + \mu_3(t)P_{110}(t)$$
$$\{100\} \rightarrow \frac{dP_{100}}{d_t} = -(\lambda(t) + \mu_3(t))P_{100}(t) + \lambda_3(t)P_{000}(t) + \mu_1(t)P_{101}(t) + \mu_2(t)P_{011}(t)$$
$$\{011\} \rightarrow \frac{dP_{011}}{d_t} = -(\lambda(t) + \mu_1(t) + \mu_2(t))P_{011}(t) + (\lambda_1(t) + \lambda_2(t))P_{001}(t) + \lambda_1(t)P_{010}(t) + \mu_3(t)P_{111}(t)$$
$$\{101\} \rightarrow \frac{dP_{101}}{d_t} = -(\lambda(t) + \mu_1(t) + \mu_3(t))P_{101}(t) + \lambda_3(t)P_{001}(t) + (\lambda_1(t) + \lambda_3(t))P_{100}(t) + \mu_2(t)P_{111}(t)$$
$$\{110\} \rightarrow \frac{dP_{110}}{d_t} = -(\lambda(t) + \mu_2(t) + \mu_3(t))P_{110}(t) + (\lambda_2(t) + \lambda_3(t))P_{010}(t) + \lambda_2(t)P_{100}(t) + \mu_1(t)P_{111}(t)$$
$$\{111\} \rightarrow \frac{dP_{111}}{d_t} = -(\lambda(t) + \mu(t))P_{111}(t) + \lambda(t)P_{011}(t) + \lambda(t)P_{101}(t) + \lambda(t)P_{110}(t)$$
$$\{S_1\} \rightarrow \frac{dP_{S_1}}{d_t} = -(\lambda(t) + \mu(t))P_{S_1}(t) + \lambda(t)P_{111}(t) + \mu(t)P_{S_2}(t)$$
$$\vdots$$
$$\{S_{c(t)}\} \rightarrow \frac{dP_{S_1}}{d_t} = -\mu(t)P_{S_{c(t)}}(t) + \lambda(t)P_{S_{c(t)-1}}(t) .$$

What,

$\lambda_i(t)$ is the call arrival rate on atom i, at the instant $t$;
$\mu_j(t)$ is the service rate of server j, at the instant $t$;
$\lambda(t) = \lambda_1(t) + \lambda_2(t) + \lambda_3(t)$ is the total rate of arrival in the system, at the instant $t$;
$\rho(t) = \frac{\lambda(t)}{\mu(t)}$ is the average workload on the system at the instant $t$.

If the system changes the number of servers over time, the number of equations in the system will be $2^{m(t)} + c(t)$. System capacity changes as the number of servers (m(t)) changes, but the number of users in the queue is fixed (c(t)) as in the model in Section 2.4.3. The solution of the system of differential equations gives the state probabilities at t (t>0) and from them we can calculate several important performance measures for the system. As in the previous sections, the average number of users in the system (L(t)), the average number of users in the queue (L_q (t)), the average waiting time in the system (W(t)) and the time average queue waiting (W_q (t)) in addition to other performance measures such as workloads, dispatch frequencies and travel times that have been adapted for this analysis as follows:

Workload, at time t, is the fraction of time the server is busy and is calculated by adding the probabilities of the states in which this server is busy.

- $\rho_i(t) = \sum_{\{B(t):b_i(t)\}} P_B(t) + P_Q(t),$ 51

What,

- ✓ $\rho_i(t)$ is the *workload* of the server *i* (*i* = 1, 2, ..., *m*), at the instant *t*;
- ✓ $\sum_{\{B(t):b_i(t)\}} P_B(t)$ is the sum of the probabilities of the states (from {000} to {111}) where server i is busy ($b_i = 1$), at the instant *t*;

- ✓ $P_Q(t)$ is the probability of queue $\left(P_Q(t) = P_{S_4}(t) + P_{S_5}(t) + \cdots\right)$ at the instant $t$.

The average workload over an interval $(t_1, t_2)$ is given by:

$$\rho_i\,(t_1, t_2) = \frac{\int_{t_1}^{t_2} \sum_{\{B(t):b_i(t)\}} P_B(t) + P_Q(t)}{t_2 - t_1} \qquad 52$$

The dispatch frequency, at instant t, indicates the fraction of dispatches in the system that is served by the server $i$ ($i$ = 1, 2, ..., $m$) in atom $j$ ($j$ = 1, 2, ..., $N_A$), and is given by the sum of two parts: $f_{ij}^{[nq]}$, fraction of dispatches in which server i is sent to atom j, but does not imply waiting time in queue; $f_{ij}^{[q]}$, fraction of dispatches in which server i is sent to atom j, and implies waiting time in queue:

$$f_{ij}(t) = f_{ij}^{[nq]}(t) + f_{ij}^{[q]}(t) = \frac{\lambda_j(t)}{\lambda(t)} \sum_{B(t) \in E_{ij}(t)} P_B\,(t) + \frac{\lambda_j(t)}{\lambda(t)} P_{Q'}(t) \frac{\mu_i(t)}{\mu(t)}. \qquad 53$$

The dispatch frequency, in an interval $(t_1, t_2)$, is given by:

$$f_{ij}\,(t_1, t_2) = \frac{\int_{t_1}^{t_2} \frac{\lambda_j(t)}{\lambda(t)} \sum_{B(t) \in E_{ij}(t)} P_B(t) + \frac{\lambda_j(t)}{\lambda(t)} P_{Q'}(t) \frac{\mu_i(t)}{\mu(t)}}{t_2 - t_1}. \qquad 54$$

The travel time in the system, at instant t, is:

$$\bar{T}(t) = \sum_{i=1}^{m} \sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) t_{ij} + P_{Q'}(t) \bar{T}_Q(t)\,. \qquad 55$$

What,

- ✓ $t_{ij}$, is the average travel time from server i to atom j;
- ✓ $\bar{T}_Q(t) = \sum_{i=1}^{m} \sum_{j=1}^{N_A} \frac{\lambda_p(t)\lambda_j(t)}{\lambda^2(t)} \tau_{pj}$, it's travel time for queued calls, at instant t;
- ✓ $\tau_{pj}$, is the average travel time between atoms p and j.

➢ The average travel time in the system, over an interval $(t_1, t_2)$, is given by:

$$\bar{T}\,(t_1, t_2) = \frac{\int_{t_1}^{t_2} \sum_{i=1}^{m} \sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) t_{ij} + P_{Q'}(t) \bar{T}_Q(t)}{t_2 - t_1}\,. \qquad 56$$

The travel time to atom j, at time t, is calculated by:

$$\overline{T}_j(t) = \frac{\sum_{i=1}^{m} f_{ij}^{[nq]}(t) t_{ij}}{\sum_{i=1}^{m} f_{ij}^{[nq]}(t)} \left(1 - P_{Q'}(t)\right) + \sum_{p=1}^{N_A} \frac{\lambda_{p(t)}}{\lambda(t)} \tau_{ij} P_{Q'}(t). \quad 57$$

The average travel time to atom j over an interval $(t_1, t_2)$, is calculated by:

$$\overline{T}_j(t_1, t_2) = \frac{\int_{t_1}^{t_2} \frac{\sum_{i=1}^{m} f_{ij}^{[nq]}(t) t_{ij}}{\sum_{i=1}^{m} f_{ij}^{[nq]}(t)} \left(1 - P_{Q'}(t)\right) + \sum_{p=1}^{N_A} \frac{\lambda_{p(t)}}{\lambda(t)} \tau_{ij} P_{Q'}(t)}{t_2 - t_1} \quad 58$$

The travel time of server i, at time t, can be approximated by:

$$\overline{TU}_i(t) = \frac{\sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) t_{ij} + \left(T_Q(t) P_{Q'}(t)/m\right)}{\sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) + \left(P_{Q'}(t)/m\right)}. \quad 59$$

The average travel time, of server i, over an interval $(t_1, t_2)$, can be approximated by:

$$\overline{TU}_i(t_1, t_2) = \frac{\int_{t_1}^{t_2} \frac{\sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) t_{ij} + \left(T_Q(t) P_{Q'}(t)/m\right)}{\sum_{j=1}^{N_A} f_{ij}^{[nq]}(t) + \left(P_{Q'}(t)/m\right)}}{t_2 - t_1}. \quad 60$$

## 3.TOY MODELS: THREE SERVERS

### 3.1 Changing the number of servers in the model *M(t)/M(t)/m(t)/C(t)*

In real systems it is common to change the number of servers depending on the arrival rate in the service, periods of high demand lead to more servers and periods of low demand make it possible to work with fewer servers without compromising the quality of service. Using the approach of Souza et. al. (2013) the day was divided into three and six periods. The M(t)/M(t)/m(t)/C(t) model with dynamic approach the arrival rates vary over time and the service rate is fixed, we are not considering server acceleration. Three analyzes will be performed, the transition rates for one day of system operation are:

a) $\lambda(t) = \begin{cases} 2, se\ 23h < t \leq 7h \\ 4, se\ 7h < t \leq 14h \\ 5, se\ 14h < t \leq 23h \end{cases}$, b) $\lambda(t) = \begin{cases} 1, se\ 2h < t \leq 6h \\ 2, se\ 6h < t \leq 10h \\ 4, se\ 10h < t \leq 15h \\ 3, se\ 15h < t \leq 18h \\ 5, se\ 18h < t \leq 22h \\ 2, se\ 22h < t \leq 2h \end{cases}$ and

c) $\lambda(t) = 5\ \ if\ 0h < t \leq 24h.$

For the three analyzes considered, it will be considered $\mu(t) = \mu = 6$. In this case, periods where the arrival fee is equal to 5 ($\lambda(t)$ = 5) we consider that $m(t)$ = 2 e $C(t)$ = 5 and in periods when the arrival rate is less than 5 ($\lambda(t) < 5$) we consider that $m(t) = 1$ e $C(t) = 6$, in both cases the queue length is the same.

The system in Section 2.1 was solved by the Runge-Kutta method (WILMER *et. al.*, 1995) with an error less than 10-6 of the exact solution, we obtain the probabilities of each state at each instant of time. The initial conditions used in this example were defined as follows: the probability that the system is empty at time t = 0 is a $(P_0(t = 0) = 1)$ the probabilities of the other states are equal to zero; when the system increases one server, the probability of that server being busy the moment it starts operating is zero; when the system downgrades a server, the probability that the retired server is busy is divided equally into the queue states at the time the server is retired. Furthermore, the results obtained have an accuracy of 10-6 of the exact solution.

Figure 12a shows the probabilities of each system state for the non-stationary model with changing the number of servers and the day divided into three periods, Figure 12b shows the probabilities of each system state for the non-stationary model with changing the number of servers and the day divided into three periods and Figure 12c shows the probabilities of each system state for the equilibrium model, the analysis was performed for a 24-hour period. We can observe that from the point where there is a change in system parameters (rate of arrival and service) and/or a change in the number of servers, the system enters the transient state again until it comes to equilibrium.

**Figure 12 -** Probabilities of queue states (P_B ) M(t)/M(t)/m(t)/C(t) a) with the day divided into three periods; b) with the day divided into six periods; c) probabilities of queue states M/M/2/5.

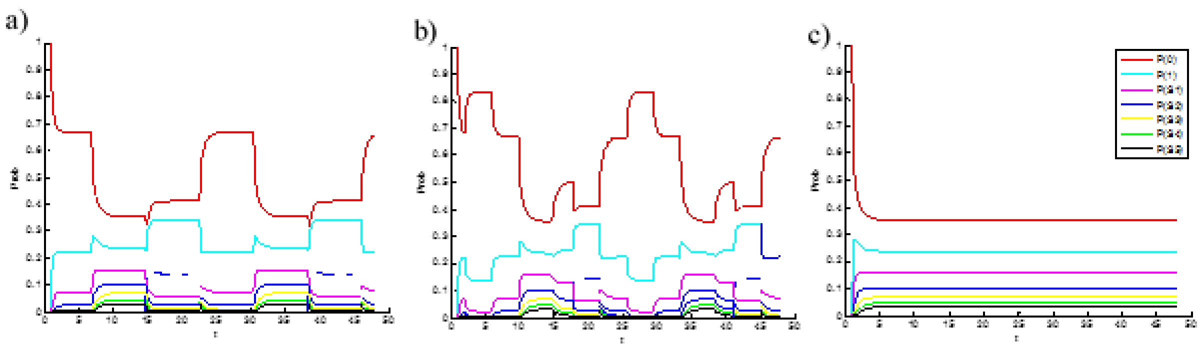



Figure 13a shows a comparison between the number of users in the M/M/2/5 model system (light green line) with the non-stationary model where there is a change in the number of servers with the day divided into three (green line dark) and six periods (blue line). Figure 13b shows a comparison between the waiting time in the system of the M/M/2/5 model (light green line) with the non-stationary model where there is a change in the number of servers with the day divided into three (green line dark) and six periods (blue line).

**Figure 13 –** a) Comparison of the number of users in the system $(\bar{L}(t))$ the M/M/2/5 model in equilibrium with the M(t)/M(t)/m(t)/C(t) model with the day divided into three and six periods; b) Comparison of waiting time in the system $(\bar{W}(t))$ of the M/M/2/5 model in equilibrium with the M(t)/M(t)/m(t)/C(t) model with the day divided into three and six periods.

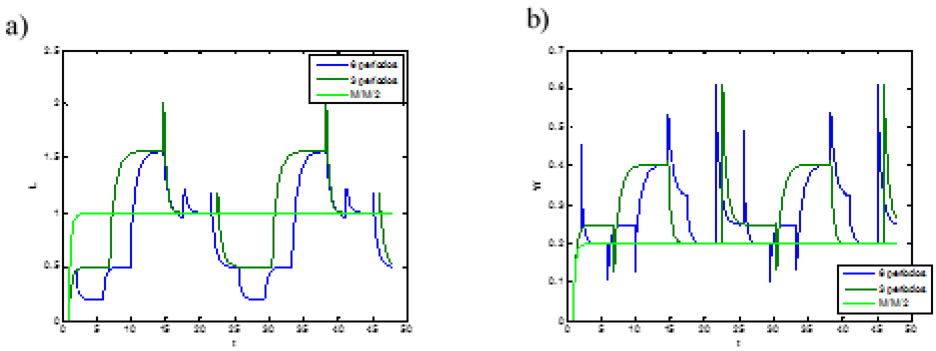
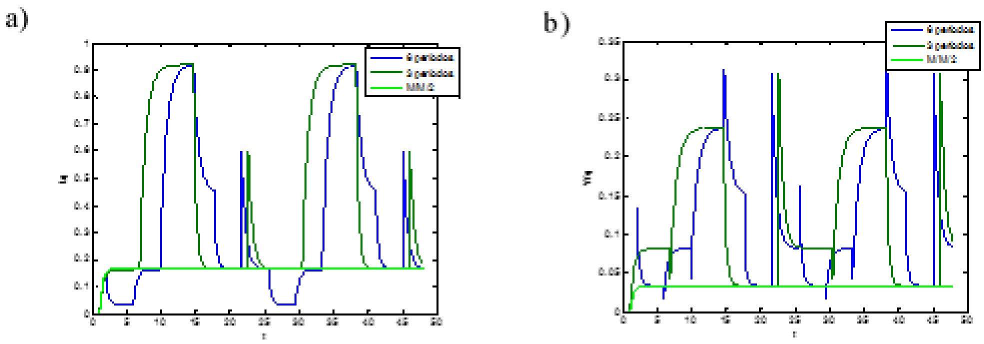



Figure 14a shows a comparison between the number of users in the system calculated by the balanced M/M/2/5 model (light green line) with the values obtained by the non-stationary models with change in the number of servers with the day divided into three (dark green line) and six periods (blue line). Figure 14b shows a comparison between the results of waiting time in the queue obtained by the balanced M/M/2/5 model (light green line) and by the non-stationary model where there is a change in the number of servers with the divided day in three (dark green line) and six periods (blue line).

**Figure 14 -** Comparison of queued user number $(\bar{L}_q(t))$ the M/M/2/5 model with the M(t)/M(t)/m(t)/C(t) model with the day divided into three and six periods; b) Comparison of waiting time in line $(\bar{W}_q(t))$ from the M/M/2/5 model to the M(t)/M(t)/m(t)/C(t) model in which there is a change in the number of employees with the day divided into three and six periods.




We can observe that the dynamic queuing analysis considers the effect of changing the arrival rate and the number of servers in the system on the performance measures analyzed, this change in parameters entails a transition period that the equilibrium model does not consider. This analysis allows for a number of issues to be analyzed, such as the practical applicability in real systems that need to change the number of servers depending on the variation in demand throughout the day. The number of periods in the day to be chosen for this analysis depends on the rate of arrivals in the system, the greater the number of tickets, the greater the possibility of dividing the day into smaller periods. If the arrival rate of the system is large enough to divide the day into small enough periods, there is also the possibility that the arrival rate will be modeled by a continuous function.

Table 2 shows the average performance measures calculated from the equilibrium analysis, where each period is analyzed assuming it is in equilibrium and is independent of its predecessor period and its successor period. In the dynamic approach, information from the immediately preceding period is considered as initial information for the model in the successor period. The results are divided into analyzes for the day divided into three periods, with deviations from the dynamic approach in relation to the equilibrium approach that reached the level of 86.40%, and six periods, with deviations from the dynamic approach in relation to the equilibrium approach which in this case reached the level of 59.14%.

**Table2**– Comparison of the average performance measures of the equilibrium models M/M/1/6 and M/M/2/7 with the non-stationary model, with the day divided into three and six periods.

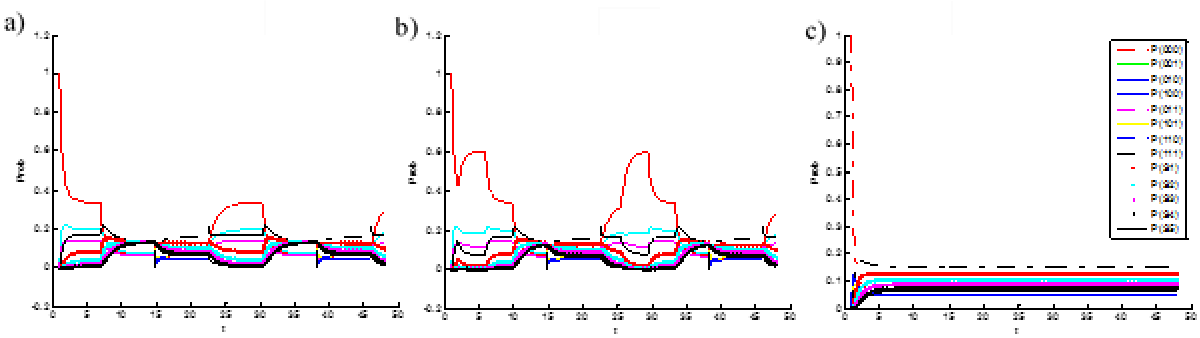
| three periods | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | balance | | | | non-stationary | | | | deviation (%) | | | |
| ***t* (min)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** |
| 1:360 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,4661 | 0,1471 | 0,2332 | 0,0736 | -6,18 | -10,16 | -6,19 | -10,17 |
| 361:840 | 1,5648 | 0,9189 | 0,4038 | 0,2371 | 1,4628 | 0,8362 | 0,3760 | 0,2150 | -6,52 | -9,00 | -6,87 | -9,30 |
| 840:1320 | 0,9985 | 0,1667 | 0,2001 | 0,0334 | 1,0538 | 0,1992 | 0,2118 | 0,0401 | 5,53 | 19,49 | 5,87 | 20,01 |
| 1321:1800 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,5447 | 0,1981 | 0,2732 | 0,0994 | 9,63 | 20,94 | 9,87 | 21,29 |
| 1801:2280 | 1,5648 | 0,9189 | 0,4038 | 0,2371 | 1,4632 | 0,8364 | 0,3761 | 0,2151 | -6,50 | -8,97 | -6,84 | -9,27 |
| 2281:2760 | 0,9985 | 0,1667 | 0,2001 | 0,0334 | 1,0538 | 0,1992 | 0,2115 | 0,0400 | 5,54 | 19,49 | 5,72 | 19,86 |
| 2761:2880 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,6909 | 0,3028 | 0,3482 | 0,1528 | 39,07 | 84,91 | 40,07 | 86,40 |
| **Six periods** | | | | | | | | | | | | |
| | **balance** | | | | **non-stationary** | | | | **deviation (%)** | | | |
| ***t* (min)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** | ***L*(*t*)** | ***Lq*(*t*)** | ***W*(*t*)** | ***Wq*(*t*)** |
| 1:60 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,3345 | 0,0796 | 0,1673 | 0,0398 | -32,67 | -51,41 | -32,73 | -51,45 |
| 61:300 | 0,2000 | 0,0333 | 0,2000 | 0,0333 | 0,2205 | 0,0429 | 0,2205 | 0,0429 | 10,25 | 28,79 | 10,26 | 28,80 |
| 301:540 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,4665 | 0,1457 | 0,2334 | 0,0729 | -6,10 | -11,01 | -6,12 | -11,02 |
| 541:840 | 1,5648 | 0,9189 | 0,4038 | 0,2371 | 1,4021 | 0,7870 | 0,3595 | 0,2019 | -10,40 | -14,36 | -10,95 | -14,84 |
| 841:1020 | 0,9449 | 0,4488 | 0,3175 | 0,1508 | 1,0920 | 0,5642 | 0,3692 | 0,1908 | 15,57 | 25,72 | 16,29 | 26,55 |
| 1021:1260 | 0,9985 | 0,1667 | 0,2001 | 0,0334 | 1,0307 | 0,1882 | 0,2067 | 0,0377 | 3,22 | 12,91 | 3,29 | 13,02 |
| 1261:1500 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,5969 | 0,2355 | 0,3000 | 0,1185 | 20,15 | 43,78 | 20,65 | 44,53 |
| 1501:1740 | 0,2000 | 0,0333 | 0,2000 | 0,0333 | 0,2248 | 0,0457 | 0,2248 | 0,0457 | 12,39 | 37,29 | 12,41 | 37,31 |
| 1741:1980 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,4666 | 0,1458 | 0,2334 | 0,0729 | -6,09 | -10,99 | -6,11 | -11,01 |
| 1981:2280 | 1,5648 | 0,9189 | 0,4038 | 0,2371 | 1,4026 | 0,7873 | 0,3597 | 0,2020 | -10,37 | -14,32 | -10,92 | -14,80 |
| 2281:2460 | 0,9449 | 0,4488 | 0,3175 | 0,1508 | 1,0925 | 0,5646 | 0,3694 | 0,1910 | 15,63 | 25,80 | 16,35 | 26,64 |
| 1461:2700 | 0,9985 | 0,1667 | 0,2001 | 0,0334 | 1,0307 | 0,1882 | 0,2067 | 0,0377 | 3,22 | 12,91 | 3,29 | 13,02 |
| 2701:2880 | 0,4968 | 0,1638 | 0,2486 | 0,0820 | 0,6297 | 0,2590 | 0,3168 | 0,1304 | 26,76 | 58,14 | 27,43 | 59,14 |

### 3.2 Changing the number of servers in the non-stationary hypercube model

To represent the example where there is a change in the number of servers, we consider that there are three servers in periods of high demand (18h<t≤22h) and one server decreases when the demand is low, server 2 is withdrawn. Queue formation of up to five users in the queue is allowed. The total system arrival rate $(\lambda(t))$ is the same as in the previous section and we consider that $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = \lambda(t)/3$. The service time rates have been defined such that, $\mu_1(t) = \mu_1 = 1.5$, $\mu_2(t) = \mu_2 = 3.0$ e $\mu_3(t) = \mu_3 = 1.5$. The dispatch preference list is the same as in table 3.1.

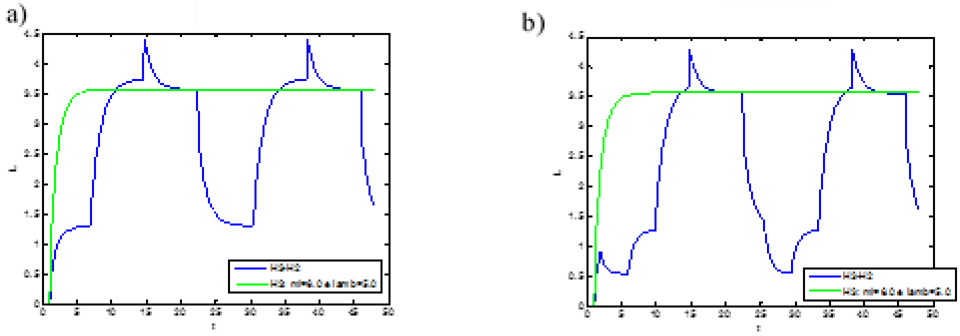
Solving the system of differential equations in Section 2.2 by the Runge-Kutta method we obtain the probabilities of each state for each instant of time (Figure 15a and 15b). The initial conditions used in this example were defined as follows: the probability that the system is empty at time t = 0 is a $(P_0(t = 0) = 1)$ the probabilities of the other states are equal to zero; when server 2 is placed on the system, the probabilities of all states that this server is busy at the time it starts operating is zero. (p.e., $P_{010}=P_{110}=P_{011}=P_{111}=0?$); when the system downs a server, server 2 is removed from the system, we have to $(P_{00}(t = t_0) = P_{000} + P_{010})$, $(P_{01}(t = t_0) = P_{001} + P_{011})$, $(P_{10}(t = t_0) = P_{100} + P_{110})$ and $(P_{11}(t = t_0) = P_{101} + P_{111})$, $t_0$ is the instant of time when server 2 is taken down. Furthermore, the results were obtained with an accuracy of $10^{-6}$.

**Figure 15** - a) states probabilities $(P_B)$ in the non-stationary hypercube model with the day divided into three periods; b) state probabilities of the non-stationary hypercube model with the day divided into six periods; c) state probabilities in the stationary hypercube model.



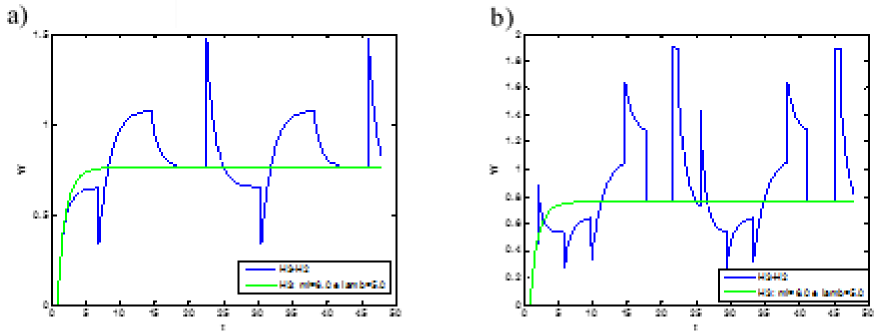

Figures 16a and 16b show a comparison between the average number of users in the system of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively. We can observe that, in both cases, in the period from 6 pm to 10 pm, the system comes into equilibrium at 8 pm.

**Figure 16** - Comparison of the number of users in the system $(\bar{L}(t))$ the non-stationary hypercube model and the stationary hypercube model a) with the day divided into three periods; b) with the day divided into six periods.



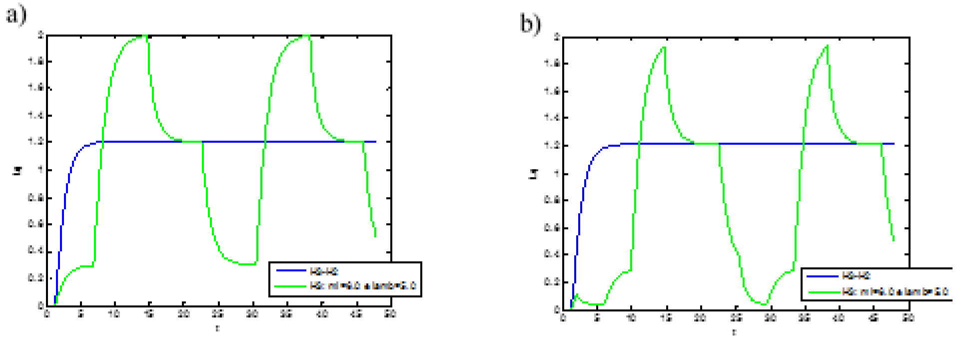

Figures 17a and 17b show a comparison between the results of waiting time in the system obtained by the dynamic approach of the hypercube model and by the approach in equilibrium with the day divided into three and six periods, respectively.

**Figure 17** - Comparison of waiting time in the system $(\bar{W}(t))$ the non-stationary hypercube model and the stationary hypercube model a) with the day divided into three periods; b) with the day divided into six periods.



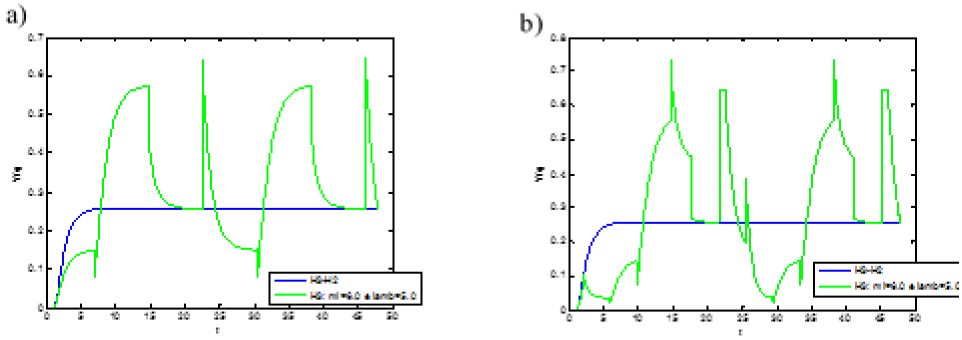

Figures 18a and 18b show a comparison between the number of users in the queue of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively.

**Figure 18** - Comparing the number of users in the queue $(\bar{L}_q(t))$ the non-stationary hypercube model and the stationary hypercube model a) with the day divided into three periods; b) with the day divided into six periods.



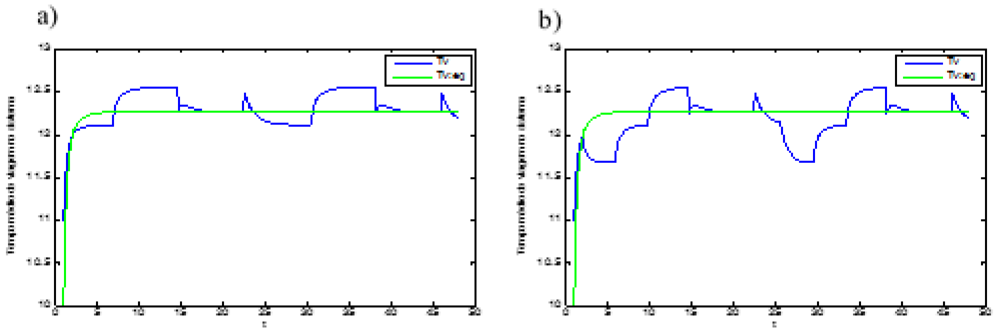
Figures 19a and 19b show a comparison between the waiting time in the queue of the dynamic approach with the equilibrium approach of the hypercube model with the day divided into three and six periods, respectively.

**Figure 19** - Comparison of waiting time in line $(\bar{W}_q(t))$ the non-stationary hypercube model and the stationary hypercube model a) with the day divided into three periods; b) with the day divided into six periods.



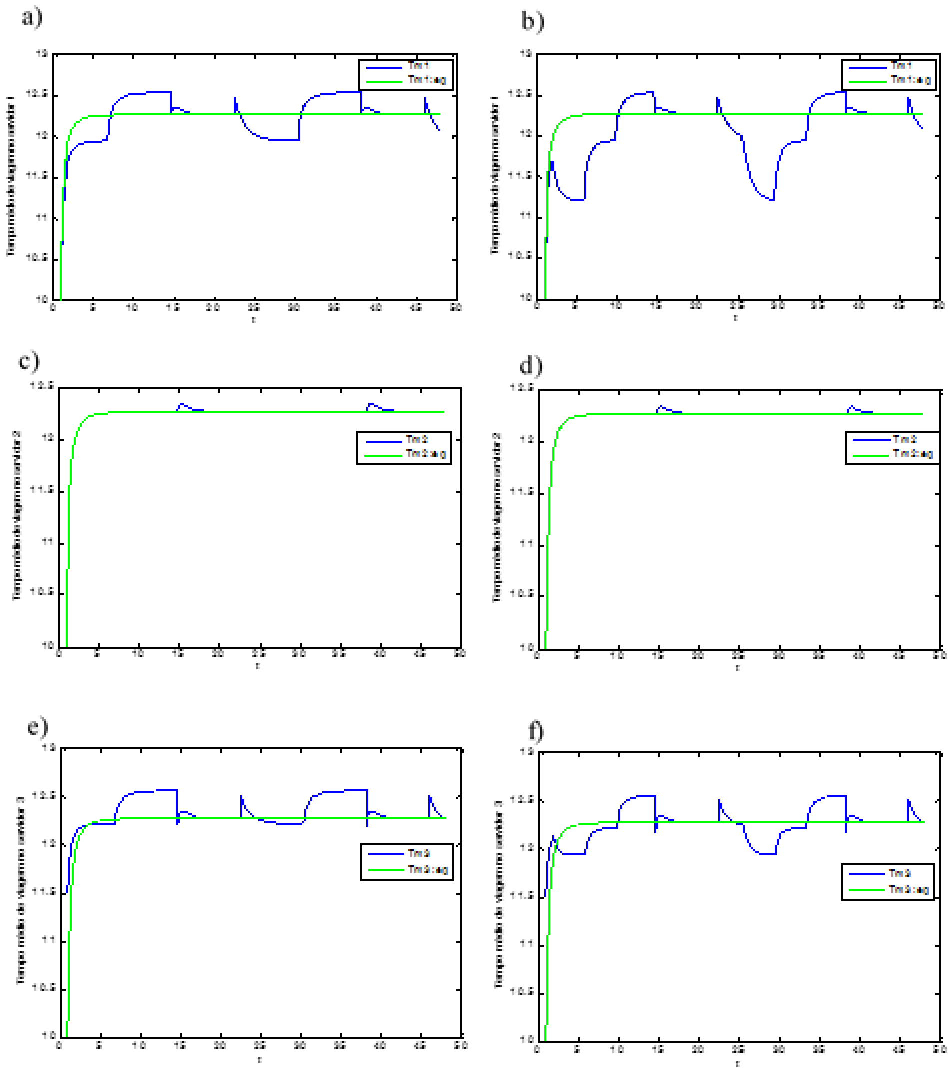
Figures 20a and 20b show the comparison between the travel time in the system of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively.

**Figure 20** - Comparison of travel time in the system $(\bar{T}(t))$ the non-stationary hypercube model and the stationary hypercube model a) with the day divided into three periods; b) with the day divided into six periods.



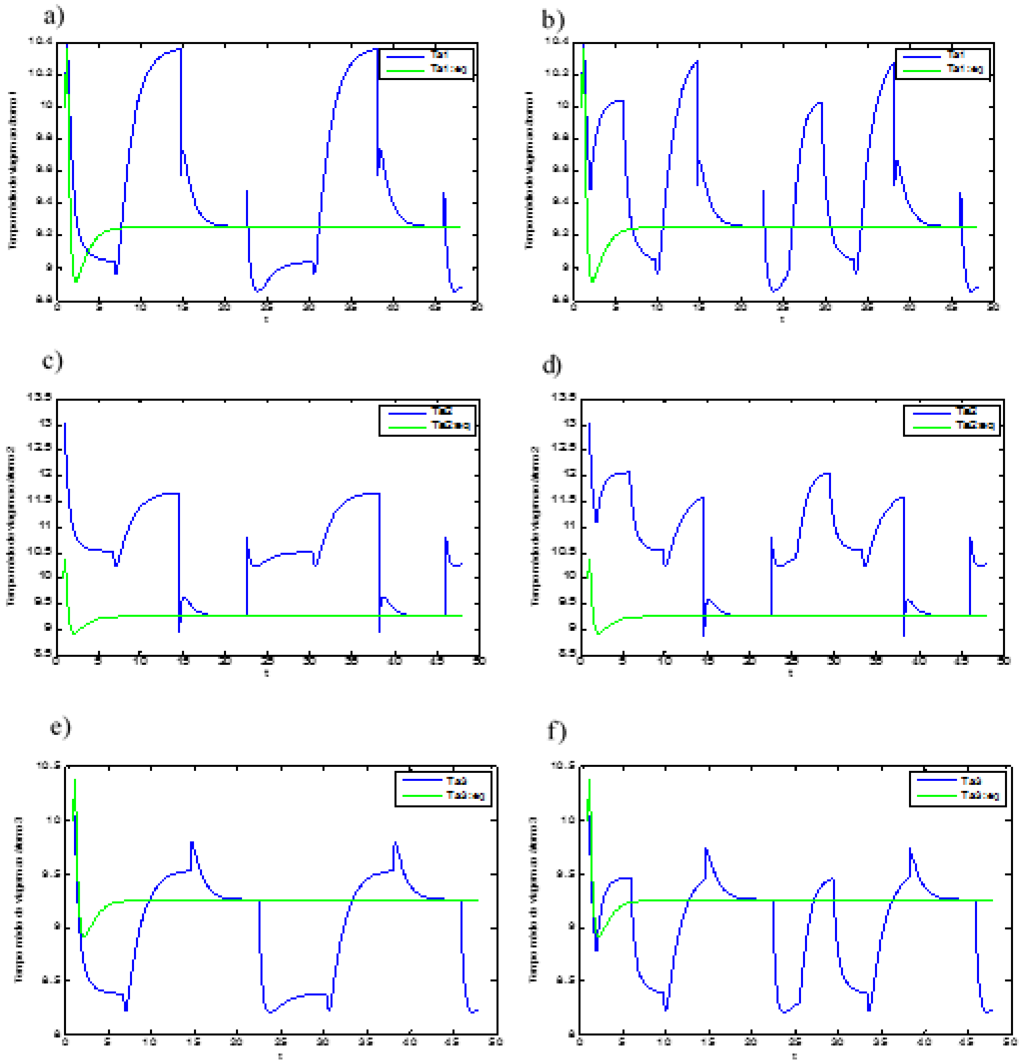
Figures 21a and 21b show the comparison between the travel time on server 1 of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively. Figures 21c and 21d show the comparison between the travel time on server 2 of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively. Figures 21e and 21f show the comparison between the travel time on server 3 of the dynamic approach of the hypercube model with the approach in equilibrium with the day divided into three and six periods, respectively.

**Figure 21** - Comparison of travel time on the server *i* $(\bar{T}_i(t))$ from the non-stationary hypercube model and the stationary hypercube model as follows: i = 1 with the day divided into three periods (a) and six periods (b); i = 2 with the day divided into three periods (c) and six periods (d); i = 3 with the day divided into three periods (e) and six periods (f).



Figures 22 and 22b show the comparison between travel time on atom 1 of the hypercube model's dynamic approach with the equilibrium approach with the day divided into three and six periods, respectively. Figures 22c and 22d show the comparison between the travel time in atom 2 of the hypercube model's dynamic approach with the approach in equilibrium with the day divided into three and six periods, respectively. Figures 22e and 22f show the comparison between the travel time in atom 3 of the hypercube model's dynamic approach with the approach in equilibrium with the day divided into three and six periods, respectively.

**Figure 22** - Comparison of travel time on the atom *j* $(\bar{\tau}_j(t))$ from the non-stationary hypercube model and the stationary hypercube model as follows: j = 1 with the day divided into three periods (a) and six periods (b); j = 2 with the day divided into three periods (c) and six periods (d); j = 3 with the day divided into three periods (e) and six periods (f).




During the system's peak period (between 6 pm and 10 pm), operated with three servers, we can verify by analyzing the non-stationary hypercube model that the system takes time to reach equilibrium, even considering that the arrival rate does not vary. In all cases, at least 2 hours are required for the system to come into balance assuming that the arrival rate does not change during this period. This study shows the importance of considering the study of non-stationary queuing models in real systems to incorporate the change in arrival rate and service in the system varying with time, as well as the change in the number of servers.

Table 3 shows the average number of users in the system (L), the average number of users in the queue (Lq), the average time spent in the system (W) and the waiting time in the queue (Wq) calculated from the equilibrium analysis. For the day divided into three periods, the deviations of the dynamic approach in relation to the equilibrium approach reached the level of 164.93% and, for the day divided into six periods, the deviations of the non-stationary approach in relation to the equilibrium approach that in this case reached the level of 220.13%.

**Table 3** - Comparison of the average performance measures of the hypercube in equilibrium model with the non-stationary model, with the day divided into three and six periods.

| Three periods | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | balance | | | | non-stationary | | | | deviation (%) | | | |
| t (min) | L | Lq | W | Wq | L | Lq | W | Wq | L | Lq | W | Wq |
| 1:360 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 1,1027 | 0,2155 | 0,5532 | 0,1082 | -14,73 | -27,79 | -14,89 | -27,90 |
| 361:840 | 3,7333 | 2,0000 | 1,0769 | 0,5769 | 3,2962 | 1,6520 | 0,9245 | 0,4650 | -11,71 | -17,40 | -14,16 | -19,39 |
| 840:1320 | 3,5547 | 1,2084 | 0,7575 | 0,2575 | 3,6876 | 1,3091 | 0,7938 | 0,2821 | 3,74 | 8,33 | 4,79 | 9,55 |
| 1321:1800 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 1,5289 | 0,4465 | 0,7758 | 0,2272 | 18,22 | 49,60 | 19,36 | 51,48 |
| 1801:2280 | 3,7333 | 2,0000 | 1,0769 | 0,5769 | 3,2977 | 1,6531 | 0,9249 | 0,4654 | -11,67 | -17,34 | -14,11 | -19,33 |
| 2281:2760 | 3,5547 | 1,2084 | 0,7575 | 0,2575 | 3,6876 | 1,3092 | 0,7938 | 0,2821 | 3,74 | 8,33 | 4,79 | 9,55 |
| 2761:2880 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 2,0427 | 0,7701 | 1,0527 | 0,3974 | 57,96 | 158,05 | 61,96 | 164,93 |
| Six periods | | | | | | | | | | | | |
| | balance | | | | non-stationary | | | | deviation (%) | | | |
| t (min) | L | Lq | W | Wq | L | Lq | W | Wq | L | Lq | W | Wq |
| 1:60 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 0,5717 | 0,0421 | 0,2859 | 0,0210 | -55,79 | -85,91 | -56,02 | -85,98 |
| 61:300 | 0,5331 | 0,0332 | 0,5332 | 0,0332 | 0,5934 | 0,0478 | 0,5935 | 0,0478 | 11,31 | 44,17 | 11,32 | 44,19 |
| 301:540 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 1,1121 | 0,2080 | 0,5576 | 0,1043 | -14,01 | -30,32 | -14,21 | -30,47 |
| 541:840 | 3,7333 | 2,0000 | 1,0769 | 0,5769 | 3,0473 | 1,4548 | 0,8389 | 0,4024 | -18,38 | -27,26 | -22,11 | -30,25 |
| 841:2020 | 2,4504 | 1,0150 | 0,8535 | 0,3535 | 3,8505 | 1,4358 | 1,3976 | 0,5217 | 57,14 | 41,46 | 63,74 | 47,57 |
| 1021:1260 | 3,5547 | 1,2084 | 0,7575 | 0,2575 | 3,5736 | 1,2217 | 0,7624 | 0,2606 | 0,53 | 1,10 | 0,64 | 1,21 |
| 1261:1500 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 2,2874 | 0,7961 | 1,1906 | 0,4147 | 76,88 | 166,75 | 83,17 | 176,49 |
| 1501:1740 | 0,5331 | 0,0332 | 0,5332 | 0,0332 | 0,7174 | 0,1044 | 0,7188 | 0,1047 | 34,56 | 214,56 | 34,81 | 215,46 |
| 1741:1980 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 1,1138 | 0,2087 | 0,5585 | 0,1047 | -13,87 | -30,05 | -14,08 | -30,20 |
| 1981:2280 | 3,7333 | 2,0000 | 1,0769 | 0,5769 | 3,0429 | 1,4518 | 0,8375 | 0,4015 | -18,49 | -27,41 | -22,23 | -30,40 |
| 2281:2460 | 2,4504 | 1,0150 | 0,8535 | 0,3535 | 3,8503 | 1,4356 | 1,3975 | 0,5217 | 57,13 | 41,44 | 63,73 | 47,55 |
| 2461:2700 | 3,5547 | 1,2084 | 0,7575 | 0,2575 | 3,5736 | 1,2217 | 0,7624 | 0,2606 | 0,53 | 1,10 | 0,64 | 1,21 |
| 2701:2880 | 1,2932 | 0,2984 | 0,6500 | 0,1500 | 2,5478 | 0,9174 | 1,3338 | 0,4802 | 97,02 | 207,40 | 105,20 | 220,13 |

Table 4 shows the average travel times of servers 1, 2 and 3 calculated from the equilibrium analysis. For the two approaches, dynamic and in equilibrium, the deviations were relatively small, reaching a maximum of 6.96%. All performance measures related to travel times had deviations of this order, which can be explained by the fact that the model works with mean values of travel times, which do not vary with time, in both approaches. Instantaneous speed values obtained, for example, from GPS devices can generate different results.

**Table 4** - Comparison of average travel times for servers in the hypercube model in equilibrium with the non-stationary model, with the day divided into three and six periods.

| Three periods | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | balance | | | non-stationary | | | deviation (%) | | |
| t (min) | tu1 | tu2 | tu3 | tu1 | tu2 | tu3 | tu1 | tu2 | tu3 |
| 1:360 | 12,0357 | | 12,1730 | 11,7926 | | 12,1620 | -2,02 | | -0,09 |
| 361:840 | 12,5473 | | 12,5637 | 12,4895 | | 12,5305 | -0,46 | | -0,26 |
| 840:1320 | 12,2860 | 12,2842 | 12,2842 | 12,2860 | 12,2842 | 12,2842 | 0,00 | 0,00 | 0,00 |
| 1321:1800 | 12,0357 | | 12,1730 | 12,0341 | | 12,2605 | -0,01 | | 0,72 |
| 1801:2280 | 12,5473 | | 12,5637 | 12,4898 | | 12,5306 | -0,46 | | -0,26 |
| 2281:2760 | 12,2860 | 12,2842 | 12,2842 | 12,2860 | 12,2842 | 12,2842 | 0,00 | 0,00 | 0,00 |
| 2761:2880 | 12,0357 | | 12,1730 | 12,2113 | | 12,3497 | 1,46 | | 1,45 |
| **Six periods** | | | | | | | | | |
| | balance | | | non-stationary | | | deviation (%) | | |
| t (min) | tu1 | tu2 | tu3 | tu1 | tu2 | tu3 | tu1 | tu2 | tu3 |
| 1:60 | 12,0357 | ----- | 12,1730 | 11,1980 | ----- | 11,9511 | -6,96 | ----- | -1,82 |
| 61:300 | 11,4359 | ----- | 11,8431 | 11,2962 | ----- | 11,9571 | -1,22 | ----- | 0,96 |
| 301:540 | 12,0357 | ----- | 12,1730 | 11,8375 | ----- | 12,1756 | -1,65 | ----- | 0,02 |
| 541:840 | 12,5473 | ----- | 12,5637 | 12,4575 | ----- | 12,5094 | -0,72 | ----- | -0,43 |
| 841:2020 | 12,3736 | ----- | 12,4210 | 12,3090 | ----- | 12,3040 | -0,52 | ----- | -0,94 |
| 1021:1260 | 12,2860 | 12,2842 | 12,2842 | 12,2689 | 12,2689 | 12,2689 | -0,14 | -0,12 | -0,12 |
| 1261:1500 | 12,0357 | ----- | 12,1730 | 12,1800 | ----- | 12,3055 | 1,20 | ----- | 1,09 |
| 1501:1740 | 11,4359 | ----- | 11,8431 | 11,3962 | ----- | 11,9900 | -0,35 | ----- | 1,24 |
| 1741:1980 | 12,0357 | ----- | 12,1730 | 11,8388 | ----- | 12,1760 | -1,64 | ----- | 0,02 |
| 1981:2280 | 12,5473 | ----- | 12,5637 | 12,4564 | ----- | 12,5087 | -0,72 | ----- | -0,44 |
| 2281:2460 | 12,3736 | ----- | 12,4210 | 12,3090 | ----- | 12,3039 | -0,52 | ----- | -0,94 |
| 2461:2700 | 12,2860 | 12,2842 | 12,2842 | 12,2689 | 12,2689 | 12,2689 | -0,14 | -0,12 | -0,12 |
| 2701:2880 | 12,0357 | ----- | 12,1730 | 12,2300 | ----- | 12,3221 | 1,61 | ----- | 1,23 |

## 4. CONCLUSION

The objective of this work is to verify the effect of the variation of the arrival and service rate during the day in detriment of the equilibrium analysis of the system. For this, we developed two illustrative examples. The models M(t)/M(t)/m(t)/C(t) and the non-stationary hypercube model were approached, considering the change in the number of servers throughout the day. The study showed how the dynamic approach is more realistic than the equilibrium approach in systems where the variation of parameters is an important factor to be considered. Furthermore, this study suggests that the greater the variation of these parameters in shorter periods, the lower the probability of the system entering into equilibrium. In future works, the incorporation of queue priority in the non-stationary model can be studied, as well as the carrying out of case studies to verify the real applicability of this analysis.

**Acknowledgments**: This study has the financial support of FAPESP – Foundation for Research Support of the State of São Paulo.

## References

ALANIS, R., INGOLFSSON, A., & KOLFAL, B. A Markov chain model for an EMS system with repositioning. **Production and Operations Management**, v.22, n.1, p.216-231, 2013.

BOFFEY B., GALVÃO R.D., ESPEJO L.G.A. A review of congestion models in the location of facilities with immobile servers. **European Journal of Operational Research**. v.178, p. 643–662, 2007.

CHIYOSHI F., GALVÃO R.D., MORABITO R. Modelo hipercubo: análise e resultados para o caso de servidores não-homogêneos**. Pesquisa Operacional**, v. 21, n.2, p.199-218, 2001.

GALVÃO, R. D., MORABITO, R. Emergency service systems: The use of the hypercube queueing model in the solution of probabilistic location problems. **International Transactions in Operational Research**, v.15, p.525-549, 2008.

IANNONI, A. P., MORABITO, R. A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways. **Transportation Research**, v. 43, n.6, p.755- 771, 2007.

IANNONI A. P., MORABITO, R., SAYDAM, C. An optimization approach for ambulance location and the districting of the response segments on highways. **European Journal of Operational Research**, v.195, p.528-542, 2009.

INGOLFSSON A., AKHMETSHINA E., BUDGE S., LI Y., WU X. A Survey and Experimental Comparison of Service-Level-Approximation Methods for Nonstationary

M_t_/M/s_t_ Queueing Systems with Exhaustive Discipline. INFORMS **Journal on Computing**, v. 19, n. 2, p. 201–214, 2007.

KIM, S. H.; WHITT, W. Online Supplement to Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes? 2013.

LARSON R.C. Hypercube queuing model for facility location and redistricting in urban emergency services. **Computers and operations research**, v.1, p. 67-95, 1974.

LARSON R.C., ODONI A.R. **Urban Operations Research**. 2 ed. Dynamic Ideas, Belmont, Massachusetts. 2007.

LITTLE J.D. A proof for the queueing formula: L = $\lambda$.W. **Operations Research**, v. 9, p. 383 – 387, 1961.

MAXWELL, M. S.; RESTREPO, M.; HENDERSON, S. G.; TOPALOGLU, H. Approximate dynamic programming for ambulance redeployment. INFORMS **Journal on Computing,** v.22, n.2, p. 266-281, 2010.

SCHMID, V. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. **European Journal of Operational Research**, v. 219, n. 3, p. 611-621, 2012.

SOUZA R. M.; MORABITO R.; CHIYOSHI F. Y.; IANONNI A. P. Análise da configuração de SAMU utilizando múltiplas alternativas de localização de ambulâncias. **Gestão & Produção**, v. 20, n. 2, p. 287-302, 2013.

SIMPSON N.C.; HANCOCK P.G. Fifty years of operational research and emergency response. **Journal of the Operational Research Society**, v. 60, p. 126-139, 2009.

SMITH, J. MACGREGOR. Robustness of State-dependent Queues and Material Handling Systems. **International Journal of Production Research**, v. 48, n. 16, p. 4631 – 4663, 2010.

SWERSEY A.J. Handbooks in OR/MS. Amsterdam: Elsevier **Science B.V**., v. 6, p.151-200, 1994.

STOLLETZ R. Approximation of the non-stationary M(t)/M(t)/c(t)-queue using stationary queueing models: The stationary backlog-carryover approach. **European Journal of Operational Research**, v.190, p. 478–493, 2008.

WHITT, W. A review of $L = \lambda W$ and extensions. **Queueing Systems**, v. 9, p.235-268, 1991.

WILMER, C. BEZERRA, C. L., VIANA P**. Equações diferenciais elementares com problemas de contorno**. 3 ed. Prentice-Hall, Rio de Janeiro, 1995.