

Capítulo 7

Uma abordagem baseada em filas dependentes do tempo para tratar um sistema de atendimento emergencial: O caso do SAMU de Ouro Preto e Mariana

Arthur Piassi Dias de Castro

Lásara Fabrícia Rodrigues

Ramon Sávio Moreira Lopes

Helton Cristiano Gomes

Cristiano Luís Turbino de França e Silva

Resumo: Esse artigo analisa o sistema do SAMU de Ouro Preto e Mariana por meio de Teoria de filas. Para tanto, são propostos modelos computacionais que visam modelar as características do sistema para tornar mais simples o entendimento da realidade e facilitar as tomadas de decisão e melhorias envolvendo o objeto de estudo. Os modelos desenvolvidos utilizam-se da abordagem SIPP (*Stationary Independent Period by Period*), que trata o sistema como dependente do tempo, onde existem taxas diferentes de chegada das ocorrências para diferentes períodos do dia, resultando em picos de congestionamento e/ou ociosidade do mesmo, de maneira que seja importante a identificação destes períodos e como seus resultados implicam no funcionamento geral dos atendimentos. Os resultados dos indicadores de desempenho foram testados para todos os modelos e comparados aos observados na realidade com o objetivo de validar um tipo de modelagem que possa ser utilizada para avaliar cenários e situações críticas dentro do sistema.

Palavras chave: Pesquisa Operacional, SIPP, Teoria de Filas, Filas dependentes do tempo, SAMU, Sistemas de saúde, Modelagem, Ouro Preto, Atendimento Emergencial

1. INTRODUÇÃO

O Serviço de Atendimento Móvel de Urgência (SAMU 192) é um serviço de atendimento médico pré-hospitalar brasileiro. Esse serviço foi idealizado na França, em 1986, e nomeado *Service d'Aide Médicale d'Urgence*, daí a sigla "SAMU". No Brasil, esse serviço foi normalizado em 2004 pelo decreto presidencial nº 5.055, apesar de já estar em atuação em algumas cidades desde 1995.

Segundo Takeda et al. (2001), a função básica de um Serviço de Atendimento Móvel de Urgência (SAMU) é responder de forma organizada, a fim de evitar o uso excessivo de recursos, a toda situação de urgência que necessite de meios médicos, desde o primeiro contato telefônico até a liberação das vítimas ou seus encaminhamentos hospitalares. O sistema deve determinar e desencadear a resposta mais adequada para o caso, assegurar a disponibilidade dos meios hospitalares, determinar o tipo de transporte exigido e preparar o acolhimento dos pacientes.

Sistemas de saúde como o do SAMU são bastante sensíveis a atrasos por atenderem a várias ocorrências graves que envolvem risco de vida para quem está sendo atendido. Nesse caso, muitas das vezes, o atraso pode acarretar em agravamento do quadro clínico do paciente. McLay e Mayorga (2010) relatam a alta associação entre taxa de resposta rápida e sobrevivência dos pacientes, principalmente em ocorrências de alto nível de risco ou alta prioridade, como os tratados em sistemas de atendimento hospitalar e pré-hospitalar.

Segundo Green et al. (2001), os atrasos são frutos de uma disparidade entre uma demanda e a capacidade de serviço disponível para atender esta demanda. Normalmente isso acontece devido à uma variabilidade natural na demanda ou na própria taxa de serviço. Essa variabilidade e a interação entre a chegada e o serviço tornam a dinâmica desses sistemas bastante complexos. Portanto, para prever os níveis de recursos disponíveis para atender determinada performance podem ser usados modelo de filas.

Ainda de acordo com Green et al. (2001), modelagens utilizando Teoria de Filas normalmente são relativamente simples de formular e, comparadas com metodologias de simulação, por exemplo, requerem menos dados e menos tempo computacional. A simplicidade da utilização também torna mais direta a análise de possíveis cenários de variação na demanda e nos recursos disponíveis por meio dos indicadores de performance relacionados ao sistema.

O SAMU Inconfidentes atua nas cidades de Ouro Preto e Mariana, além de todos distritos pertencentes às duas cidades. Para isso, utiliza 4 ambulâncias, sendo 3 USBs (Unidades de Serviço Básico) e 1 USA (Unidade de Serviço Avançado), para realizar o atendimento à população. A região de Ouro Preto corresponde a uma área de 1.245,86 km² e uma população estimada de 73.994 pessoas, enquanto a de Mariana e seus distritos possui uma área de 1.194,21 km² e população de 60.142 pessoas.

O objetivo do estudo é utilizar conceitos da Teoria de Filas para modelar e analisar o funcionamento do SAMU na região dos Inconfidentes. Para isso, foram utilizados dados de ocorrências registradas no ano de 2017. Também foram analisados indicadores de performance do sistema, além de analisar a flutuação da chegada dos chamados no tempo.

O restante deste artigo está organizado da seguinte maneira. A Seção 2 apresenta uma revisão de literatura sobre Teoria de Filas e estuda alguns trabalhos que tratam de sistemas semelhantes ao do SAMU no que se diz respeito ao comportamento dos atendimentos. A Seção 3 se aprofunda em métodos de modelagem de sistemas de filas dependentes do tempo, apresentando exemplos e ressalvas sobre sua utilização. Já na Seção 4, é apresentado o Estudo de Caso, onde observa-se todos os parâmetros de funcionamento real do SAMU, ao passo que, a Seção 5 mostra os modelos propostos para a modelagem deste Estudo de Caso. A Seção 6 realiza um apanhado dos resultados obtidos pela modelagem e propõe alguns cenários alternativos com o objetivo de analisar como o sistema se comporta à situações críticas, além de comparar os indicadores de desempenho obtidos com o sistema com o objetivo de perceber se os modelos ilustram bem a realidade. Finalmente, a Seção 7 apresenta uma conclusão dos resultados obtidos e sugere caminhos para trabalhos futuros envolvendo o mesmo objeto de estudo.

2. SISTEMAS DE FILAS DEPENDENTES DO TEMPO

A teoria de filas é um ramo da probabilidade estudada dentro da Pesquisa Operacional. Seu estudo consiste em analisar a formação de filas através de cálculos matemáticos que mensuram determinadas propriedades das filas. Segundo Green et al. (2001), um sistema de filas básico é um sistema de serviços, onde "clientes" esperam algum tipo de serviço de um ou mais servidores que podem estar ou não disponíveis.

Green e Kolesar (1991) afirmam que a maioria dos sistemas de filas não apresentam taxas de chegada constantes e que é comum encontrar sistemas onde a demanda varia de acordo com o tempo de maneira cíclica, as quais podem apresentar variações diárias, semanais, mensais, etc. Isso pode ser visto em problemas envolvendo telecomunicações, bancos, sistemas policiais e emergenciais.

Os mesmos autores sugerem que qualquer sistema não estacionário pode ser tratado por meio de simulação ou utilizando-se de modelos de integração numérica, se forem Markovianos. Porém esses métodos apresentam, de maneira geral, um custo e tempo computacional altos. Portanto, vários métodos de aproximação utilizando filas foram desenvolvidos com o objetivo de simplificar e reduzir o custo de modelar esse tipo de sistema.

Segundo Schwarz et al. (2016), o estudo de sistemas de filas dependentes do tempo começou com Kolmogorov (1931) e desde então vem sendo utilizado em diferentes áreas de pesquisa, incluindo matemática, ciência da computação e pesquisa operacional. Os autores ainda ressaltam que utilizar entradas dependentes com o tempo geralmente gera um impacto grande na performance do sistema de filas, o que indica que é importante considerar essa variável ao modelar e controlar tais sistemas.

Schwarz et al. (2016) desenvolveram um esquema de classificação de sistemas de filas dependentes do tempo. O objetivo do esquema é classificar as diferentes abordagens encontradas na literatura, avaliar a performance dessas abordagens e estabelecer um paralelo entre os diferentes modelos. Os autores identificaram três principais categorias de abordagens:

- (1) Soluções numéricas ou analíticas: esse tipo de abordagem utiliza sistemas de equações para descrever o comportamento dependente do tempo de sistema de filas.
- (2) Abordagens que dividem o tempo total em intervalos menores: essa técnica modela cada período de tempo separadamente e trata os períodos como independentes, conectados ou transientes.
- (3) Abordagens baseadas em modificações nas características do sistema: utilizam-se de modificações nas propriedades do sistema, como mudanças no número de servidores ou nas características do serviço para aproximar o modelo do sistema real sem a necessidade de dividir os períodos.

Dentro das abordagens que dividem o tempo total em intervalos menores e tratam os períodos como independentes, Schwarz et al. (2016) citam o SSA (*Simple Stationary Approximation*) o PSA (*Pointwise Stationary Approximation*) e o SIPP (*Stationary Independent Period by Period*). Essas abordagens diferem entre si quanto a duração do período e determinação dos parâmetros de entrada do sistema. No SSA, os parâmetros são uma média utilizada para todo o horizonte de tempo, o PSA, determina um período $l=0$ para se utilizar de valores paramétricos instantâneos, realizando suas análises de desempenho, já o SIPP modela cada período como independente dos outros, considerando uma taxa de chegada estacionária para cada período.

3.STATIONARY INDEPENDENT PERIOD BY PERIOD (SIPP)

O SIPP é uma abordagem de resolução de um sistema de filas que utiliza-se do artifício de separar o horizonte de tempo T em intervalos pré-definidos $[a_i, b_i]$, $(i=1, 2, \dots, I)$ de maneira que sejam permitidas mudança em parâmetros de entrada do sistema em cada período. Dessa forma, pode-se aplicar um modelo estacionário de filas separadamente para cada um destes intervalos. Segundo Green (2011), a abordagem utilizada mais frequentemente na literatura é aplicar um modelo $M/M/c$ para cada intervalo de tempo, obtendo assim um número de servidores mínimos para atender ao período em questão. O SIPP vêm sendo usado por muitos autores e apresenta bons resultados para resolver sistemas de filas $M(t)/M/c(t)$ e $M(t)/M/c$ (ver, por exemplo, Green et al. (2001); Stoltetz (2008, 2011)).

Green et al. (2001) estabelece três requisitos para a aplicação do SIPP:

- (1) O tempo entre a chegada e o atendimento (*delay times*) devem ser estatisticamente independentes entre si em intervalos consecutivos;
- (2) Em cada intervalo, o sistema entra em estado estacionário;
- (3) A taxa de chegada é constante dentro de cada intervalo de tempo.

É importante ressaltar que a escolha da duração de cada período pode influenciar diretamente no desempenho do método, bem como na sua aderência ao modelo de filas. Segundo Green (2011), quanto menor a duração do intervalo de tempo definido, é mais provável que o sistema adira ao SIPP. Os autores utilizaram períodos de $\frac{1}{4}$ de hora e 2 horas e os resultados mostraram que 79% e 28% dos cenários são

confiáveis para o SIPP, respectivamente. Outro resultado interessante é que a confiabilidade do SIPP tende a se reduzir quando a taxa de chegada aumenta ou quando ela apresenta maior variação de um período para outro. Esses resultados sugerem que o SIPP seja confiável com taxas de utilização baixas e curtos intervalos de tempo.

Stolletz (2011) mostra que a modelagem por meio do SIPP pode superestimar a utilização de recursos e tamanho da fila em sistemas onde se observa alta dependência entre as taxas de chegada de períodos consecutivos. Para estes casos, sugere-se a utilização da abordagem PSA que considera uma ligação entre as taxas de diferentes períodos. Já Stolletz (2013) utiliza uma variante do SIPP para resolver ou minimizar esse problema, o *lagged SIPP*, onde uma parte da taxa de chegada do período $i-1$ é incorporada ao período i , considerando que há um atraso entre a chegada de uma ocorrência e a realização do serviço.

4. ESTUDO DE CASO

O SAMU de Ouro Preto e Mariana atua de forma integrada compartilhando os recursos dos municípios e utilizando políticas de despacho específicas dependendo do tipo e localização do chamado. A chegada das ocorrências ocorre pela discagem ao número 192, que encaminha os chamados para uma central localizada na cidade de Belo Horizonte. Durante o atendimento telefônico, a equipe decide acionar ou não uma equipe de resgate. Além disso, é registrado o local da ocorrência e acionada a base mais próxima do mesmo. Nesse atendimento, os chamados também são classificados em:

- Chamados A - Prioridade alta: Ocorrências graves que sugerem um potencial risco de vida do paciente, que incluem traumas graves ou quadros clínicos relacionados a doenças com alta probabilidade de óbito ou que requerem cuidados especializados.
- Chamados B - Prioridade baixa: Ocorrências relacionadas à traumas menos graves, pequenos e médios ferimentos ou quadros clínicos estáveis.

As ambulâncias utilizadas pelo sistema incluem uma Unidade de Serviço Avançado (USA) e três Unidades de Serviço Básico (USB). As USBs contam com equipes formadas por motorista, socorrista ou enfermeiro e um técnico. A USA atua com os mesmos profissionais da USB mais um médico especializado e é melhor equipada, contando com aparelhos de ressuscitação. As ambulâncias USBs geralmente atendem à chamados B, porém, em casos da USA estar ocupada podem ser acionadas para chamados de prioridade alta. A USA atende somente os chamados de prioridade A ou transferências. As transferências são realizadas de maneira arbitrária pelos gerentes do sistema, porém afetam o desempenho e as taxas de utilização dos recursos, portanto têm que ser consideradas na análise.

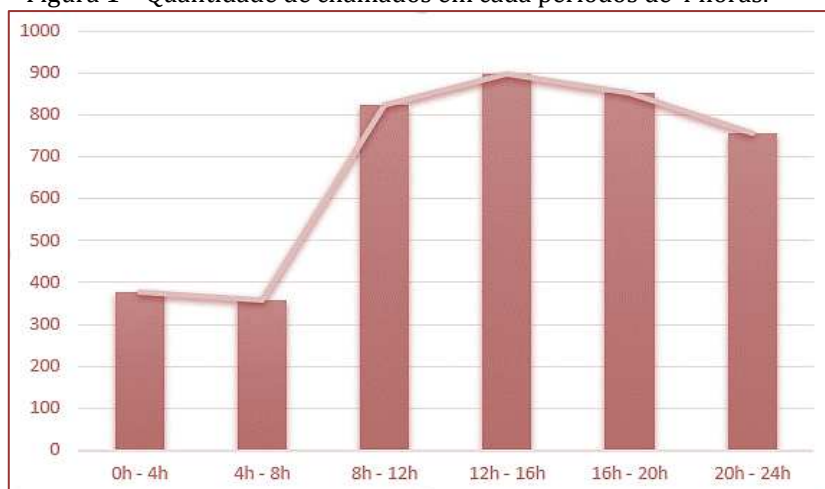
Atualmente, as USB, quando disponíveis, estão localizadas nas UPA (Unidades de Pronto Atendimento) de Mariana, Ouro Preto e Cachoeira do Campo e a USA na UPA de Ouro Preto. Ressalta-se que até setembro de 2017, a USB localizada na UPA de Cachoeira do Campo estava localizada na UPA de Ouro Preto.

A coleta e análise de dados foram autorizadas pelas Secretarias Municipais de Saúde de Ouro Preto e Mariana e pelos coordenadores do SAMU de ambos os municípios. A coleta de dados foi realizada nas sedes do SAMU, localizadas nas UPAs de Ouro Preto e Mariana. Os dados coletados correspondem aos atendimentos realizados pelas quatro ambulâncias durante o ano de 2017. Esses dados foram essencialmente coletados em prontuários de atendimentos.

No total, foram identificados 4269 prontuários, dos quais 4145 foram consideradas válidos. Desse total, 1560 correspondiam à cidade de Mariana e seus distritos, enquanto 2331 à Ouro Preto e seus distritos, além de 147 ocorrências em rodovias, 14 em outras cidades e 93 consideradas como indeterminadas.

A Figura 1 mostra a quantidade de chamados que aconteceram no sistema por intervalo de 4 horas durante o ano de 2017. Pode-se perceber claramente nesses dados uma menor quantidade de chamados nos dois primeiros intervalos do dia.

Figura 1 – Quantidade de chamados em cada períodos de 4 horas.



5. MODELO

Nesta seção, propõem-se quatro modelos utilizados para representar o sistema do SAMU. Os modelos se utilizam de modelos clássicos de filas encontrados na literatura e de algumas aproximações com o objetivo de se adequar à realidade. A modelagem foi realizada com o objetivo de se obter indicadores de desempenho para o sistema, como tempo em fila para cada classe de chamados e taxa de utilização do sistema. Todos os modelos foram desenvolvidos no *software* Rstudio.

5.1 MODELO 1

O Modelo 1 modela de maneira separada os atendimentos de alta prioridade realizados pela USA e os atendimentos de baixa prioridade realizados pelas três USBs. Esse modelo utiliza modelos de filas $M(t)/M/c$, em que as taxas de chegada e atendimento são markovianas e a quantidade de servidores para atendimento é fixa e estes são assumidos como homogêneos. Já as taxas de chegada mudam de acordo com o período do dia em que se encontram, caracterizando sistemas dependentes do tempo. Assim, as taxas de chegada (λ_a e λ_b) são tratadas como um conjunto de taxas de chegada para cada período do dia $\lambda_a = (\lambda_{a1}, \lambda_{a2}, \lambda_{a3}, \dots, \lambda_{ai})$ e $\lambda_b = (\lambda_{b1}, \lambda_{b2}, \lambda_{b3}, \dots, \lambda_{bi})$, em que i representa a quantidade de períodos.

Utilizou-se o princípio da modelagem do SIPP, que considera o sistema $M(t)/M/c$ e divide a análise do mesmo em períodos que são tratados como independentes entre si. Desta forma, o modelo analisa cada período i como um sistema $M/M/c$, onde cada taxa de chegada é considerada como estacionária.

Visto isso, para a modelagem do atendimento de chamados básicos utilizou-se o modelo $M(t)/M/3$, a taxa de atendimento média foi calculada para as três ambulâncias do sistema e a taxa de chegada foi calculada somando os atendimentos das três USBs. Para a USA, utilizou-se o modelo $M(t)/M/1$ e as taxas de atendimento e chegada foram obtidas diretamente da USA em cada período.

5.2 MODELO 2

O modelo 2 admite que as chamadas de alta prioridade devem ser atendidas por uma USB caso a USA esteja ocupada. Nesse modelo, os chamados de alta prioridade que aguardariam em fila são transferidos para serem atendidos por USBs. Para tal, continua-se tratando os modelos de alta e baixa prioridade separadamente, porém, no atendimento de chamadas de alta prioridade não é admitida fila de espera e, com isso, o modelo utilizado passa a ser um $M(t)/M/1/1$. As chamadas perdidas nesse modelo são repassadas para o sistema que atende chamadas de baixa prioridade. Essa perda é calculada pela Equação (1) e adicionada ao vetor de chegada do sistema de baixa prioridade, que continua sendo modelado como $M(t)/M/3$.

$$Perda = \frac{\lambda_a}{m * \mu_a} \quad (1)$$

onde λ_a e μ_a são as taxas de chegada e atendimento da classe de alta prioridade e m é a quantidade de servidores no sistema que neste caso é 1 por se tratar de um $M(t)/M/1/1$.

Esse modelo também utiliza o SIPP tratando os períodos como independentes entre si. Assim, utiliza o modelo $M/M/1/1$ para os chamados de alta prioridade e o modelo $M/M/3$ para os chamados de baixa prioridade e para as chamadas de alta prioridade não atendidas pelo modelo anterior.

5.3 MODELO 3

O modelo 3 considera as prioridades entre os chamados ao selecionar o próximo chamado a ser atendido. Esse modelo utiliza disciplina não preemptiva, na qual o atendimento de um chamado de menor prioridade não é interrompido ao chegar ao sistema um chamado de maior prioridade.

Para tal, foi utilizado o modelo $M(t)/M/c/NRRP$ e, em cada período, utilizando os princípios do SIPP o modelo $M/M/c/NRRP$ (ARENALES et al., 2007).

Esse modelo considera a prioridade, mas apresenta algumas limitações que o distanciam do sistema real. Primeiramente, não há como diferenciar os servidores, ou seja, as ambulâncias são assumidas como homogêneas, utilizando-se da taxa média de serviço de todas as ambulâncias. Dito isso, pode-se perceber que este modelo ilustra uma situação hipotética para esse sistema do SAMU na qual todas as ambulâncias são USAs.

5.4 MODELO 4

O modelo 4 trata as chamadas de alta prioridade independente das chamadas de baixa prioridade assim como no modelo 2. Entretanto, admite prioridade no atendimento das chamadas de alta prioridade quando atendidas por USBs assim como no Modelo 3. Assim, também utilizando os princípios do SIPP, para as chamadas de alta prioridade é utilizado o modelo $M(t)/M/1/1$ e modelo $M/M/1/1$ em cada período e a perda é calculada pela Equação (1). Esta perda é transferida para ser atendida juntamente com as chamadas de baixa prioridade, onde é utilizado o modelo $M(t)/M/c/NRRP$, em que há prioridade no atendimento das ocorrências, e o modelo $M/M/c/NRRP$ em cada período.

6.RESULTADOS

A partir dos modelos construídos na seção anterior, procurou-se utilizar os resultados para comparar os modelos e o sistema em estudo.

O tempo de resposta dos chamados é um importante indicador de desempenho de sistemas emergenciais. Visto que utilizando esses modelos de filas não é possível calcular diretamente esse tempo, analisou-se o tempo em fila, já que o tempo de resposta nada mais é que o tempo que o servidor espera em fila adicionado ao tempo de deslocamento do mesmo. Nesse caso, pode-se inferir que o tempo de deslocamento dos servidores não se altera para todos os modelos, portanto o único fator que influencia no tempo de resposta a título de comparação entre os quatro tipos de modelagem é o tempo médio em fila.

A Tabela 1 apresenta o tempo médios que um paciente espera em fila para cada tipo de modelagem, período do dia e classe de atendimento. Pode-se perceber que, no geral, os dois primeiros períodos do dia, apresentam tempos em fila baixos e os três períodos seguintes apresentam valores mais altos, sugerindo que o sistema está mais congestionado. Esses resultados estão de acordo com o esperado em função da quantidade de chamados em cada intervalo mostrada na Tabela 1.

Tabela 1 – Tempos médio em fila (minutos).

Período	Tempo médio em fila para alta prioridade				Tempo médio em fila para baixa prioridade			
	Modelo 1	Modelo 2*	Modelo 3	Modelo 4**	Modelo 1	Modelo 2	Modelo 3	Modelo 4
0-4 horas	7.4748	-	0.0096	0.0738	0.0816	0.0894	0.0102	0.0816
4-8 horas	7.4748	-	0.0078	0.0612	0.0672	0.0738	0.0084	0.0672
8-12 horas	21.7542	-	0.1794	0.6588	0.7596	0.8472	0.2222	0.8256
12-16 horas	28.8114	-	0.2382	0.7656	0.8724	0.9948	0.3006	0.9708
16-20 horas	32.0952	-	0.2022	0.6222	0.7092	0.8268	0.2526	0.7758
20-24 horas	24.4944	-	0.1242	0.4470	0.5124	0.5880	0.1506	0.543

*Classe não aceita filas no modelo.

**Resultados referentes aos chamados perdidos adicionados no modelo.

Em relação ao tempo em fila dos chamados de alta prioridade, os resultados apresentaram diferenças significativas entre os modelos. O modelo 1 apresenta resultados muito superiores aos demais modelos. Esse modelo utiliza apenas uma USA a qual apresenta tempo de atendimento alto. O modelo 2 não admite fila de espera e os chamados não atendidos são encaminhados para as USBs, porém esse atendimento é realizado sem diferenciar as classes dos chamados. O modelo 3 assume que existem quatro servidores homogêneos, ou seja, todos os servidores podem atender chamados dessa classe, além de prioridade entre as classes. O modelo 4 teoricamente se aproxima mais da realidade do sistema em estudo. Vale destacar que é preciso refinar os modelos para calcular o tempo em fila de chamados de alta prioridade.

Já para os chamados de baixa prioridade, o modelo 4 mostra tempos de fila próximos aos modelos 1 e 2. Pelos mesmos motivos apresentados para os chamados de alta prioridade, os resultados do Modelo 3 se distanciam dos demais modelos.

Ao comparar-se os resultados de tempo em fila com os tempos de resposta do sistema, percebeu-se que, em geral, nos chamados de alta prioridade, o tempo em fila representa uma menor parcela do tempo de resposta, variando entre 0,37% em um período mais ocioso e 4,13% no período mais congestionado. Já na classe de baixa prioridade, esta mesma parcela variou de 0,42% até 6,82%. Os resultados sugerem que a fila tem mais impacto nos chamados de baixa prioridade, principalmente em períodos de pico.

Outra medida de desempenho para o SAMU é a taxa de utilização dos servidores. A Tabela 2 apresenta um comparativo entre as taxas de utilização para cada tipo de modelagem proposta e período do dia, além taxa de utilização amostral, calculada pela amostragem de ocorrências. Os resultados confirmam os resultados obtidos para o tempo em fila, ou seja, há maior ociosidade nos dois primeiros períodos do dia seguida de um pico nos três períodos seguintes.

Percebe-se também que os resultados, principalmente dos Modelos 2 e 4 se aproximaram da taxa obtida na amostra. O Modelo 4 apresentou uma diferença média de 5,21% entre a taxa de utilização encontrada e a calculada com os dados amostrais. Esses resultados sugerem que esse modelo pode ser utilizado para representar o sistema estudado, uma vez que considera mais características do sistema real em sua modelagem.

Tabela 2 – Taxas de utilização dos recursos.

Período	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Amostra
0-4 horas	0.1529	0.1414	0.0871	0.1407	0.1302
4-8 horas	0.1469	0.1354	0.0821	0.1347	0.1365
8-12 horas	0.3570	0.3265	0.1924	0.3249	0.3164
12-16 horas	0.4072	0.3691	0.2073	0.3673	0.3658
16-20 horas	0.4102	0.3686	0.1983	0.3669	0.3229
20-24 horas	0.3474	0.3139	0.1739	0.3125	0.2896

7.CONCLUSÃO

Com base nos resultados obtidos nesse estudo, pode-se concluir que é possível modelar o sistema SAMU por meio de Teoria de Filas. Entretanto, deve-se atentar para os pressupostos e simplificações de cada modelo frente ao sistema analisado.

Dentre os modelos propostos, o modelo 4 apresentou os melhores resultados em comparação com os demais modelos. Esse modelo é que mais se aproxima das características do sistema estudado, dentre as quais pode-se citar: a USA atende apenas chamados de alta prioridade e as USBs funcionam como *backup* da USA, ou seja, são acionados quando a USA está ocupada; os atendimentos a chamados de alta prioridade realizados pela USB são prioritários em relação aos de baixa prioridade; e a modelagem trata USA e USBs como heterogêneos e independentes. Esses resultados também sugerem que esse modelo pode ser utilizado futuramente para analisar cenários e propor melhorias de maneira geral.

Para trabalhos futuros, pretende-se comparar os resultados obtidos com modelos de simulação a serem desenvolvidos, refinar a modelagem por meio do SIPP, introduzindo o conceito do SIPP PRI (*Priority Stationary Period by Period*), afim de refinar os modelos desenvolvidos de forma que se aproximem mais das características reais do sistema.

8.AGRADECIMENTOS

Os autores gostariam de agradecer a FAPEMIG pelo apoio financeiro.

REFERÊNCIAS

- [1] Arenales, Marcos, Armentano, Vinícius, Morabito, Reinaldo, Yanasse, Horácio. Pesquisa Operacional, Rio de Janeiro, Editora Campus, 2007
- [2] Decreto Nº 5.055, de 27 De Abril de 2004. Palácio do Planalto. Acesso em 19 de outubro de 2015. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Decreto/D5055.htm
- [3] Green, Linda V., Kolesar, Peter J., Soares, João. Improving the Sipp Approach for Staffing Service Systems That Have Cyclic Demands. Operations Research, p. 549-564, 2001. Disponível em: <http://dx.doi.org/10.1287/opre.49.4.549.11228>.
- [4] Green, Linda V. Queueing theory and modeling. Graduate School of Business, Columbia University, New York, New York, 16p., 2011.
- [5] Green, Linda V., Kolesar, Peter J. The pointwise stationary approximation for queues with nonstationary arrivals. Management Science, v. 37, n. 1, 1991.
- [6] Mclay, Laura A., Mayorga, Maria E. Evaluating emergency medical service performance measures, Health Care Management Science, v. 13, n. 2, p. 124-136, 2009.
- [7] Schwarz, Justus Arne, Selinka, Gregor, Stolltritz, Raik. Performance analysis of time-dependent queueing systems: Survey and classification. Omega. p. 170-185, 2016.

- [8] Stolletz, Raik. 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*. p. 478–493, 2008.
- [9] Stolletz, Raik, Analysis of passenger queues at airport terminals. *Research in Transportation Business & Management*, v. 1, p. 144–149, 2011.
- [10] Stolletz, Raik, Lagershausen, Svenja. Time-dependent performance evaluation for loss-waiting queues with arbitrary distributions. *International Journal of Production Research*. p. 1366-1378, 2013.
- [11] Takeda, Renata Algisi, Widmer, João Alexandre, MORABITO, Reinaldo. Uma proposta alternativa para avaliação do desempenho de sistemas de transporte emergencial de saúde brasileiros. *Transportes*, v. 9, n. 2, p. 9-27, 2001.
- [12] Vile, Julie Leanne, Gillard, Jonathan William, Harper, Paul Robert, Knight, Vincent Anthony. Time-dependent stochastic methods for managing and scheduling Emergency Medical Services. *Operations Research for health care*. p. 42-52, 2016.
- [13] Vile, julie leanne, Gillard, Jonathan William, Harper, Paul Robert, Knight, Vincent Anthony. A queueing theoretic approach to limit excessive waiting times in time-dependent dual-class service systems, 2015, submetido a publicação. Disponível em: <http://www.julievile.co.uk/publications.html>.