

ANÁLISE DO SERVIÇO DE ATENDIMENTO MÓVEL DE URGÊNCIA (SAMU) DE BELO HORIZONTE VIA SIMULAÇÃO E OTIMIZAÇÃO

Pedro Marinho Sizenando Silva

Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte - Brasil
pedromarinhosilva@gmail.com

Luiz Ricardo Pinto

Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte - Brasil
luiz@dep.ufmg.br

RESUMO

Diversos trabalhos publicados na literatura médica e na área de pesquisa operacional demonstram que existe uma relação direta entre o tempo de resposta das ambulâncias e a probabilidade de sobrevivência das vítimas envolvidas em acidentes. Apresenta-se, neste trabalho, o desenvolvimento de um modelo de simulação representativo das operações do Serviço de Atendimento Móvel de Urgência de Belo Horizonte. Com o modelo verificado e validado, apresentam-se estimativas para o tempo de resposta médio do sistema. Alguns cenários foram avaliados para determinação dos ganhos e eventuais perdas nas diversas etapas do processo de resgate decorrentes do aumento na demanda e de mudanças nas lógicas de operação. Um estudo também foi conduzido integrando o modelo de simulação com um otimizador, de maneira a identificar a melhor configuração do sistema, levando em consideração restrições relativas ao tempo de resposta.

PALAVRAS CHAVE. SAMU. Simulação. Otimização. Simulação.

ABSTRACT

Several studies published in the medical literature and in the area of operational research shows that there is a direct relation between response time of rescue units and the probability of survival for victims involved in accidents. This study presents the development of a simulation model representing the operations of the Emergency Medical System of Belo Horizonte. The average response time of the system is estimated with the model verified and validated. Some scenarios were evaluated to determine the gains and losses in the various stages of the rescue process related to service demand growth and changes in the operations procedures. A study was also conducted by integrating the simulation model with an optimizer, in order to identify the best configuration of the system, taking into consideration restrictions concerning the response time.

KEYWORDS. SAMU. Simulation. Optimization. Simulation.

1. Introdução

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 aos hospitais. 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 (TAKEDA *et al.*, 2001).

Serviços emergenciais como os SAMU's apresentam altos graus de incerteza, e normalmente suas eficiências são medidas através do tempo médio de resposta a um chamado, ou seja, o tempo que uma vítima espera em média para começar a receber algum tipo de atendimento. Quanto maior o grau de incerteza envolvido e maior a necessidade de se obter respostas rápidas, menor deve ser a taxa de utilização dos operadores e equipamentos do sistema. Caso contrário, o nível de serviço oferecido pode ser deteriorado. Neste contexto, quando bem dimensionados, geralmente ocorrem longos períodos em que os operadores e equipamentos permanecem desocupados (GONÇALVES, 1994).

O serviço de atendimento móvel de urgência no Brasil segue o modelo francês de tratamento pré-hospitalar, que presta socorro às vítimas no local do acidente, não se atendo somente ao transporte da vítima para o hospital. Lançado em setembro de 2003 pelo Governo Federal, está disponível 24 horas por dia, sendo constituído por equipes de profissionais da saúde que incluem médicos, enfermeiras e assistentes, que respondem a emergências de diferentes naturezas, incluindo trauma, pediátrica, cirúrgica, ginecológica, obstétrica e saúde mental (TIMERMAN *et al.*, 2006).

O problema da análise de serviços emergenciais pode ser abordado de diferentes maneiras: localização de bases (Toregas *et al.*, 1971), projeto das áreas de cobertura específicas (Church e ReVelle, 1974), simulação (Shih e Su, 2003), modelo hipercubo (Oliveira e Gonçalves, 2004), dentre outros. O presente trabalho utiliza das técnicas de simulação a eventos discretos para condução do estudo de análise do serviço de atendimento móvel de urgência de Belo Horizonte – MG.

2. Revisão de Literatura

A utilização das técnicas de simulação na área de saúde como ferramenta para análise e melhoria de processos tem demonstrado adequação em diferentes aplicações envolvendo avaliação de projetos, dimensionamento de recursos, revisão de procedimentos e protocolos de atendimento a pacientes, avaliação da capacidade instalada e políticas operacionais.

De acordo com Brailsford (2007), modelos de simulação na área de saúde podem ser classificados em três grupos: modelos do corpo humano, representativos de processos biológicos em indivíduos saudáveis ou análise clínica da eficiência de alguma intervenção/tratamento; modelos operacionais ou táticos, cujo foco está na modelagem dos fluxos de pacientes pelo sistema e identificação dos gargalos; e modelos estratégicos, focados em responder questões a longo prazo relativas ao planejamento estratégico da organização (configuração do sistema de atendimento, projeções com relação à quantidade de funcionários, etc.).

Os primeiros esforços de aplicação dos conceitos de simulação na área de saúde remontam à década de 60, com o trabalho de Fetter e Thompson (1966), que desenvolveram modelos para solucionar problemas relacionados ao agendamento de consultas de pacientes. Foram capazes de identificar variáveis importantes na avaliação da utilização do quadro médico, tais como taxa de chegada de pacientes, chegada antecipada ou tardia, adiamentos, intervalos de agendamento e paradas devido à troca de turnos dos médicos.

Um dos primeiros trabalhos envolvendo a aplicação de técnicas de simulação para modelagem e análise de sistemas de atendimento móvel de urgência foi desenvolvido na década

de 70 por Fitzsimmons (1971). Um modelo de simulação foi implementado utilizando a linguagem SIMSCRIPT para uso geral por parte de gestores de sistemas de saúde, de maneira a auxiliar a avaliação de sistemas médicos de emergência existentes ou em fase de planejamento.

Shih e Su (2003) desenvolveram um modelo de simulação a eventos discretos para avaliar o sistema de atendimento médico de emergência em operação na cidade de Taipei, Taiwan. O estudo focou em 23 hospitais habilitados para atendimento de emergência e 36 centros de atendimento. O sistema na época do estudo atuava em conjunto com o corpo de bombeiros, que fornecia o atendimento médico básico (no caso de ocorrências menos graves) através do envio de uma unidade básica de atendimento (BLS). Com o modelo desenvolvido foi possível identificar e propor melhorias que resultaram na diminuição do atraso entre a chegada de uma chamada e o despacho de uma ambulância em 50% e na diminuição da probabilidade de pacientes terem de esperar pelo despacho do resgate para menos de 1%.

Diferentes políticas de operação de sistemas de atendimento podem ser avaliadas com o auxílio das técnicas de simulação. Koch e Weigl (2003) apresentam um modelo baseado em simulação de eventos discretos para estudo da logística de transporte do serviço de resgate da Cruz Vermelha na Áustria, comparando políticas de centralização e descentralização das unidades de resgate. A rede de tráfego representativa do modelo foi implementada em SIMAN pela possibilidade da utilização de algoritmos de grafos integrados com a linguagem para resolver problemas de caminho mínimo. Os resultados mostraram que existe potencial para redução do tempo de espera para os pacientes e da distância percorrida se a organização for centralizada.

Uma abordagem interessante para análise de sistemas emergenciais é a combinação de modelos de localização de facilidades e técnicas de simulação para análise das soluções geradas. Aringhieri *et al.* (2007) desenvolveram um modelo de programação linear inteira para localização de unidades de atendimento médico e um procedimento de simulação para teste das soluções, tendo como base a cidade de Milano, Itália.

Maxwell, Henderson and Topaloglu (2009) modelaram o problema de distribuição de ambulâncias como um processo de decisão de Markov, de maneira a avaliar a performance de uma dada política de alocação. Um modelo de simulação foi formulado e integrado com um método aproximado de programação dinâmica para computar políticas de distribuição consideradas de alta qualidade, apresentando resultados para a cidade de Edmonton, no Canadá.

3. Samu de Belo Horizonte

De acordo com dados recentes retirados do site do IBGE – Instituto Brasileiro de Geografia e Estatística, a cidade de Belo Horizonte (BH) possui uma área de 330.954 km² e uma população estimada de 2.412.937 habitantes (ano 2007). A cidade está dividida em nove regionais, conforme apresentado na figura 1.

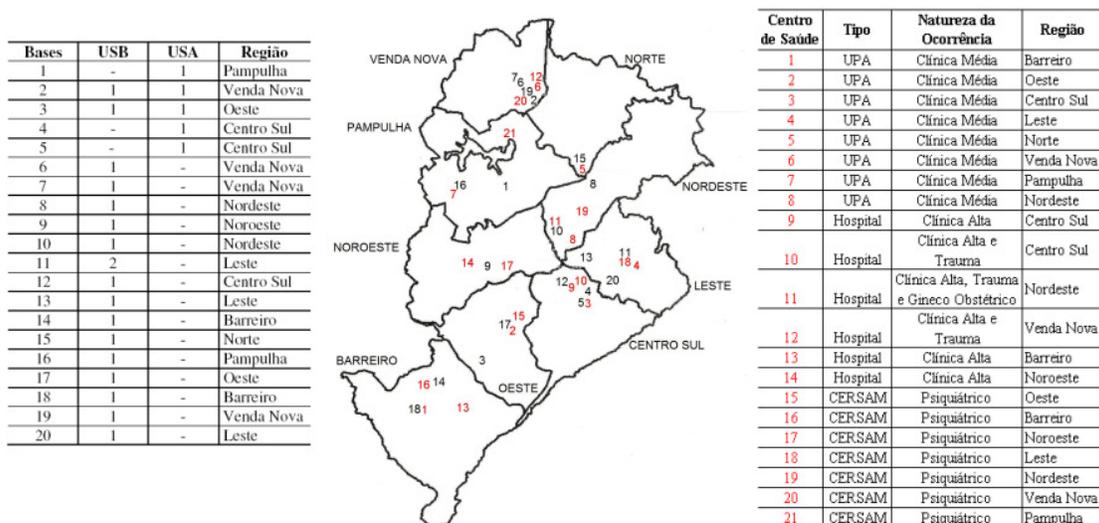
O serviço de atendimento móvel de urgência de BH opera 24 horas por dia e conta com uma base central de regulação além de outras 19 bases espalhadas pela cidade, todas com um número fixo de ambulâncias (figura 1).

Ao todo são 17 ambulâncias básicas (USB) e 5 avançadas (USA), sendo que a ambulância na base 20 é exclusiva para atendimento de casos psiquiátricos mais graves (ambulância de saúde mental). As unidades básicas são acionadas para atender casos de gravidade leve e média ou então em auxílio à uma unidade avançada, por estar mais próxima do local do incidente, sendo composta por motorista socorrista e auxiliar de enfermagem. As unidades avançadas só são empenhadas em casos realmente graves, por serem em menor quantidade, compostas por motorista, enfermeiro e médico.

A central de regulação (base 5) é onde acontece toda a triagem e atendimento das chamadas. Atualmente, conta com 6 teledigifonistas (atendentes que fazem o primeiro contato com o solicitante), 5 médicos reguladores (decidem sobre o envio ou não de ambulância, podendo passar orientações médicas pelo telefone caso não seja necessário o envio) e 3

despachantes (responsáveis pelo despacho da ambulância uma vez solicitado pelo médico regulador bem como o controle da localização e situação exatas de cada unidade).

Figura 1 – Distribuição espacial das bases e centros de saúde de Belo Horizonte



Com relação aos centros de saúde, o Samu trabalha em conjunto com 8 Unidades de Pronto Atendimento (UPA), 6 hospitais e 8 Centros de Referência em Saúde Mental (CERSAM). A decisão da escolha do centro de atendimento é baseada no tipo de ocorrência (clínico – média ou alta, trauma, psiquiátrico ou gineco-obstétrico), sendo feita sempre pelo médico regulador da central, nunca pela equipe da ambulância que realiza o atendimento. A relação dos centros e tipos de ocorrência que recebem pode ser visualizada na figura 1.

4. Modelo Computacional

O modelo computacional foi desenvolvido no software Arena 11.00 da Rockwell Automation Technologies. A figura 2 apresenta um fluxograma básico de seu funcionamento.

O processo começa com a chegada de uma chamada na central de atendimentos que é recebida por um dos atendentes (teledigifonistas). Antes de iniciar o atendimento na central de regulação, determina-se a região de origem da chamada, seu tipo, conduta médica e as coordenadas UTM (*Universal Transverse Mercator*). O atendente faz a análise inicial da chamada buscando identificar sua natureza (trote, solicitação de informações, solicitação de atendimento ou engano). Em caso de solicitação de atendimento, são coletadas informações iniciais como local do acidente e alguns dados da vítima, e a chamada é encaminhada para o médico regulador para avaliar a gravidade da situação. Neste momento, pode ser necessário o envio de uma ambulância ou então a ligação é encerrada sem a necessidade de atendimento local, sendo feita somente a orientação pelo telefone. No primeiro caso, a ambulância mais próxima (considerando sua situação corrente e o tempo de deslocamento) é acionada por um dos despachantes e inicia o deslocamento até o local do incidente. Algumas chamadas são canceladas após a chegada da ambulância ao local devido a ocorrência de trotes. Uma vez no local, inicia-se o atendimento médico após o contato com a vítima e em casos de maior gravidade, esta é encaminhada para o hospital mais próximo de acordo com sua natureza e levando em conta a proximidade dos centros de atendimento. A escolha do centro é feita sempre pelo médico regulador na central, nunca pela equipe que está realizando o atendimento.

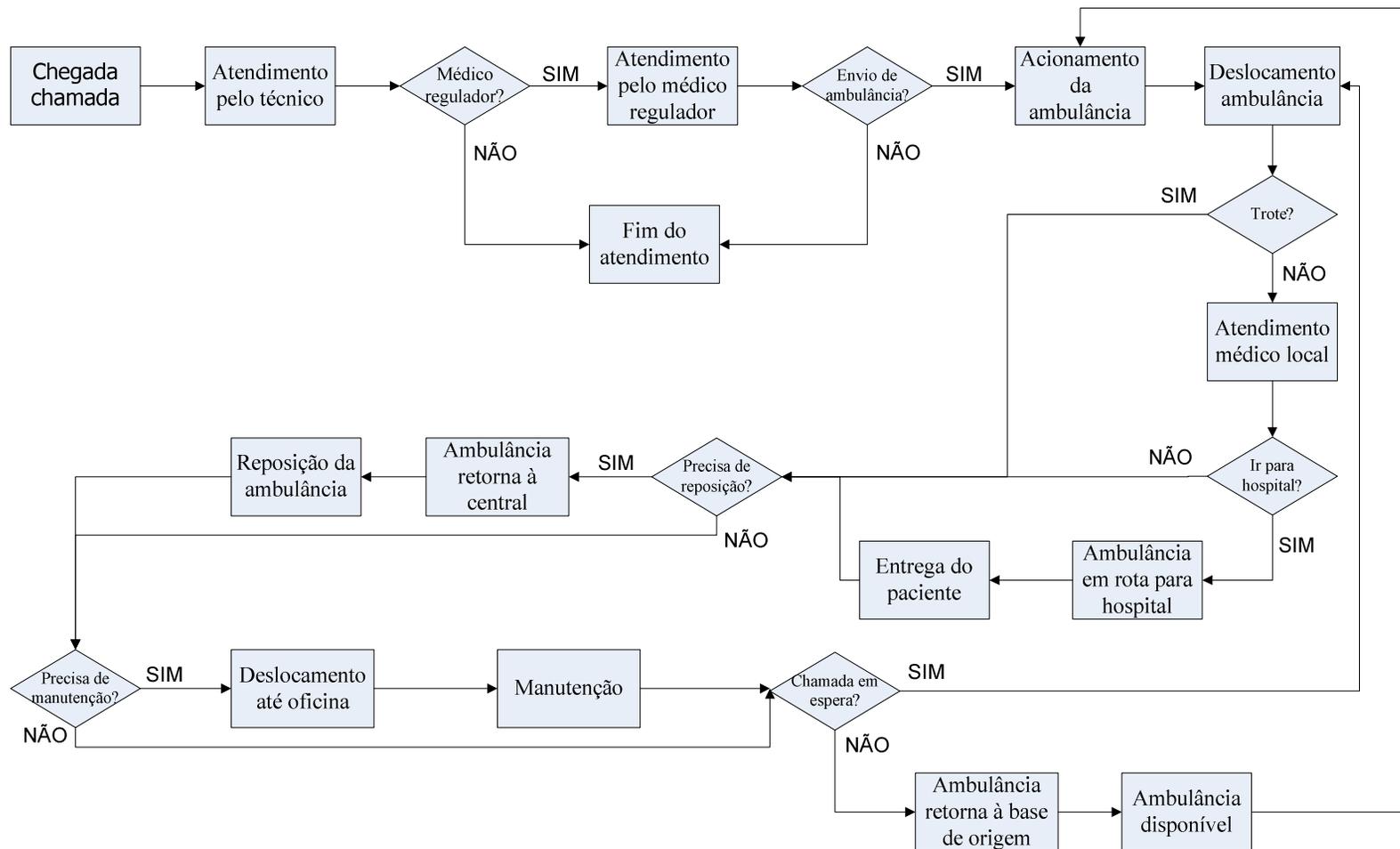


Figura 2 - Fluxograma do processo de atendimento de uma chamada de emergência

Após deslocamento da ambulância, entrega do paciente e preenchimento dos relatórios requeridos, a equipe verifica a necessidade de preparação, de maneira a repor equipamentos utilizados ou até mesmo avariados durante o processo de atendimento. Existe também a possibilidade de a ambulância necessitar de manutenção devido a avarias sofridas, por exemplo, durante os trajetos até o incidente e até o hospital designado. Para o modelo desenvolvido, a manutenção nunca acontece durante o processo de resgate, mas sempre ao final do ciclo (atendimento + hospital + reposição). Caso exista alguma chamada em espera após o processo de atendimento e eventuais reposição e manutenção, a ambulância dirige-se diretamente até o local do incidente, dado que a mesma foi escolhida para realizar o atendimento. Caso contrário, a ambulância retorna à base de origem (ponto zero) ficando então disponível para novo atendimento.

Para análise das características das chamadas e processos de operação da central de regulação, utilizou-se o banco de dados do Samu com informações do período de outubro de 2006 até agosto de 2009, contendo todos os dados relativos à descrição da chamada (quem atendeu, horários, região de origem, tipo da chamada, procedimentos adotados, ambulância despachada, etc.). Ao todo foram registradas 1.734.416 chamadas neste período. De todas as chamadas que chegaram à central de regulação do Samu no ano de 2009, apenas 31.93% resultaram em atendimento pelo médico regulador (as demais podem ser divididas em trote, abuso, engano, chamada interna, informações ou ligação interrompida).

Os tempos das diversas operações do serviço de atendimento móvel tiveram de ser estimados a partir de dados coletados diretamente na central de regulação. Para a atividade de reposição, utilizou-se o padrão de uma reposição a cada 24 horas para as unidades básicas e duas reposições diárias para as unidades avançadas. Para a atividade de manutenção, utilizou-se uma taxa média de 3 manutenções por dia, sendo que a escolha da ambulância é feita de maneira aleatória. Este dado foi estimado pela equipe do SAMU.

As duas atividades que tiveram o tempo de serviço estimado com base na percepção dos funcionários envolvidos foram as de reposição e entrega do paciente no hospital, sendo utilizada, para estes casos, a distribuição triangular.

Para determinação das coordenadas dos incidentes, a cidade de Belo Horizonte e suas nove regionais foram mapeadas através do sistema UTM de coordenadas. As coordenadas UTM também foram utilizadas para determinar a localização exata das bases e centros de atendimento, além de facilitar o cálculo do tempo de deslocamento, baseado na distância euclidiana entre dois pontos. O contorno de cada regional foi aproximado por um polígono, buscando considerar da melhor maneira possível os limites da região. Para sorteio das coordenadas do incidente, utilizou-se os valores máximo e mínimo das coordenadas UTM do eixo horizontal e do eixo vertical, amostrando valores dentro do retângulo formado por estes extremos com igual probabilidade. Uma rotina em VBA foi implementada para verificar se o ponto sorteado se encontra dentro do polígono traçado para representar aquela região. Caso não esteja, um novo ponto é sorteado e verificado até que a condição seja satisfeita.

De maneira a controlar a alocação das ambulâncias, foi criada uma matriz de controle com o número de linhas igual à quantidade total de ambulâncias, armazenando em suas células informações em tempo real referentes a:

- hora que em a ambulância estará disponível;
- coordenadas UTM do local em que a ambulância estará disponível;
- número de chamadas em espera pela ambulância;
- tipo de ambulância.

A distância euclidiana, na maioria das aplicações reais de transportes, constitui-se em uma abstração útil para os cálculos e estruturação dos modelos. A razão reside na sua simplicidade de representação analítica e na sua característica de unicidade (isto é, há somente uma ligação euclidiana entre dois pontos). Através de coeficientes corretivos médios pode-se relacionar matematicamente as distâncias efetivas com as distâncias euclidianas, possibilitando assim o tratamento mais realista das aplicações (NOVAES, 1989).

O cálculo do tempo de deslocamento T_d de uma ambulância leva em consideração a

distância euclidiana ponderada pelo fator de correção e a velocidade média da ambulância, de acordo com a relação:

$$T_d = D * F / V$$

onde D é a distância euclidiana, F é o fator de correção e V a velocidade média.

O cálculo da velocidade média de uma ambulância é influenciado pela intensidade do tráfego, que por sua vez é variável e dependente do tempo, resultando em um valor médio para cada período de tempo considerado ao longo do dia. Uma vez acionada, o sistema determina à priori o tempo de serviço e as coordenadas UTM em que a ambulância estará disponível. Após o término do atendimento, o sistema determina, baseado em dados históricos, a necessidade de reposição e manutenção para a unidade.

Desta maneira, quando uma nova chamada chega, os tempos de liberação de cada ambulância estarão atualizados, mesmo que esta esteja em atendimento. Quando o processo de atendimento termina, o sistema verifica se existe alguma chamada em espera por aquela unidade específica para decidir seu destino. Se não existe chamada, a ambulância retorna para a base e atualiza o tempo e coordenadas de liberação. O procedimento de despacho verifica, em primeiro lugar, se o tipo requerido de ambulância está disponível. Desta maneira, a escolha de uma unidade específica é baseada no menor tempo de resposta. Se o tempo de liberação da ambulância for maior do que o tempo de simulação corrente, significa que a ambulância está em atendimento. O tempo de resposta, para este caso, é calculado como apresentado a seguir:

$$T = T_r - T_c + T_d$$

onde T é o tempo de resposta, T_r é o tempo de liberação, T_c é o tempo corrente de simulação.

No caso de estar disponível, o tempo de resposta da ambulância fica restrito ao tempo de deslocamento T_d .

Para o caso da cidade de Belo Horizonte, o procedimento para estimativa do fator de correção baseou-se no levantamento das coordenadas UTM de alguns pontos espalhados ao longo das nove regionais e no trabalho de Novaes (1989). Foi possível estabelecer uma relação aproximada para a estimativa do fator de correção, sendo neste caso considerado igual a 1.366.

Para a velocidade média foram utilizados valores de 25 km/h para o período de 7:00 às 21:00 e 50 km/h para o período de 21:00 às 7:00.

4.1 Validação do Modelo

O processo de validação foi conduzido desde as etapas iniciais do projeto, com a participação direta da gerência do Samu. Todo o processo de funcionamento do sistema foi discutido com os responsáveis (modelagem conceitual) bem como as simplificações adotadas para a modelagem computacional. Além da participação e acompanhamento de especialista, foram utilizados dados reais coletados em campo para validação dos resultados gerados pelo modelo desenvolvido, tendo como base os parâmetros atuais de operação do sistema. Foram utilizados três parâmetros para esta etapa do processo de validação: tempo de espera que uma chamada gasta entre a verbalização da necessidade de envio de uma ambulância até o empenho efetivo da mesma, o tempo de espera acrescido dos tempos de deslocamento e tempo de atendimento local e a quantidade de chamadas atendidas discretizadas por mês. Para os dois primeiros, foram coletados 100 dados distribuídos ao longo dos cinco dias úteis e do período de 07h00min até 21h00min em cada dia (tabelas 1 e 2).

Tabela 1 – Tempo de espera e tempo total

	Tempo Espera [min]		Tempo Total [min]	
	Simulação	Real	Simulação	Real
Mínimo	0,50	0,33	8,37	9,14
Médio	6,07	6,13	35,35	31,82
Máximo	52,66	48,91	96,07	79,75

Tabela 2 – Número de ambulâncias despachadas

Mês	Real	Simulado	Diferença
dez/08	7.437	7.453	-0,21%
jan/09	7.266	7.470	-2,81%
fev/09	6.752	6.722	0,44%
mar/09	7.503	7.458	0,60%
abr/09	7.281	7.196	1,17%
mai/09	7.528	7.416	1,49%
jun/09	6.934	7.140	-2,97%
jul/09	7.417	7.508	-1,23%
ago/09	7.429	7.388	0,55%

4.2 Cenário Atual de Operação

Para o cenário atual e testes apresentados a seguir, o modelo computacional desenvolvido foi rodado com 10 replicações com um período de aquecimento de 100 horas e duração da simulação de 7 dias. Os intervalos de confiança foram de 95%.

Os dois parâmetros analisados foram o tempo de resposta (tempo entre chegada da chamada na central e chegada da ambulância no local do incidente) e tempo de espera (tempo que uma chamada fica em espera até ser alocada a uma ambulância, dado que já foi atendida pelo médico regulador na central). Os resultados estão apresentados na tabela 3 a seguir.

Tabela 3 – Tempo de resposta e tempo de espera para cenário atual

	Tempo de Resposta [min]	Tempo de Espera [min]
Mínimo	4,34	0,50
Médio	21,21	4,96
Máximo	78,50	52,66

O tempo de resposta mínimo diz respeito aos incidentes que ocorrem no mesmo local ou bastante próximos de alguma base de operação do Samu, resultando em tempo de deslocamento nulo ou muito pequeno. O tempo de resposta máximo é justificado pelas chamadas que apresentam um elevado tempo de espera por serem de baixa prioridade. O intervalo de confiança para o tempo de resposta médio é apresentado a seguir: I.C. = [20.72 ; 21.71] minutos.

Para o tempo de espera, o valor mínimo se refere ao envio imediato da ambulância, não tendo de esperar nenhuma outra atividade terminar, sendo composto somente pelo tempo gasto pelos despachantes. O tempo máximo diz respeito à priorização de chamadas mais importantes em horários críticos, conforme apresentado. O intervalo de confiança para o tempo de espera médio calculado é: I.C. = [4.51 ; 5.42] minutos.

As figuras 3 e 4 apresentam a distribuição dos dois parâmetros analisados. Com relação à distribuição dos valores para o tempo de resposta considerando as faixas apresentadas, pode-se observar que a maior quantidade de valores está entre 10 e 20 minutos, seguido pela faixa entre 20 e 30 minutos. Os valores extremos são de baixa frequência de ocorrência. Para a distribuição dos tempos de espera, a grande maioria dos valores se encontra na faixa entre zero e cinco minutos (72,96%), sendo que valores muito altos são de frequência relativa bastante baixa (entre 30 e 40 minutos, somente 1,23% das chamadas).

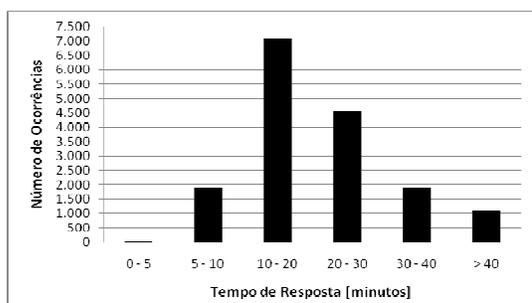


Figura 3 – Distribuição do tempo de resposta (cenário atual)

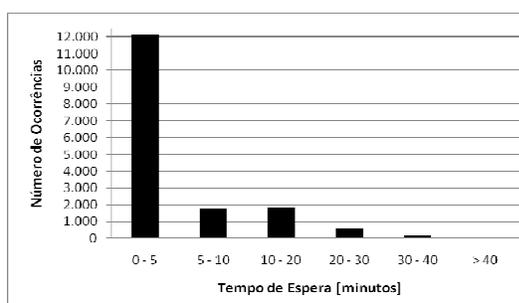


Figura 4 – Distribuição do tempo de espera (cenário atual)

Com relação à taxa de utilização das ambulâncias, das 10 replicações realizadas foi possível obter uma média de 55.32% para as unidades básicas e 11.75% para as unidades avançadas. Gonçalves (1994) afirma que quando os sistemas de atendimento móveis de urgência estão bem dimensionados, geralmente ocorrem longos períodos em que os operadores e equipamentos permanecem desocupados. No trabalho de Takeda (2000), o índice de ocupação dos veículos encontrado de 60% foi considerado elevado para este tipo de sistema. De acordo com Fitzsimmons e Fitzsimmons (2004), uma ambulância do serviço de emergência raramente está ocupada por mais de 30% de seu tempo. Essa baixa utilização, no entanto, é indispensável para permitir o atendimento imediato. O excesso de capacidade da ambulância é, assim, necessário, pois o custo implícito da espera para esse serviço pode ser exorbitante em termos de vidas humanas.

5. Análise de Cenários

5.1 Aumento de Demanda

O primeiro cenário buscou avaliar o comportamento do sistema frente ao aumento na demanda pelos serviços (figura 5). Foram analisadas as variáveis tempo de resposta médio, tamanho médio da fila de espera (número médio de chamadas que estão esperando alocação à alguma ambulância), tamanho máximo da fila de espera e tempo de espera médio. O tempo de resposta médio sofre um leve crescimento até o aumento de 20% na demanda por serviços, quando então começa a crescer de maneira mais acentuada. Com o crescimento de 30%, o tempo de resposta médio aumenta em mais de 17 minutos, sendo que para 40% de aumento este valor mais do que dobra quando comparado ao cenário atual. O tempo de espera médio mais do que dobra com um aumento de 20% sobre a demanda, impactando no cálculo apresentado para o tempo de resposta médio. Para um aumento de 40%, a variável analisada já apresenta valores médios próximos de meia hora, totalmente incompatíveis com parâmetros de desempenho para este tipo de sistema. O tamanho de fila pode ser um bom indicador da relação entre o desempenho das operações dentro da central de regulação e as operações envolvendo especificamente uma unidade móvel. O cenário de aumento de 30% na demanda, indicando um tamanho médio de fila de 4,09, representa ou um sistema operando em condições limites durante as 24 horas de operação ou então um sistema bastante congestionado. O tamanho máximo da fila de espera foi utilizado para aumentar o entendimento do comportamento do sistema e aumentar a confiança na definição da qualidade dos parâmetros de resposta estimados frente ao aumento da demanda. Pode acontecer de um valor médio para o tamanho da fila de espera ser considerado aceitável, mas estar mascarando valores máximos (ocorrendo, por exemplo, em horários de pico) que são insustentáveis para a operação regular do sistema.

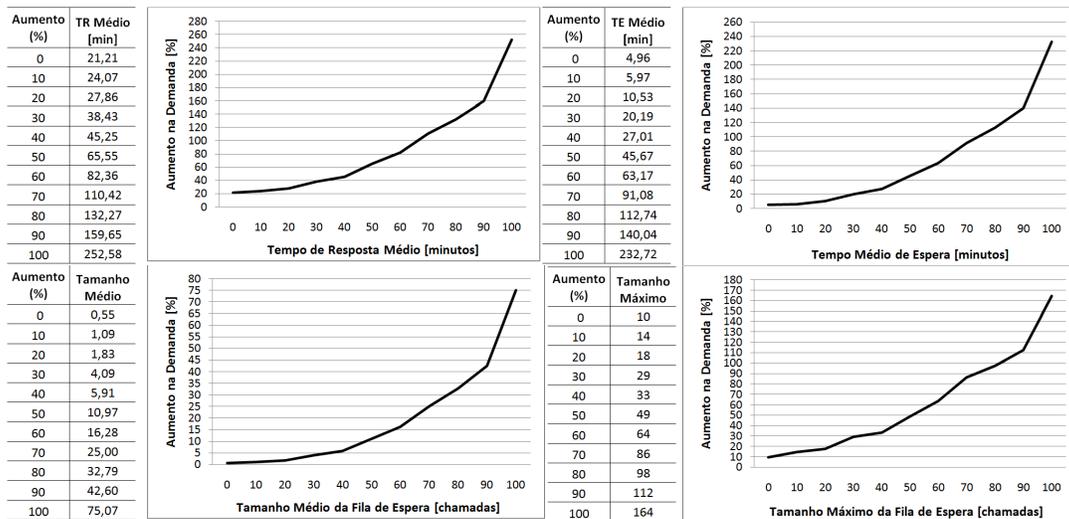


Figura 5 – Resultados para o cenário de aumento de demanda

5.2 Reposição nos Hospitais

O segundo cenário considerou a atividade de reposição podendo ser feita em qualquer um dos centros de atendimento existentes, motivado por trabalhos semelhantes na literatura. Após a entrega do paciente, o modelo verifica a necessidade ou não de reposição, de acordo com o tempo decorrido da última vez em que a unidade repôs equipamento. Se acontecer de uma necessidade por reposição não coincidir com o momento de entrega do paciente, a ambulância retorna para a central de regulação, independente de sua posição, não utilizando nenhum dos centros de atendimento considerados. O tempo de reposição permaneceu o mesmo, tanto para a central quando para os hospitais e unidades de pronto atendimento. O teste *Intervalo de Confiança para a Diferença das Médias com Amostras Pareadas* foi utilizado para identificar possíveis ganhos na variável tempo de resposta quando comparada com o cenário atual. O intervalo de confiança para a média das diferenças foi calculado, indicando com 95% de confiança que o tempo de resposta para o cenário de reposição nos hospitais realmente é menor quando comparado com o cenário atual. O tempo total médio gasto com deslocamentos para reposição de material considerando todas as unidades é aproximadamente oito vezes maior no caso da reposição ser feita somente na central, sendo que para o segundo cenário, em média apenas 9.98% das reposições feitas necessitaram deslocamento para a central de regulação.

5.3 Otimização baseada em Simulação

O terceiro cenário procurou integrar o modelo de simulação desenvolvido com um procedimento de otimização para determinar a quantidade e localização ideal das unidades nas bases atuais de operação do Samu.

A formulação matemática do modelo implementado consiste na determinação da localização ideal das unidades móveis dentro de restrições relativas ao tempo de resposta do sistema, de maneira a minimizar uma função de custo. Para o problema tratado, foram considerados os custos mensais de se manter uma unidade básica ou avançada em operação. O primeiro passo é o levantamento de quais bases poderão ser utilizadas para modelagem e os limites relativos à quantidade de unidades que podem permanecer em espera até o momento de empenho em cada uma dessas bases. Dependendo do *range* utilizado o problema pode resultar em um número de cenários extremamente elevado, fazendo-se necessário o uso de métodos

heurísticos de solução. Para o presente estudo, foram consideradas as 20 bases atuais de operação com um *range* de 0 a 3 ambulâncias de cada tipo em cada base, não tendo sido variada a quantidade de ambulâncias de saúde mental. Os tempos de resposta médio foram calculados de maneira independente para cada tipo de unidade, de maneira que o otimizador não subestimasse o número de unidades avançadas buscando reduzir o valor da função objetivo, uma vez que este tipo de unidade é mais dispendiosa. Além disso, a modelagem desta maneira permite que os critérios de resposta sejam diferentes para cada tipo de ambulância, de acordo com interesses dos gestores do sistema. O problema de otimização pode ser formulado da seguinte maneira:

$$\underset{\theta \in \Theta}{\text{Min}} J(\theta) = E[C(\theta, \omega)]$$

onde θ representa o vetor de variáveis de entrada, neste caso representado pelo conjunto de bases para alocação e a quantidade de ambulâncias de cada tipo em cada base; $J(\theta)$ representa a função objetivo, ω representa a replicação, $C(\theta, \omega)$ representa a medida de desempenho (neste caso o custo total) da alternativa θ na replicação ω e $E[C(\theta, \omega)]$ é o valor esperado de $C(\theta, \omega)$. As restrições são:

$$\begin{aligned} R^a(\theta, \omega) &\leq RT^a \\ R^b(\theta, \omega) &\leq RT^b \\ N_i^a(\theta, \omega) &\leq NALS && \forall i \\ N_i^b(\theta, \omega) &\leq NBLs && \forall i \end{aligned}$$

onde $R^a(\theta, \omega)$ e $R^b(\theta, \omega)$ são os tempos de resposta para a configuração θ na replicação ω para unidades avançadas e básicas, respectivamente; $N_i^a(\theta, \omega)$ e $N_i^b(\theta, \omega)$ são a quantidade de unidades avançadas e básicas alocadas em casa base i na configuração θ e replicação ω ; RT^a e RT^b são limites superiores para os tempos de resposta das unidades avançadas e básicas; $NALS$ e $NBLS$ são os limites superiores para o total de unidades em cada base.

Para cada cenário, variou-se a restrição relativa ao tempo de resposta médio, sendo que o modelo de otimização foi implementado através do *OptQuest for Arena (OptTek Systems)*. Para $RT^a \leq 15$ minutos e $RT^b \leq 18$ minutos, a melhor solução encontrada correspondeu a 20 unidades básicas e 8 unidades avançadas (os valores atuais são de 17 e 5, respectivamente). Para este caso, $R^a(\theta, \omega) = 14.77$ minutos e $R^b(\theta, \omega) = 17.97$ minutos.

Considerando valores mais restritivos para os tempos de resposta, o cenário com RT^a e $RT^b \leq 15$ minutos retornou como melhor solução uma configuração com 24 unidades básicas e 8 unidades avançadas, com $R^a(\theta, \omega) = R^b(\theta, \omega) = 14.95$ minutos.

A situação mais restritiva analisada, considerando RT^a e $RT^b \leq 10$ minutos não encontrou solução viável dentro das configurações atuais do sistema. É possível observar, analisando os cenários anteriores, que existe uma tendência de acúmulo de unidades em determinadas bases. Por exemplo, no cenário atual as 20 bases de operação possuem pelo menos uma ambulância alocada, já no cenário com $RT^a \leq 15$ e $RT^b \leq 18$, temos 4 bases que ficaram desocupadas. Este cenário indica que para ganhos mais expressivos na variável tempo de resposta, uma possibilidade interessante seria o reposicionamento das bases atuais através de um estudo de localização das unidades de resgate, evitando aumentar exageradamente o número de ambulâncias em cada base.

6. Conclusão

Serviços de atendimento móveis de urgência possuem uma grande aleatoriedade associada aos tempos dos diversos processos que fazem parte das operações rotineiras do sistema.

Mesmo aquelas operações que possuem regras aparentemente claras de funcionamento apresentam componentes que não são possíveis de mensurar e que influenciam diretamente no funcionamento das diversas atividades. Considerando as características específicas deste tipo de sistema, a técnica de simulação a eventos discretos juntamente com procedimentos de otimização foi utilizada dentro de uma metodologia para análise de sistemas médicos de emergência.

O uso combinado de técnicas de otimização e simulação melhora a busca por configurações ótimas do sistema, permitindo a simulação e análise de um número grande de alternativas.

Como sugestão para trabalhos futuros de análise de sistemas de atendimento móveis de urgência, podemos citar o desenvolvimento de algoritmos ou métodos heurísticos para definição de políticas de despacho, buscando otimizar a utilização das unidades e reduzir o tempo de resposta.

Referências

- Aringhieri, R., Carello, G. e Morale, D.** (2007), Ambulance location through optimization and simulation: the case of Milano urban area. *XXXVIII Annual Conference of the Italian Operations Research Society Optimization and Decision Sciences*, 29 pp.
- Brailsford, S. C.** (2007), Tutorial: advances and challenges in healthcare simulation modeling. *Proceedings of the 2007 Winter Simulation Conference*, p. 1436-1448.
- Church, R. L. e Reville, C.** (1974), The maximal covering location problem. *Papers of the Regional Science Association*, v. 32, p. 101-118.
- Fetter, R. B. e Thompson, J. D.** (1966), Patients' waiting time and doctors' idle time in the outpatient setting. *Health Services Research* 1, p. 66-90.
- Fitzsimmons, J. A.** (1971), An emergency medical system simulation model. *Proceedings of the 5th Conference on Winter Simulation*, p. 18-25.
- Fitzsimmons, J. A. e Fitzsimmons, M. J.** *Administração de serviços: operações, estratégia e tecnologia da informação*. 4. ed. Porto Alegre: Artmed Editora S. A., 2004. 564 p.
- Gonçalves, M. B.** (1994), Métodos de pesquisa operacional em serviços emergenciais. *Anais do XXVI^o Simpósio Brasileiro de Pesquisa Operacional*, SOBRAPO, Florianópolis, v. 1, p. 597-601.
- Koch, O. e Weigl, H.** (2003), Modeling ambulance service of the Austrian red cross. *Proceedings of the 2003 Winter Simulation Conference*, p. 1701-1706.
- Maxwell, M. S., Henderson, S. G. e Topaloglu, H.** (2009), Ambulance redeployment: an approximate dynamic programming approach. *Proceedings of the 2009 Winter Simulation Conference*, p. 1850-1860.
- Novaes, A. G.** *Sistemas logísticos: transporte, armazenagem e distribuição física de produtos*. São Paulo: Editora Edgard Blucher, 1989. 372 p.
- Oliveira, L. K. de e Gonçalves, M. B.** (2004), Metodologia para avaliar centro de emergência: aplicação ao centro de emergência da Polícia Militar de Santa Catarina. *Revista Produção Online*, v. 4, n. 2.
- Shih, C. L. e Su, S.** (2003), Modeling an emergency medical services system using computer simulation. *International Journal of Medical Informatics* 72(3), p. 57-72.
- Takeda, R. A.** (2000), Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde. 2000. 210 pp. *Tese (Doutorado em Transportes)* – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo. 2000.
- Takeda, R. A., Widmer, J. A. e Morabito, R.** (2001), Uma proposta alternativa para avaliação do desempenho de sistemas de transporte emergencial de saúde brasileiros. *Transportes*, v. 9, n. 2, p. 9-27.
- Timerman, S., Gonzalez, M. M. C., Zaroni, A. C. e Ramires, J. A. F.** (2006), Emergency medical services: Brazil. *Resuscitation* 70, p. 356-359.
- Toregas, C., Swain, R., Reville, C. e Bergman, L.** (1971), The location of emergency service facilities. *Operations Research* 19, p. 1363-1373.