



Análise da configuração de SAMU utilizando múltiplas alternativas de localização de ambulâncias

Analysis of SAMU configuration using multiple alternatives of ambulance location

Regiane Máximo de Souza¹

Reinaldo Morabito²

Fernando Yassuo Chiyoshi³

Ana Paula Iannoni⁴

Resumo: O Sistema de Atendimento Móvel de Urgência (SAMU) no Brasil é um sistema médico emergencial de responsabilidade do poder público, em que a demanda de usuários em uma região urbana é usualmente separada por subregiões e classes de chamados emergenciais. Essa demanda pode mudar de forma significativa ao longo do dia, geograficamente e temporalmente, devido à sua natureza aleatória, mas também devido aos diferentes padrões de comportamentos da população ao longo do dia. Por exemplo, tipicamente há menos demanda durante a noite do que de dia. Os objetivos deste trabalho são: verificar se o conhecido modelo hipercubo de filas espacialmente distribuídas é adequado para analisar medidas de desempenho do SAMU, tais como tempos médios de resposta aos usuários, e utilizar este modelo para analisar múltiplas alternativas de localização das ambulâncias, explorando variações importantes da demanda e do serviço ao longo do dia. Para verificar a viabilidade e a aplicabilidade desta abordagem, foi realizado um estudo de caso no SAMU de Ribeirão Preto-SP.

Palavras-chave: Modelo hipercubo. SAMU. Localização de ambulâncias. Atendimento médico emergencial.

Abstract: *The Brazilian emergency medical system SAMU (Sistema de Atendimento Móvel de Urgência) is an emergency medical system of public government liability, in which the users' service demand in an urban region is usually separated into subregions and classes of emergency calls. This demand can change substantially during the day, geographically and temporally, due to its random nature and also to the different behavior patterns of the population throughout the day. For instance, typically there is less demand during the night hours than during the day. The goals of this study are to verify whether the hypercube queuing model is adequate to analyze performance measures of SAMU, such as mean response times to the users, and use this model to analyze multiple alternatives of ambulance location considering significant variations in the demand and service throughout the day. In order to verify the feasibility and applicability of this approach, a case study was conducted in the SAMU of Ribeirão Preto-SP.*

Keywords: *Hypercube queuing model. SAMU. Ambulance location. Emergency medical system.*

1 Introdução

A qualidade de vida da população está diretamente ligada ao acesso à saúde. Principalmente em áreas urbanas, há um grande número de acidentes e outras ocorrências, como, por exemplo, infarto, intoxicações, queimaduras, afogamentos e quedas acidentais. Nos Sistemas de Atendimento de Emergência (SAEs), de forma geral, o tempo médio de resposta ao usuário é de fundamental importância, pois a demora no atendimento pode implicar em prejuízos

e sequelas importantes para os usuários. Em especial, o SAMU é um programa do governo federal que tem a finalidade de prestar socorro médico emergencial às pessoas e garantir a qualidade no atendimento em centros urbanos. Esses sistemas são caracterizados essencialmente por incertezas quanto a disponibilidade, localização e tempo de serviço dos servidores, e quanto à demanda dos usuários e o tempo de resposta para

¹ Departamento de Engenharia de Produção, Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP, CEP 17033-360, Bauru, SP, Brasil, e-mail: regiane@feb.unesp.br

² Departamento de Engenharia de Produção, Universidade Federal de São Carlos – UFSCar, CP 676, CEP 13565-905, São Carlos, SP, Brasil, e-mail: morabito@ufscar.br

³ Programa de Engenharia de Produção – COPPE, Universidade Federal do Rio de Janeiro – UFRJ, CP 68507, CEP 21945-970, Rio de Janeiro, RJ, Brasil, e-mail: chiyoshi@pobox.com

⁴ Laboratoire Genie Industriel, Ecole Centrale Paris, Chatenay Malabry, 92295, Paris, França, e-mail: iannoni93@hotmail.com

atendimento. Observa-se que, nas últimas décadas, vem aumentando o interesse de pesquisadores em estudos de SAEs urbanos (SIMPSON; HANCOCK, 2009; TAKEDA; WIDMER; MORABITO, 2007).

Para projetar e analisar a configuração de SAEs e SAMUs, gestores e tomadores de decisão destes sistemas precisam balancear o nível de serviço oferecido aos usuários com os investimentos requeridos no sistema para poder oferecer esse nível de serviço. Eles também precisam definir e analisar medidas de desempenho de maneira a serem capazes de avaliar adequadamente a configuração do sistema. Isso pode ser realizado com o auxílio de modelos analíticos de teoria de filas e modelos experimentais de simulação, com usuários e servidores do sistema espacialmente distribuídos na região de interesse e servidores considerados como distinguíveis ou indistinguíveis, dependendo da natureza do serviço em consideração (SWERSEY, 1994; GALVÃO; MORABITO, 2008).

O modelo hipercubo, proposto originalmente por Larson (1974) e estendido por vários autores (GALVÃO; MORABITO, 2008) é um modelo descritivo baseado na teoria de filas espacialmente distribuídas, utilizado para analisar e planejar SAEs. O modelo considera os servidores do sistema individualmente e disciplinas de prioridade no despacho destes servidores, se houverem servidores disponíveis para despacho imediato, e fila com disciplina de atendimento FCFS (*First Come First Served*), no caso de todos os servidores estarem ocupados. Exemplos de aplicação do modelo hipercubo para análise da configuração de SAEs nos Estados Unidos podem ser encontrados em Chelst e Barlach (1981), Brandeau e Larson (1986), Burwell, Jarvis e McKnew (1993), Sacks e Grieff (1994), Swersey (1994) e Larson e Odoni (2007). No Brasil, alguns exemplos aparecem em Gonçalves, Novaes e Albino (1994), Gonçalves, Novaes e Schmitz (1995), Mendonça e Morabito (2001), Oliveira (2003), Chiyoshi, Galvão e Morabito (2000), Costa (2004), Figueiredo e Lorena (2005), Luque (2006), Takeda, Widmer e Morabito (2007), Iannoni e Morabito (2006, 2007) e Morabito, Chiyoshi e Galvão (2008). Neste trabalho, o modelo hipercubo foi aplicado para analisar a configuração de um SAMU, e um modelo de simulação discreta também foi utilizado para verificar a validade de algumas hipóteses simplificadoras do modelo hipercubo.

Durante um dia típico de operação de um SAMU, em determinados períodos pode haver diferenças significativas com relação ao número e à localização dos chamados. Diversas medidas de desempenho relevantes para se analisar SAMUs podem ser estimadas pelo modelo hipercubo. Elas podem ser separadas em medidas externas do ponto de vista do usuário, tais como tempos médios de resposta a um chamado, tempos médios de viagem para cada área da cidade (referida neste estudo como *átomo*)

e frequências de chamadas atendidas em um tempo inferior a um limite determinado; e medidas internas do ponto de vista do gerente do sistema, tais como cargas de trabalho das ambulâncias, frequências de despacho das ambulâncias para os átomos, frações de atendimentos realizados fora da área de cobertura de cada ambulância e tempos médios de viagem para cada ambulância.

Dada a relevância dos SAMUs em cidades brasileiras, os objetivos deste trabalho foram: (i) avaliar a aplicabilidade e utilidade do modelo hipercubo para analisar o SAMU da cidade de Ribeirão Preto (SAMU-RP) na sua configuração; (ii) analisar possibilidades de múltiplas alternativas de localização de ambulâncias do SAMU-RP ao longo do dia de operação do sistema por meio do modelo hipercubo, considerando as variações na demanda e no serviço do sistema. Um estudo anterior foi realizado em Takeda, Widmer e Morabito (2007) com o SAMU-Campinas, em que foi sugerido desenvolver mais pesquisas, aplicando o modelo hipercubo em outros SAMUs, para verificar e analisar melhor sua aplicabilidade nestes sistemas. No presente estudo, pretendemos seguir esta linha de pesquisa e investigar vantagens da utilização de múltiplas alternativas de localização de ambulâncias ao longo do dia de operação do sistema, o que não foi explorado em Takeda, Widmer e Morabito (2007). Além disso, o SAMU-RP tem algumas características operacionais diferentes do SAMU-Campinas estudado em Takeda, Widmer e Morabito (2007), que foram aqui consideradas. Por exemplo, além da prioridade no despacho de ambulâncias quando o sistema não se encontra saturado (que é tratado no modelo hipercubo por meio da estratégia de *layering*; LARSON; ODoni, 2007), o SAMU-RP também considera prioridade neste despacho quando o sistema se encontra com fila de espera de usuários.

Desta maneira, neste estudo, pretendemos também mostrar que o modelo hipercubo clássico (sem prioridade na fila) também pode ser aplicado em SAMUs com prioridade na fila, desde que o nível de utilização do sistema seja relativamente baixo. Alguns experimentos foram realizados para se mostrar isso, em que são discutidos os limites desta análise com modelo hipercubo clássico (sem prioridade na fila) em sistemas com prioridade na fila. Para isso, os resultados destes experimentos foram comparados com os obtidos com um modelo de simulação discreta que considera prioridade na fila. Na próxima seção, apresentamos uma breve descrição do SAMU-RP; na seção 3, descrevemos resumidamente o modelo hipercubo (para que o texto fique mais autocontido); e na seção 4, discutimos a aplicação do modelo hipercubo no SAMU-RP. Na seção 5, analisamos os resultados obtidos para a configuração do SAMU-RP e, também, investigando-se cenários alternativos

para o sistema. Finalmente, a seção 6 apresenta as conclusões deste estudo e algumas perspectivas para pesquisas futuras.

2 O SAMU de Ribeirão Preto (SAMU-RP)

Conforme mencionado, o SAMU tem a responsabilidade de socorrer vítimas e, nos casos mais urgentes, fazer o pré-atendimento, no sentido de estabilizar a vítima, e depois encaminhá-la a um hospital próximo. O SAMU de Ribeirão Preto conta com 1 veículo de suporte avançado (VSA) e 9 veículos de suporte básicos (VSB), que são despachados, dependendo da gravidade do chamado e da sua localização. Estes veículos estão distribuídos em cinco bases de atendimento no município, situadas em hospitais ou postos de saúde. O VSA e os VSBs diferem basicamente com relação aos equipamentos e à equipe enviada ao chamado. O VSA é constituído por uma equipe formada por um médico, um enfermeiro e um motorista, e os equipamentos são mais sofisticados que os dos VSBs, a fim de que haja melhor atendimento e seja possível estabilizar o paciente até a sua chegada ao hospital. Os VSBs são constituídos por uma equipe formada por um auxiliar de enfermagem e um motorista, que são enviados preferencialmente para atender casos urgentes, mas também atendem a emergências quando o VSA está ocupado.

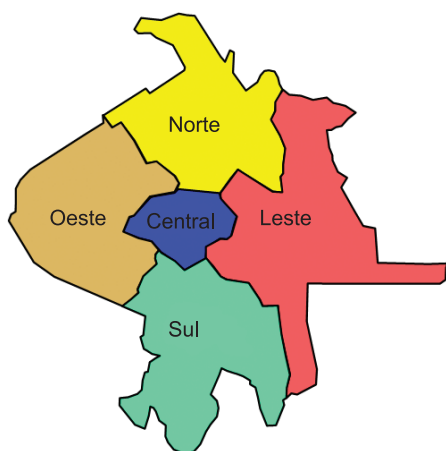
A cidade de Ribeirão Preto está dividida pelo SAMU-RP em cinco regiões: Norte, Sul, Leste, Oeste e Central (Figura 1), e as ambulâncias são distribuídas ao longo delas. A Região Central possui três ambulâncias, sendo dois VSBs e o VSA. As Regiões Norte, Sul e Oeste possuem dois VSBs cada uma e a Região Leste possui um VSB. Uma característica importante do SAMU-RP é o atendimento a mais de uma classe de chamados emergenciais de usuários. As urgências são divididas em duas classes: moderada (chamados

da classe *b*), como fraturas expostas, bronquites, quadros pneumônicos, choques cardiogênicos, entre outros; e leve (chamados da classe *c*), situações em que a o paciente apresenta um quadro menos preocupante como, lombalgias, febres, dores de cabeça, entre outros. Para um chamado caracterizar uma emergência (chamados da classe *a*), é necessário que haja risco de vida ao paciente, em casos como: comprometimentos da via aérea, comprometimentos da dinâmica respiratória, comprometimentos da dinâmica circulatória, comprometimentos da função neurológica ou comprometimentos funcionais de extremidades.

Além dos atendimentos dos chamados dos tipos *a*, *b* e *c*, descritos anteriormente, o SAMU-RP também realiza atendimentos agendados de remoção de pacientes. Esses chamados são caracterizados principalmente pelo transporte de pacientes (de hospital para hospital, de casa para hospital ou vice-versa). No presente estudo, esses atendimentos de remoção não foram considerados nas análises por, em geral, não se tratarem de serviços emergenciais.

3 O modelo hipercubo

O modelo hipercubo é um modelo analítico que expande o espaço de estados de um sistema de fila *M/M/m* a fim de representar cada servidor individualmente, desta maneira, podendo considerar políticas de despacho mais complicadas no sistema. A solução do modelo é dada partindo-se da construção de um conjunto de equações de equilíbrio (*steady state*) para o sistema. Os resultados baseiam-se nos valores das probabilidades de estado de equilíbrio do sistema, possibilitando o cálculo de diversas medidas de desempenho, tais como: cargas de trabalho dos servidores, tempo médio de resposta do sistema ou de cada servidor, frequência de atendimento de cada servidor em cada região, entre outras. O modelo



Ribeirão Preto

Área: 652 km²

População: 547.417 habitantes

SAMU - 192

Média mensal de atendimentos emergenciais 10000 chamados, sendo 4100 chamados de alta, média e baixa complexidade. O sistema realiza 5900 atendimentos de remoção.

Médicos reguladores: 20

Enfermeiros técnicos: 75

Motoristas: 68

Hospitais: 1 público

Figura 1. Distribuição espacial do sistema SAMU - Ribeirão Preto em 2005.

baseia-se na divisão da região atendida pelo sistema em átomos geográficos (regiões de demanda). Cada átomo é considerado uma fonte de chamados pontual e independente das demais e o atendimento a cada átomo é realizado por servidores que estão distribuídos na região. A localização dos servidores deve ser conhecida ou estimada por meio de técnicas de probabilidade geométrica. Se um servidor estiver ocupado, outros servidores poderão atender ao chamado, mesmo que estejam fora de sua região preferencial do chamado, prevalecendo a cooperação entre os servidores.

A disponibilidade dos servidores é representada por meio do espaço de estados dos servidores. Um estado particular do sistema sem fila é dado pela lista dos servidores que estão livres ou ocupados. Considere um sistema com apenas $m = 3$ servidores e sejam $\{000\}$, $\{001\}$, $\{010\}$, ..., $\{111\}$ os $2^3 = 8$ possíveis estados do sistema, em que os 0 s e 1 s indicam se cada um dos três servidores está livre ou ocupado, respectivamente. Por exemplo, o estado $\{011\}$ representa o estado em que o servidor 1 está livre e os servidores 2 e 3 estão ocupados (note que $\{011\}$ descreve o estado dos servidores da esquerda para a direita). Assim, o espaço de estados desse sistema com três servidores pode ser representado por um cubo; no caso de haverem mais de três servidores, temos um hipercubo. A Figura 2 ilustra o espaço de estados desse sistema com três servidores. O modelo hipercubo trata tanto sistemas em que não é permitida a formação de fila, como aqueles em que quando todos os servidores estão ocupados, os chamados que chegam esperam em uma fila por meio da qual os usuários são atendidos à medida em que os servidores tornam-se desocupados segundo a disciplina FCFS. Os demais estados S_4, S_5, S_6, \dots , da Figura 2 representam os estados com 1, 2, 3, ..., usuários na fila de espera do sistema.

No caso dos m servidores serem homogêneos, com mesma taxa de atendimento, o modelo hipercubo, que considera os m servidores distinguíveis, possui uma relação direta com o modelo clássico de fila $M/M/m$, que considera os m servidores indistinguíveis. Segundo Larson e Odoni (2007), existem nove hipóteses principais que devem ser verificadas para a aplicação do modelo hipercubo:

- A região deve ser dividida em N_A átomos;
- As solicitações por serviço em cada átomo j ($j = 1, 2, \dots, N_A$) chegam independentemente e de acordo com uma distribuição de Poisson;
- Os tempos de viagem de um átomo i para um átomo j ($i, j = 1, \dots, N_A$) devem ser conhecidos ou estimados;
- O sistema opera com m servidores espacialmente distribuídos, homogêneos ou não, que podem se deslocar para atender a qualquer um dos átomos;
- Quando disponíveis, a localização dos servidores deve ser conhecida, ao menos probabilisticamente;
- Apenas um servidor é despachado para atender a cada chamado. Quando chamados estiverem esperando em fila, a escolha do chamado a ser atendido utiliza a disciplina FCFS;
- Há uma lista de preferências de despacho de servidores para cada átomo;
- O tempo total de atendimento de um chamado é exponencialmente distribuído e composto pela somatória dos seguintes tempos: tempo de preparo do servidor (*setup time*), tempo de viagem do servidor até o local da ocorrência, tempo de execução do serviço junto ao usuário (tempo em cena) e o tempo de viagem de retorno à base; e

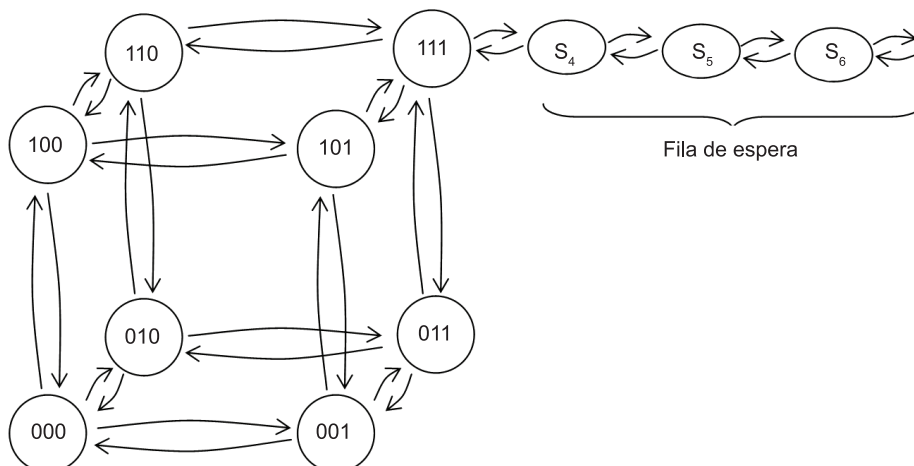


Figura 2. Estados do sistema com três servidores.

- Variações no tempo total de atendimento devido às variações no tempo de viagem são consideradas de segunda ordem, quando comparadas às variações dos tempos em cena e/ou tempo de preparação da equipe.

Diversas medidas de desempenho podem ser calculadas a partir das probabilidades do sistema estar em cada estado, obtidas pela solução das equações de equilíbrio do sistema. Essas medidas auxiliam na configuração e análise do sistema sob a hipótese de que o sistema esteja em equilíbrio. Alguns exemplos são:

- Carga de trabalho (*workload*): é a fração de tempo em que o servidor está ocupado e é calculada somando-se as probabilidades dos estados em que este servidor estiver ocupado;
- A frequência de despacho: indica a fração de despachos no sistema que é atendida pelo servidor i ($i = 1, 2, \dots, m$) no átomo j ($j = 1, 2, \dots, N_A$);
- O tempo médio de viagem no sistema;
- O tempo médio de viagem ao átomo j ; e
- O tempo médio de viagem do servidor i .

Outras medidas de desempenho podem ser definidas e calculadas, como a fração de despachos do servidor i como *backup*, a fração de despachos de *backup* para o átomo j e a fração total de despachos *backup* no sistema. Para mais detalhes do modelo hipercubo, o leitor pode consultar, por exemplo, Larson e Odoni (2007).

4 Aplicação do modelo hipercubo no SAMU-RP

4.1 Descrição dos atendimentos

Um estudo do número e distribuição dos atendimentos ao longo de um ano de operação do sistema SAMU-RP foi realizado a fim de se verificar se houve diferenças significativas no número de

solicitações, por exemplo, em períodos de férias escolares. Para analisar as variações da frequência dos chamados, considerou-se a distribuição dos atendimentos realizados neste ano do SAMU-RP. A Figura 3 mostra a distribuição dos chamados de janeiro a dezembro deste ano. Para verificar se o número médio de atendimentos foi estatisticamente igual em todos os meses deste ano, foi aplicada a análise de variância (ANOVA) com nível de significância $\alpha = 0,05$. Foi verificado que apenas o mês de abril tem média diferente dos outros meses. Conforme comunicação com os gestores do SAMU-RP, esse mês foi considerado atípico, uma vez que houve dois feriados prolongados, neste ano.

Também se verificou se houve diferenças estatísticas significativas com relação ao número de atendimentos em cada mês, por exemplo, se nos finais de semana houve mais ou menos atendimentos em relação aos dias úteis. Foi observado que não houve diferenças significativas em todos os meses deste ano. Então sorteou-se aleatoriamente um mês (agosto) e depois dez dias deste mês também foram escolhidos aleatoriamente para verificar diferenças com relação ao número de chamados ao longo do dia. Foi feita a análise estatística do número médio de atendimentos destes dez dias, e o resultado mostrou que não houve diferenças significativas do número médio de chamados, com $\alpha = 5\%$ de significância. Para fins de ilustração, a Figura 4 mostra, por exemplo, a distribuição dos chamados (por hora) no 4º dia observado em agosto. Pode-se verificar que há dois períodos de pico distintos: por volta de 12 horas e 20 horas.

4.2 Períodos de pico

Cada dia foi então dividido em três partes: manhã (8h00 às 16h00), tarde (16h00 às 24h00) e noite (00h00 às 8h00), de acordo com discussão com os gestores do sistema. Em cada parte do dia, foi determinado um período de pico a ser estudado, conforme descrito a

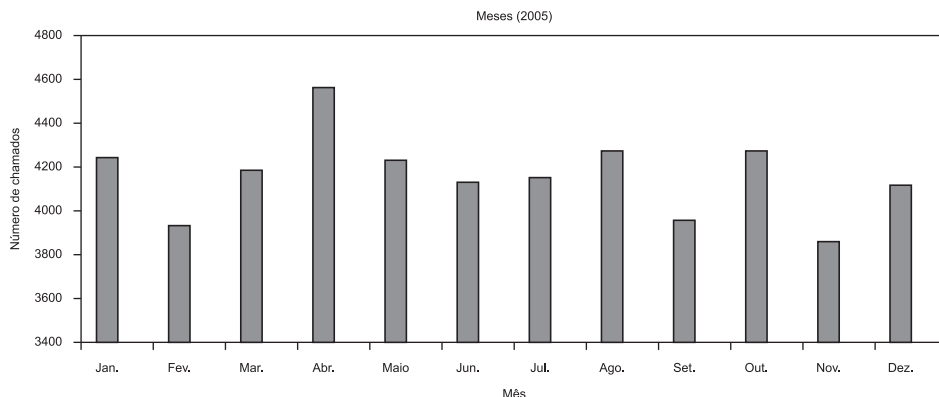


Figura 3. Atendimentos de urgência e emergência realizados pelo SAMU_RP em 2005.

seguir. A análise dos períodos de pico foi feita com os dados obtidos a partir do controle dos chamados do SAMU-RP. A Tabela 1 mostra a taxa média de chegada (número de chamados dividido pelo período de tempo considerado, em minutos) para vários períodos do dia, durante os 10 dias estudados. Foram analisados vários períodos de diferentes tamanhos, e o período de pico em cada parte do dia foi identificado pela análise feita a partir da média e do desvio padrão da taxa de chegada. Os períodos escolhidos (linhas destacadas em negrito da Tabela 1) foram os que apresentaram maior taxa média de chegada (penúltima coluna da Tabela 1) e, em caso de empate, o menor desvio padrão (última coluna), similarmente ao que foi feito no estudo em Takeda, Widmer e Morabito (2007) para o SAMU-Campinas. A seguir, são mencionados apenas períodos manhã, tarde e noite, fazendo referência aos três períodos de pico correspondentes.

Observa-se que, caso a análise fosse feita em apenas um período do dia, o escolhido seria o período da

tarde (das 18h00 às 23h00). Mais adiante é mostrado que esse não é o período mais desfavorável do ponto de vista das medidas de desempenho, como *workload* e tempos de resposta, se comparado com o período da manhã.

4.3 Hipóteses do modelo hipercubo para o SAMU-RP

4.3.1 Área dividida em átomos geográficos do sistema e processo de chegada

Há várias maneiras de se fazer a representação da área estudada em átomos geográficos, como divisão política, bairros, setores policiais, entre outros. Neste trabalho, optou-se por utilizar a divisão por setores, já utilizada pelo SAMU-RP: Norte, Sul, Leste, Oeste e Central. Cada átomo geográfico (Centro – 1; Norte – 2; Sul – 3; Leste – 4; Oeste – 5) foi dividido em três subátomos, um para cada classe de usuários (grave (a),

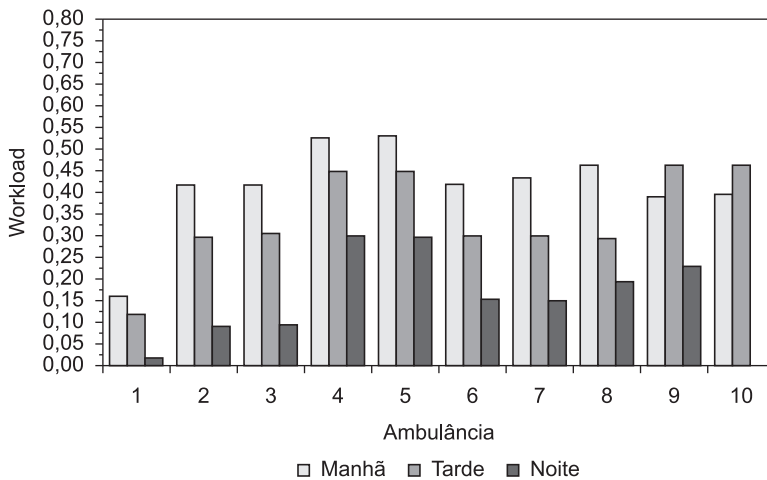


Figura 4. atendimentos realizados pelo SAMU_RP no 4º dia observado em agosto de 2005.

Tabela 1. Análise do período de pico para os três períodos do dia: manhã, tarde e noite.

Período (horas)	Dia										Média	D-P	
	1	2	3	4	5	6	7	8	9	10			
Manhã													
9 - 14	0,103	0,103	0,110	0,093	0,130	0,113	0,083	0,133	0,117	0,110	0,110	0,110	0,015
10 - 14	0,117	0,117	0,113	0,104	0,129	0,121	0,083	0,121	0,113	0,108	0,113	0,113	0,012
11 - 14	0,106	0,106	0,083	0,106	0,139	0,139	0,089	0,111	0,122	0,106	0,111	0,111	0,018
Tarde													
17 - 23	0,100	0,142	0,133	0,111	0,111	0,144	0,128	0,106	0,125	0,111	0,121	0,121	0,016
18 - 23	0,113	0,133	0,147	0,130	0,117	0,157	1,333	0,103	0,140	0,113	0,249	0,381	0,381
19 - 23	0,108	0,117	0,146	0,125	0,100	0,167	0,138	0,113	0,125	0,129	0,127	0,020	0,020
Noite													
24 - 6	0,047	0,056	0,047	0,058	0,078	0,053	0,058	0,075	0,036	0,039	0,055	0,055	0,014
24 - 7	0,045	0,069	0,045	0,060	0,076	0,055	0,052	0,083	0,038	0,040	0,056	0,056	0,015
1 - 6	0,040	0,043	0,050	0,053	0,073	0,043	0,053	0,070	0,037	0,043	0,051	0,051	0,012

moderada (b) e leve (c)), totalizando quinze subátomos no sistema (ver hipótese 1 do modelo hipercubo na seção 3): Centro a (1a), Centro b (1b), Centro c (1c), Norte a (2a), Norte b (2b), Norte c (2c), Sul a (3a), Sul b (3b), Sul c (3c), Leste a (4a), Leste b (4b), Leste c (4c), Oeste a (5a), Oeste b (5b), Oeste c (5c). Conforme Larson e Odoni (2007) e Takeda, Widmer e Morabito (2007), esta estratégia de dividir átomos em subátomos, chamada de *layering*, é utilizada para diferenciar classes de usuários e permitir que o modelo hipercubo considere prioridade no despacho das ambulâncias para cada classe, quando o sistema não se encontra saturado (i.e., com todas as ambulâncias ocupadas). Assim, são devidamente representadas no modelo as três classes de usuários em cada átomo do SAMU-RP (Figura 5). Entretanto, convém notar que esta abordagem torna-se uma aproximação em sistemas congestionados que permitam a formação de filas de espera, pois, conforme hipótese 6 na seção 3, o modelo hipercubo considera fila com disciplina de atendimento FCFS (i.e., não considera explicitamente prioridades na fila).

De todos os dez dias considerados, apenas os dias em que estavam operando dez ambulâncias nos períodos da manhã e da tarde e os dias em que estavam operando nove ambulâncias no período da

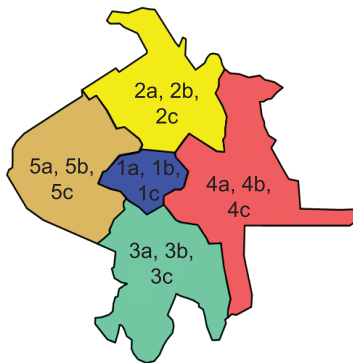


Figura 5. Átomos geográficos do SAMU-RP.

noite foram considerados, de forma que fizeram parte da análise 9, 10 e 6 dias nos períodos da manhã, tarde e noite, respectivamente. Para se verificar se o processo de chegada de chamadas pode ser considerado como um processo de Poisson (ver hipótese 2 do modelo hipercubo na seção 3), testes de aderência foram realizados utilizando-se o *software* BestFit. Esses testes mostraram que, com um nível de significância de 5%, não se pode rejeitar a hipótese de que os intervalos entre chegadas sucessivas tenham distribuição exponencial nos três períodos do dia. O gráfico ilustrado na Figura 6 mostra a porcentagem do número de chamados em cada subátomo nos três períodos: manhã, tarde e noite. A Tabela 2 mostra os números de chamados e as proporções de chegadas destes chamados em cada subátomo do sistema, assim como as taxas médias de chegada (chamados por hora) em cada subátomo, em cada período do dia.

4.3.2 Tempos de viagem e localização dos servidores

Os tempos de viagem dos servidores entre os átomos foram obtidos a partir dos dados coletados do próprio SAMU-RP (ver hipótese 3 do modelo hipercubo na seção 3). Conforme mencionado, o SAMU-RP é composto por uma frota de dez ambulâncias distintas: nove VSBs e um VSA. Estas ambulâncias estão descentralizadas, localizadas, quando disponíveis, em cinco postos de saúde distribuídos um em cada setor da cidade (ver hipótese 5 do modelo hipercubo na seção 3), podendo deslocar-se para qualquer átomo do sistema para realizar um atendimento. As ambulâncias estão distribuídas da seguinte forma: dois VSBs e um VSA na Região Central, as Regiões Norte, Sul e Oeste possuem dois VSBs cada e a Região Leste possui um VSB, nos períodos da manhã e tarde. À noite, a Região Oeste tem apenas um VSB, totalizando 9 ambulâncias, neste período.

Por conveniência, a seguir, o VSA é chamado de veículo 1, enquanto os VSBs são chamados de

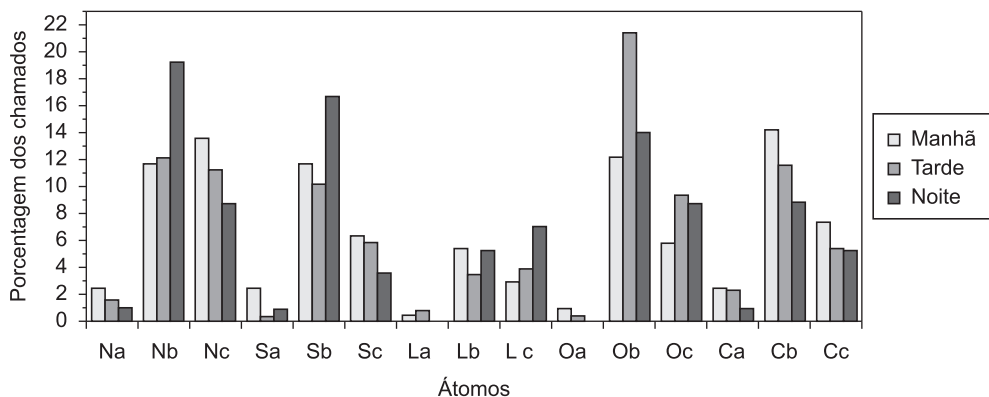


Figura 6. Porcentagem do número de chamados em cada subátomo nos períodos manhã, tarde e noite.

Tabela 2. Taxas médias de chegada (chamados/hora) para cada subátomo.

Subátomo	Nº de chamados			Proporção			Taxa de chegada			
	Manhã	tarde	noite	manhã	Tarde	noite	manhã	tarde	noite	
C	1a	5	4	1	0,024	0,016	0,009	0,116	0,085	0,026
	1b	24	31	22	0,117	0,121	0,193	0,559	0,660	0,563
	1c	28	29	10	0,137	0,113	0,088	0,652	0,617	0,256
N	2a	5	6	1	0,024	0,023	0,009	0,116	0,128	0,026
	2b	29	30	10	0,142	0,117	0,088	0,675	0,638	0,256
	2c	15	14	6	0,073	0,055	0,053	0,349	0,298	0,154
S	3a	5	1	1	0,024	0,004	0,009	0,116	0,021	0,026
	3b	24	26	19	0,117	0,101	0,167	0,559	0,553	0,486
	3c	13	15	4	0,063	0,058	0,035	0,303	0,319	0,102
L	4a	1	2	0	0,005	0,008	0,000	0,023	0,043	0,000
	4b	11	9	6	0,054	0,035	0,053	0,256	0,192	0,154
	4c	6	10	8	0,029	0,039	0,070	0,140	0,213	0,205
O	5a	2	1	0	0,010	0,004	0,000	0,047	0,021	0,000
	5b	25	55	16	0,122	0,214	0,140	0,582	1,170	0,409
	5c	12	24	10	0,059	0,093	0,088	0,279	0,511	0,256
Total	205	257	114	1,000	1,000	1,000	4,772	5,469	2,916	

Tabela 3. Tempos e taxas médias de atendimento para cada ambulância.

Ambulância	Tempo médio de serviço (minutos)			Desvio padrão			Coeficiente de variação			Taxa de serviço (chamados/hora)			Tempo médio de viagem (minutos)			Desvio padrão		
	M	T	N	M	T	N	M	T	N	M	T	N	M	T	N	M	T	N
1	41,1	47,2	34,3	12,4	14,1	12,3	0,30	0,30	0,36	1,46	1,27	1,75	10,1	9,8	10,0	3,9	3,4	5,7
2	44,1	30,4	21,0	26,2	18,8	7,8	0,59	0,62	0,37	1,36	1,97	2,86	9,5	8,8	8,4	3,5	4,6	3,3
3	44,1	30,4	21,0	26,2	18,8	7,8	0,59	0,62	0,37	1,36	1,97	2,86	9,5	8,8	8,4	3,5	4,6	3,3
4	60,8	45,6	45,0	36,2	28,7	33,1	0,60	0,63	0,74	0,99	1,32	1,33	12,3	11,7	10,9	8,0	5,7	4,9
5	60,8	45,6	45,0	36,2	28,7	33,1	0,60	0,63	0,74	0,99	1,32	1,33	12,3	11,7	10,9	8,0	5,7	4,9
6	51,4	33,1	26,5	40,3	13,6	9,8	0,78	0,41	0,37	1,17	1,81	2,26	9,3	9,7	9,1	6,1	5,1	2,9
7	51,4	33,1	26,5	40,3	13,6	9,8	0,78	0,41	0,37	1,17	1,81	2,26	9,3	9,7	9,1	6,1	5,1	2,9
8	79,6	38,3	35,6	58,4	18,9	29,1	0,73	0,49	0,82	0,75	1,57	1,68	12,1	10,2	10,0	3,8	5,6	5,1
9	46,5	39,6	25,6	31,2	30,0	11,0	0,67	0,76	0,43	1,29	1,51	2,34	9,0	9,0	8,3	6,4	5,3	3,4
10	46,5	39,6	...	31,2	30,0	...	0,67	0,76	...	1,29	1,51	...	9,0	9,0	...	6,4	5,3	...
VSA	41,1	47,2	34,3	12,4	14,1	12,3	0,30	0,30	0,36	1,46	1,27	1,75	10,1	9,8	10,0	3,9	3,4	5,7
VSB	53,9	37,3	30,8	36,2	22,3	17,7	0,67	0,59	0,53	1,15	1,64	2,12	10,3	9,8	9,4	5,8	5,2	3,9
Total	52,6	38,3	31,2	33,8	21,5	17,1	0,63	0,56	0,51	1,18	1,61	2,08	10,3	9,8	9,5	5,6	5,1	4,1

veículos 2, 3, ..., 10, respectivamente, para os períodos da manhã e tarde. No período da noite, o VSA é chamado de veículo 1, e os VSBs são chamados de 2, 3, ..., 9, dado que a ambulância 10 não opera à noite.

4.3.3 Política de despacho dos servidores

O SAMU-RP em geral envia apenas uma equipe para atender a um chamado, que é um dos requisitos para a aplicação do modelo hipercubo (ver hipótese 6 na seção 3). Além disso, o VSA pode atender apenas a chamados de emergência, enquanto os VSBs podem atender a qualquer tipo de chamado. Nos experimentos com modelo hipercubo da próxima

seção, simplificamos esta característica do sistema, considerando que todos os servidores podem atender a qualquer tipo de chamado (ver hipótese 4 do modelo hipercubo na seção 3). Conforme visto adiante, essa simplificação é razoável, uma vez que as frequências de despacho do VSA para chamados do tipo *b* e *c* são bem pequenas: 0,0073, 0,0023, e menores que 10^{-4} para os períodos da manhã, tarde e noite, respectivamente.

A política de despacho dos servidores depende da distribuição espacial e da localização dos servidores, implementada no modelo hipercubo por meio de listas de preferência de despachos correspondentes (ver hipótese 7 na seção3). A escolha do servidor

preferencial no SAMU-RP é feita analisando a origem e a gravidade do chamado, e a preferência é dada ao servidor localizado na mesma área. Se o chamado for do tipo *a*, o VSA é imediatamente enviado ao local da ocorrência. Caso o VSA esteja ocupado, o VSB mais próximo do local do chamado é enviado. Se o chamado for de qualquer um dos tipos *b* ou *c*, o VSB disponível mais próximo do local é enviado. Caso todos os servidores da área do chamado estiverem ocupados, é escolhido aleatoriamente o primeiro servidor *backup* disponível da lista de despacho. Tanto o VSA quanto os VSBs podem se deslocar para qualquer átomo. Se todos os servidores estiverem ocupados, o chamado entra em fila e fica aguardando por um servidor disponível.

4.3.4 Tempos de atendimento dos servidores

Os tempos de atendimento foram obtidos a partir do intervalo de tempo entre os instantes de saída e retorno à base. Esses valores estão disponíveis nas fichas de regulação médica. Os tempos de atendimento incluem o tempo de preparo da equipe, o tempo de viagem até o local da ocorrência, o tempo em cena e o tempo de viagem de retorno à base. Os valores estão apresentados na Tabela 3 (em minutos). Para verificar estatisticamente a hipótese do modelo hipercubo de que os tempos de atendimento sejam exponencialmente distribuídos (ver hipótese 8 na seção 3), foi aplicado o teste de Kolmogorov-Smirnov do *software* BestFit. Esta hipótese foi rejeitada com $\alpha = 5\%$. Desta maneira, a utilização de tempos de atendimento com distribuições exponenciais no modelo hipercubo é uma simplificação que pode comprometer os resultados da análise, o que é estudado na próxima

seção. Além disso, as análises mostraram que as diferenças estatísticas entre as médias dos tempos de atendimento dos servidores são significativas nos três períodos e em todas as ambulâncias, com $\alpha = 5\%$ de significância. Dessa forma, a aplicação do modelo hipercubo considerou que os servidores não são homogêneos nos três períodos considerados (ver hipótese 4 do modelo hipercubo na seção 3). De acordo com Larson e Odoni (2007), para que o modelo hipercubo possa ser aplicado, os tempos médios de viagem devem ser suficientemente pequenos em relação aos tempos médios de serviço para cada ambulância, para que as variações nos tempos de atendimento, que são devidas às variações nos tempos de viagem, possam ser consideradas como sendo de segunda ordem, se comparadas com as variações dos tempos de atendimento (ver hipótese 9 do modelo hipercubo na seção 3). A Tabela 4 mostra o tempo médio de serviço, o tempo médio de viagem e a relação entre o tempo médio de serviço e o tempo médio de viagem para cada servidor. Pode-se notar que os tempos médios de viagem são relativamente pequenos com relação aos tempos médios de serviço. Para mais detalhes destas e outras análises realizadas sobre os processos de chegada e serviço do SAMU-RP, e destas e outras considerações e análises estatísticas envolvidas para aplicação do modelo hipercubo, o leitor pode consultar Souza (2010).

5 Aplicação do modelo hipercubo

O modelo hipercubo envolve a solução de um sistema linear de 2^n (n é o número de servidores), cujas variáveis são as probabilidades de equilíbrio de cada estado do sistema. O modelo hipercubo implementado para este trabalho é o mesmo que em Takeda, Widmer e Morabito (2007), que utilizou a

Tabela 4. Relação entre o tempo de atendimento e o tempo de viagem para as ambulâncias.

Ambulância	Tempo médio de serviço (minutos)			Tempo médio de viagem (minutos)			Relação: Tempo médio de viagem/tempo médio de serviço		
	Manhã	Tarde	Noite	Manhã	Tarde	Noite	Manhã	Tarde	Noite
1	41	47	34	10	10	10	0,25	0,21	0,29
2	44	30	21	10	9	8	0,22	0,29	0,40
3	44	30	21	10	9	8	0,22	0,29	0,40
4	61	46	45	12	12	11	0,20	0,26	0,24
5	61	46	45	12	12	11	0,20	0,26	0,24
6	51	33	27	9	10	9	0,18	0,29	0,34
7	51	33	27	9	10	9	0,18	0,29	0,34
8	80	38	36	12	10	10	0,15	0,27	0,28
9	47	40	26	9	9	8	0,19	0,23	0,33
10	47	40	-----	9	9	-----	0,19	0,23	-----
VSA	41	47	34	10	10	10	0,25	0,21	0,29
VSB	54	37	31	10	10	9	0,19	0,27	0,32
Total	53	38	31	10	10	9	0,20	0,26	0,32

estratégia de *layering* que incorpora prioridade nos estados do hiper-cubo e, para chamados em fila, a disciplina FCFS. Diversas fórmulas são derivadas a partir das probabilidades de equilíbrio obtidas da solução deste sistema linear para estimar as medidas de desempenho de interesse do sistema, que podem ser encontradas em detalhes em Larson e Odoni (2007) e Souza (2010). Os resultados da análise do SAMU-RP apresentados a seguir foram obtidos por meio da solução desse sistema de equações de equilíbrio para cada um dos períodos: manhã, tarde e noite.

Um modelo de simulação detalhado foi desenvolvido e implementado utilizando o *software* ARENA® para se avaliar os possíveis desvios dos resultados aproximados do modelo hiper-cubo. Este modelo de simulação representa um sistema de atendimento emergencial com todas as características do SAMU-RP, descritas na Seção 2. A simulação foi realizada considerando as distribuições mais aderentes para os tempos de atendimentos, além de considerar prioridade na fila de espera por classe de usuário. Por exemplo, nos testes realizados com o *software* BestFit®, a distribuição Lognormal foi a que melhor

aderiu para descrever os tempos de atendimentos de várias ambulâncias, sendo então utilizada no modelo de simulação (conforme hipótese 8 na seção 3, o modelo hiper-cubo admite que estes tempos sejam exponencialmente distribuídos). Os detalhes da implementação computacional do modelo hiper-cubo e do modelo de simulação podem ser encontrados em Souza (2010).

5.1 Resultados para o SAMU-RP

A Figura 7 mostra a localização das ambulâncias nas cinco bases do SAMU-RP, nos períodos da manhã e tarde (apenas uma ambulância opera na Região Oeste no período da noite). A Figura 8 mostra o comportamento das cargas de trabalho (*workloads*) de todas as ambulâncias, nos três períodos estudados. Vale salientar que dados incompletos ou inconsistentes (como, por exemplo, tempos de viagem excessivamente pequenos ou grandes - *outliers*) do controle dos chamados ou do controle de tráfego foram eliminados da amostra, de forma que a comparação dos *workloads* do modelo e da amostra não pôde ser realizada sem prejuízo da análise, uma vez que o *workload* da amostra seria subestimado porque esses períodos em que as ambulâncias estavam ocupadas na amostra não seriam considerados.

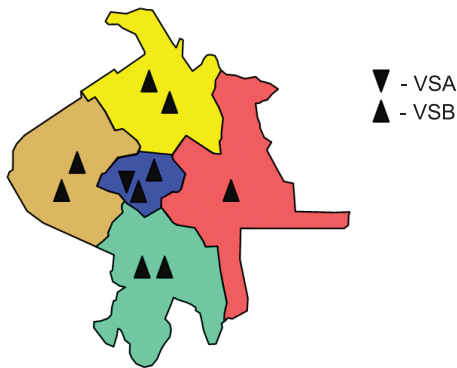


Figura 7. Localização das ambulâncias no SAMU-RP.

5.1.1 Medidas de desempenho do SAMU-RP

A probabilidade de encontrar o sistema vazio e a probabilidade de todos os servidores estarem ocupados foram computadas e os valores foram de 0,0002 e 0,0020, respectivamente. A probabilidade de fila no sistema e o tempo de espera na fila também foram calculados e resultaram em 0,0009 e 0,0205 minutos, respectivamente. Os tempos médios de espera em fila para todos os chamados do sistema calculados pelo modelo hiper-cubo e pela simulação não foram

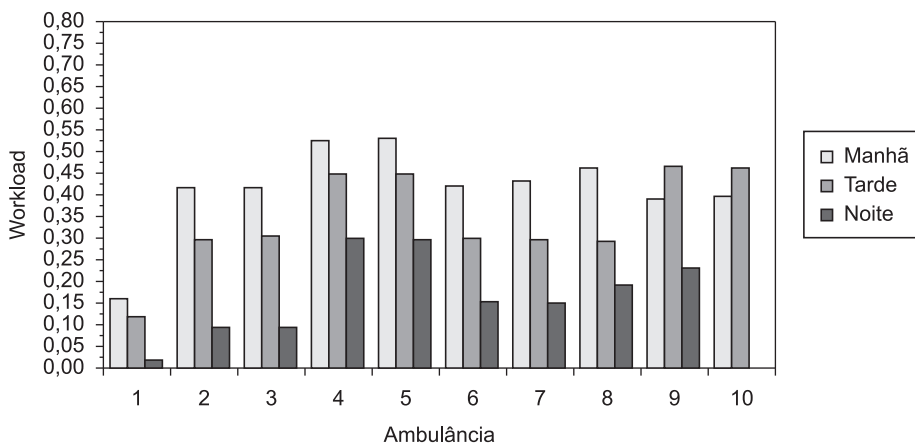


Figura 8. Workloads das ambulâncias.

significativos (menores que 10^{-2} minutos). Pode-se notar que os desvios absolutos do modelo em relação à simulação para o tempo médio de espera na fila não são grandes. A Tabela 5 mostra os *workloads* dos três períodos analisados. Os resultados do modelo no período da tarde são comparados com os resultados da simulação, uma vez que esse era o período com maior quantidade de dados disponíveis, viabilizando a realização dos testes de aderência. Pode-se verificar que o desvio percentual da ambulância 1 (VSA) é alto, 38,21%. Isso pode ser devido à dificuldade de se obter uma distribuição de probabilidade que represente bem o processo de atendimento desta ambulância, dado que a amostra continha uma pequena quantidade de dados para essa ambulância (apenas 5 observações).

Pode-se observar que os *workloads* da manhã são maiores que os do período da tarde. Esse comportamento é devido ao fato de que, apesar da taxa de chegada dos chamados ser maior no período da tarde, a taxa de serviço nesse período também é bem maior que no período da manhã. Dessa forma, o tempo de atendimento das ambulâncias no período da manhã é maior que no período da tarde, fazendo com que a taxa de utilização das ambulâncias seja maior no período da manhã. Em uma análise mais imediata, quando apenas a taxa de chegada seria observada, o período da tarde seria o escolhido. A probabilidade de fila no sistema é muito baixa (0,0038), fazendo com que os tempos de viagem sejam muito próximos dos tempos de resposta no sistema (i.e., os tempos de espera em fila são próximos de zero).

Conforme esperado, o período da noite é o menos gerenciado dos três analisados, com probabilidade de fila menor que 10^{-4} , de forma que os tempos de viagem e de resposta são coincidentes. O *workload* médio dos VSBs é 0,19, enquanto que do VSA é 0,02. A Tabela 6 mostra os resultados dos tempos médios

de viagem para cada subátomo, obtidos pelo modelo, pela simulação e pela amostra. Em geral os desvios do modelo em relação à simulação e à amostra são relativamente baixos (tipicamente menores que 5%), de maneira que o modelo pode ser considerado validado com base nos resultados da simulação e da amostra. Análises das outras medidas de desempenho para o SAMU-RP podem ser encontradas em Souza (2010).

5.2 Cenários alternativos

Os próximos cenários consideram uma ambulância a menos em cada região para o atendimento das classes *a*, *b* e *c*. Desta maneira, analisa-se o impacto da perda de uma ambulância no sistema devido a algum imprevisto de operação, ou a realocação desta ambulância para, por exemplo, fazer atendimentos de remoção. Outros cenários também poderiam ter sido analisados, como: deslocar ambulâncias entre as regiões para analisar outras configurações alternativas com mesmo número de ambulâncias, ou mesmo analisar cenários incluindo uma nova ambulância no sistema, para avaliar os benefícios obtidos. Em todos os casos a análise por meio do modelo hipercubo seria similar.

O critério para a escolha da ambulância a ser desconsiderada consistiu em simplesmente retirar uma ambulância do átomo que apresenta a menor taxa de chamados, desde que a base contenha mais de uma ambulância do mesmo tipo. Vale salientar que um procedimento de calibração dos tempos médios de serviço (consequentemente, das taxas médias de serviço) foi utilizado para se analisar estes cenários alternativos. Por meio deste procedimento, ajustaram-se os tempos médios de viagem separadamente de cada servidor, de forma a considerar os fatores geográficos que influenciam a viagem de cada servidor. Uma

Tabela 5. *Workload* das ambulâncias.

<i>Workload</i> Ambulância	Manhã	Noite	Tarde		
	Modelo	Modelo	Modelo	Simulação	Desvio (%)
1	0,16	0,02	0,12	0,1942	-38,21
2	0,42	0,09	0,3	0,2895	3,63
3	0,42	0,1	0,31	0,2888	7,34
4	0,53	0,3	0,45	0,443	1,58
5	0,53	0,3	0,45	0,44	2,27
6	0,42	0,15	0,3	0,2958	1,42
7	0,43	0,15	0,3	0,2938	2,11
8	0,46	0,19	0,29	0,2838	2,18
9	0,39	0,23	0,47	0,4619	1,75
10	0,39	...	0,46	0,4609	-0,2
VSA	0,16	0,02	0,12	0,1942	-38,21
VSB	0,44	0,19	0,37	0,3619	2,45
Total	0,42	0,17	0,345	0,3452	-1,61

Tabela 6. Tempos médios de viagem para cada subátomo no período da tarde.

Subátomo	Tempo médio de viagem (m.)			Desvio em relação à simulação		Desvio em relação à amostra	
	Modelo	Simulação	Amostra	Minutos	%	Minutos	%
1a	8,915	8,744	9,000	0,171	1,951	-0,0850	-0,944
1b	8,846	8,890	8,481	-0,043	-0,488	0,3652	4,306
1c	8,730	8,892	8,583	-0,162	-1,821	0,1471	1,714
2a	10,722	9,753	10,250	0,969	9,935	0,4720	4,605
2b	10,978	10,939	11,464	0,039	0,354	-0,4861	-4,240
2c	10,469	10,946	11,125	-0,477	-4,357	-0,6561	-5,898
3a	8,941	10,532	9,000	-1,591	-15,104	-0,0589	-0,654
3b	9,041	8,894	8,700	0,147	1,656	0,3411	3,921
3c	9,598	8,894	9,500	0,704	7,912	0,0975	1,026
4a	10,879	11,239	11,000	-0,360	-3,199	-0,1209	-1,099
4b	10,941	10,607	9,750	0,334	3,148	1,1907	12,212
4c	10,945	10,607	10,143	0,338	3,189	0,8021	7,908
5a	9,204	9,822	9,000	-0,618	-6,288	0,2044	2,271
5b	9,097	9,500	9,182	-0,403	-4,242	-0,0852	-0,928
5c	9,620	9,499	9,278	0,121	1,278	0,3420	3,686
Média a	9,732	10,018	9,650	-0,286	-2,541	0,0823	0,836
Média b	9,781	9,766	9,515	0,015	0,086	0,2651	3,054
Média c	9,872	9,767	9,726	0,105	1,240	0,1465	1,687

descrição detalhada deste procedimento pode ser encontrada em Larson e Odoni (2007) (ver também discussão de convergência deste procedimento em Chiyoshi, Galvão e Morabito (2000).

Na avaliação dos três cenários alternativos desta seção, analisamos subperíodos do dia que podem ter diferentes características temporais e geográficas importantes de serem consideradas. Para isto, utilizamos a abordagem baseada no modelo hipercubo para sistemas em equilíbrio, estudando os períodos escolhidos de forma independente. Por meio desta abordagem, analisamos a configuração do SAMU-RP a partir das medidas de desempenho obtidas pelo modelo para cada período crítico do dia.

5.2.1 Cenário 1 – Período da manhã com uma ambulância a menos

O cenário 1 representa o período da manhã com uma ambulância a menos na Região Oeste. A localização das ambulâncias para este cenário está representada na Figura 9. Os resultados mostraram que o aumento dos tempos médios de espera na fila no cenário 1 em relação ao SAMU-RP, em minutos, não são significativos com relação aos desvios absolutos (menores que 0,3 minutos). A Tabela 7 contém os *workloads* das ambulâncias para o cenário 1 e a comparação com o SAMU-RP. Neste cenário, os *workloads* aumentaram em média 10,4% nos VSBs, 21,6% no VSA e 11,6% nas ambulâncias em geral. Além disso, os resultados mostraram que há um

aumento dos *workloads* das ambulâncias no cenário 1 em relação ao SAMU-RP, com desvios relativos de, no mínimo, 4,2% e, no máximo, 23,5%. Com a análise do cenário 1, de forma geral, a diminuição de uma ambulância na Região Oeste no período da manhã não tem grande impacto nas medidas de desempenho analisadas, principalmente considerando os desvios absolutos no tempo de resposta ao usuário.

5.2.2 Cenário 2 – Período da tarde com uma ambulância a menos

O cenário 2 representa o período da tarde com uma ambulância a menos na Região Sul. A localização das ambulâncias para este cenário está representada na Figura 8. Os tempos médios de espera na fila para o cenário 2 não são significativos (menores que 0,08 minutos). Os resultados mostraram que o aumento dos tempos médios de espera na fila no cenário 2 em relação ao SAMU-RP, em minutos, não são significativos com relação aos desvios absolutos. As porcentagens dos desvios são bastante altas, de, no mínimo, 77,5% e, no máximo, 78,7%. A Tabela 10 contém os *workloads* das ambulâncias para o cenário 2 e a comparação com o SAMU-RP. Neste cenário, os *workloads* aumentaram em média 19,3% nos VSBs, 22,0% no VSA e 19,6% nas ambulâncias em geral. Além disso, os resultados indicaram que há aumento dos *workloads* das ambulâncias no cenário 2 em relação ao SAMU-RP, com desvios relativos de, no mínimo, 5,9% e, no máximo, 34,7%. Similarmente

ao cenário 1, a diminuição de uma ambulância na Região Oeste no período da manhã não tem grande impacto nas medidas de desempenho analisadas, principalmente considerando os desvios absolutos no tempo de resposta ao usuário, com exceção do subátomo 4b.

5.2.3 Cenário 3 – Período da noite com uma ambulância a menos

O cenário 3 representa o período da noite com uma ambulância a menos na Região Central. A localização das ambulâncias para este cenário está representada na Figura 11. A Tabela 9 contém os *workloads* das ambulâncias para o cenário 3 e a comparação com o SAMU-RP. Neste cenário, os *workloads* aumentaram em média 18,3% nos VSBs, 9,5% no VSA e 17,2% nas ambulâncias em geral. Além disso, os resultados mostraram que há aumento dos *workloads* das ambulâncias no cenário 3 em relação ao SAMU-RP, com desvios relativos de, no mínimo, 9,5% e, no máximo, 61,04%. As análises

do cenário 3 são similares às dos cenários 1 e 2. Os tempos médios de espera na fila para o cenário 3 não são significativos (menores que 0,02 minutos). Os resultados mostram que o aumento dos tempos médios de espera na fila no cenário 3 em relação ao SAMU-RP, em minutos, não são significativos com relação aos desvios absolutos. Porém, as porcentagens dos desvios relativos são bastante altas, de, no mínimo, 90,9% e, no máximo, 93,1%. Com a análise do cenário 3, de forma geral, a diminuição de uma ambulância na Região Oeste no período da manhã não tem grande impacto nas medidas de desempenho analisadas, principalmente considerando os desvios absolutos no tempo de resposta ao usuário, com exceção do subátomo 4b. Como mencionado antes, a ambulância retirada da análise pode ser realocada para fazer somente atendimentos de remoção, dada a importância social desse serviço.

A partir das análises das medidas de desempenho do sistema SAMU-RP realizadas para cada período, é possível determinar múltiplas alternativas de configuração de localização de ambulâncias nesses

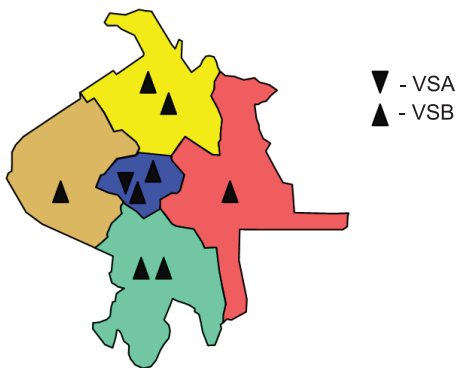


Figura 9. Localização das ambulâncias no cenário 1.

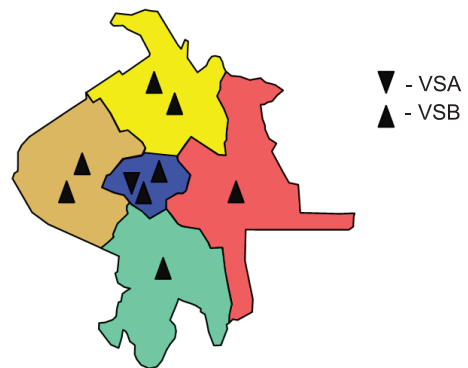


Figura 10. Localização das ambulâncias no cenário 2.

Tabela 7. *Workloads* do cenário 1.

Ambulância	Workload		Desvio (%)
	cenário 1	SAMU-RP	
1	0,1945	0,1600	21,6
2	0,4548	0,4149	9,6
3	0,4761	0,4187	13,7
4	0,5570	0,5290	5,3
5	0,5500	0,5280	4,2
6	0,4646	0,4265	8,9
7	0,4705	0,4312	9,1
8	0,4997	0,4590	8,9
9	0,4868	0,3943	23,5
10	...	0,3902	...
VSA	0,1945	0,1600	21,6
VSB	0,4949	0,4435	10,4
Total	0,4616	0,4152	11,6

Tabela 8. *Workloads* do cenário 2.

Ambulância	Workload		Desvio (%)
	Cenário 2	SAMU-RP	
1	0,1464	0,1200	22,0
2	0,3599	0,3021	19,1
3	0,3464	0,3051	13,5
4	0,4793	0,4526	5,9
5	0,5031	0,4494	11,9
6	0,3503	0,2989	17,2
7	...	0,2977	...
8	0,4009	0,2831	34,7
9	0,3681	0,4637	30,0
10	0,5669	0,4637	22,3
VSA	0,1464	0,1200	22,0
VSB	0,4219	0,3773	19,3
Total	0,3913	0,3436	19,6

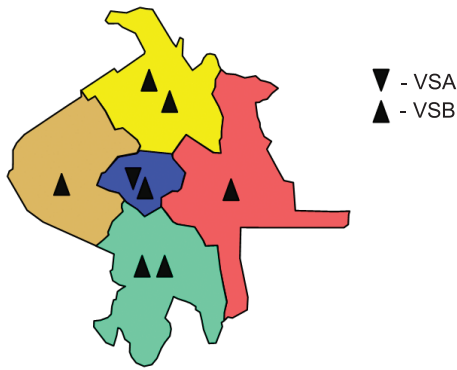


Figura 11. Localização das ambulâncias no cenário 3.

Tabela 9. Workloads do cenário 3.

Ambulância	Workload		Desvio (%)
	cenário 3	SAMU-RP	
1	0,0196	0,0179	9,5
2	0,1502	0,0941	59,6
3	...	0,0942	...
4	0,2364	0,2971	-20,4
5	0,3560	0,2972	19,8
6	0,2451	0,1522	61,0
7	0,1739	0,1505	15,5
8	0,2357	0,1962	20,1
9	0,2017	0,2296	-12,2
VSA	0,0196	0,0179	9,5
VSB	0,2036	0,1889	18,3
Total	0,1806	0,1699	17,2

períodos, melhorando tanto o desempenho do sistema do ponto de vista dos usuários (medidas externas) como o desempenho do ponto de vista das medidas internas da gerência do sistema. Os demais detalhes dessas análises encontram-se em Souza (2010), assim como análises da aplicação do modelo em outros cenários do SAMU-RP.

5.3 Análise do tempo de resposta do modelo hipercubo para as classes de usuários

O tempo médio de resposta aos usuários depende da taxa de utilização do sistema, dado que os chamados aguardam em fila de espera se todos os servidores estiverem ocupados. Como o tempo de resposta inclui o tempo de espera na fila, para sistemas mais congestionados, é importante considerar o tempo de espera para cada classe de usuários. A Figura 12 mostra as diferenças dos tempos médios de espera na fila por classe de usuários, utilizando o modelo hipercubo, que considera disciplina de fila FCFS (linha azul da Figura 12), e utilizando o modelo de simulação, que considera prioridade na fila por classe de usuários e calcula o tempo médio de espera na fila para as três classes (linhas verde, vermelha e preta da Figura 12). O limite de tamanho da fila foi arbitrariamente fixado em 10 usuários.

Por meio desta análise, pretende-se verificar até que ponto (i.e., até qual taxa de utilização) pode-se utilizar o modelo hipercubo utilizando a estratégia de *layering* (sem prioridade na fila) para se analisar um sistema com prioridade. A configuração do

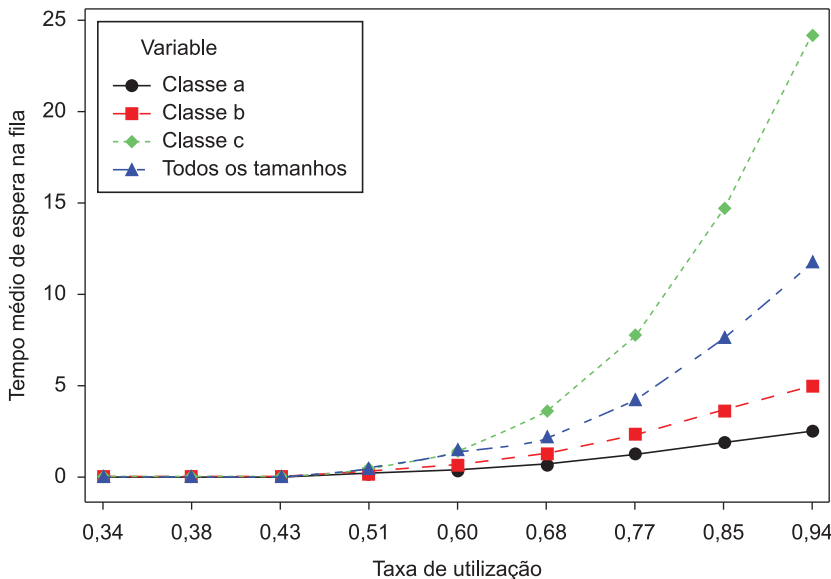


Figura 12. Tempos médios de espera na fila para todos os chamados e para as classes a, b e c.

SAMU-RP apresenta a taxa média de utilização de 0,34. Os resultados mostram que, quando a taxa de utilização ultrapassa 50%, as diferenças dos tempos médio de espera para os chamados das três classes começam a ser significativas. Note que os tempos de espera são superestimados no modelo hipercubo para os chamados das classes *a* e *b*, enquanto que para os chamados da classe *c* estes tempos são subestimados. Essa análise reforça a importância de considerar prioridade na fila do modelo hipercubo em sistemas em que a taxa de utilização é alta o suficiente para a formação de fila.

6 Conclusões

Neste estudo, avaliou-se, por meio de um estudo de caso no SAMU-RP, a aplicabilidade do modelo hipercubo para analisar sistemas de atendimento médico de urgência em áreas urbanas. Investigaram-se também possibilidades de múltiplas alternativas de localização de ambulâncias ao longo do dia de operação do sistema, considerando as variações na demanda e no serviço do sistema. Essa análise reforçou a importância de se considerar diferentes alternativas de localização de ambulâncias no sistema para cada período do dia: manhã, tarde e noite. Em particular, mostrou-se que, caso a análise fosse feita com base apenas no período de pico de demanda do dia, o período escolhido seria a tarde. No entanto, apesar de o período da tarde ser o período com maior demanda, o período da manhã foi o mais desfavorável do ponto de vista de certas medidas de congestão do sistema, pois o tempo de atendimento dos chamados é maior nesse período. Além da configuração do sistema SAMU-RP, foram também estudados alguns cenários alternativos para se avaliar diferentes aspectos desse sistema.

Na configuração estudada do SAMU-RP, o sistema opera com 10 ambulâncias nos períodos da manhã e tarde e 9 ambulâncias no período da noite, descentralizadas da seguinte maneira: 1 VSA e 2 VSBs na Região Central, 2 VSBs na Região Norte, 2 VSBs na Região Sul, 2 VSBs na Região Oeste e 1 VSB na Região Leste. Os resultados obtidos pelo modelo hipercubo com a análise dos cenários alternativos 1, 2 e 3 mostraram que, desconsiderando os atendimentos de remoção de pacientes, no período da manhã, o SAMU-RP poderia operar com um VSB a menos na Região Oeste; no período da tarde, com um VSB a menos na Região Sul; e no período da noite, com um VSB a menos na Região Central, sem maiores prejuízos nas medidas de desempenho internas e externas do sistema. As ambulâncias não utilizadas nestas configurações alternativas do sistema poderiam ser utilizadas para, por exemplo, realizar atendimentos agendados de remoção.

Uma pesquisa interessante a partir deste estudo seria incorporar políticas de prioridade na fila do

modelo hipercubo, para que o modelo represente de forma mais acurada sistemas mais congestionados em função de aumentos nas demandas dos usuários, ou em situações em que o agendamento da remoção de pacientes comprometa a disponibilidade das ambulâncias, tornando o sistema mais congestionado. Pretende-se reportar em breve os resultados desta pesquisa.

Agradecimentos

Os autores agradecem aos dois revisores anônimos os úteis comentários e sugestões. Também agradecem ao SAMU-RP a colaboração com esta pesquisa e ao CNPq o apoio financeiro.

Referências

- BRANDEAU, M.; LARSON, R. C. Extending and applying the hypercube queueing model to deploy ambulances in Boston. In: SWERSEY, A. J.; INGNALL, E. J. (Eds.). **Delivery of Urban Services**. TIMS Studies in the Management Science, Elsevier, 1986. v. 22, p. 121-153.
- BURWELL, T. H.; JARVIS J. P.; McKNEW, M. A. Modeling co-located servers and dispatch ties in the hypercube model. **Computers & Operations Research**, v. 20, n. 2, p. 113-119, 1993. [http://dx.doi.org/10.1016/0305-0548\(93\)90067-S](http://dx.doi.org/10.1016/0305-0548(93)90067-S)
- CHELST, K. R.; BARLACH, Z. Multiple unit dispatches in emergency services: models to estimate system performance. **Management Science**, v. 27, n. 12, p.1390-1409, 1981. <http://dx.doi.org/10.1287/mnsc.27.12.1390>
- CHIYOSHI, F.; GALVÃO, R. D.; MORABITO, R. O uso do modelo hipercubo na solução de problemas de localização probabilísticos. **Gestão & Produção**, v. 7, n. 2, p.146-174, 2000. <http://dx.doi.org/10.1590/S0104-530X2000000200005>
- COSTA, D. M. **Uma metodologia iterativa para determinação de zonas de atendimento de serviços emergenciais**. 2004. Tese (Doutorado em Engenharia de Produção)-Universidade Federal de Santa Catarina, Florianópolis, 2004.
- FIGUEIREDO, A. P. S.; LORENA, L. A. N. Localização de ambulâncias: uma aplicação para a cidade de São José dos Campos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO – SBSR, 12., 2005, Goiânia. **Anais...** Goiânia: INPE, 2005. p. 1965-1972.
- 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. <http://dx.doi.org/10.1111/j.1475-3995.2008.00654.x>
- GONÇALVES, M. B.; NOVAES, A. G.; ALBINO, J. C. C. Modelos para localização de serviços emergenciais em rodovias. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL – SBPO, 26. 1994, Florianópolis. **Anais...** Florianópolis, 1994. p. 591-596.
- GONÇALVES, M. B.; NOVAES, A. G.; SCHMITZ, R. Um modelo de otimização para localizar unidades de serviço emergenciais em rodovias. In: CONGRESSO DE

- PESQUISA E ENSINO EM TRANSPORTES, 9., 1995, São Carlos. **Anais...** São Carlos, 1995. p. 962-972.
- IANNONI, A. P.; MORABITO, R. Modelo hipercubo integrado a um algoritmo genético para análise de sistemas médicos emergenciais em rodovias. **Gestão & Produção**, v. 13, n. 1, p. 93-104, 2006. <http://dx.doi.org/10.1590/S0104-530X2006000100009>
- IANNONI, A. P.; MORABITO, R. A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways. **Transportation Research E**, v. 43, n. 6, p. 755-771, 2007. <http://dx.doi.org/10.1016/j.tre.2006.05.005>
- 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. [http://dx.doi.org/10.1016/0305-0548\(74\)90076-8](http://dx.doi.org/10.1016/0305-0548(74)90076-8)
- LARSON, R. C.; ODoni, A. R. **Urban Operations Research**. 2nd ed. Belmont: Dynamic Ideas, 2007.
- LUQUE, L. **Análise da aglutinação de estados em cadeias de markov do modelo hipercubo de filas com servidores co-localizados**. 2006. Dissertação (Mestrado)-Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2006.
- MENDONÇA, F.; MORABITO, R. Analysing emergency medical service ambulance deployment on a Brazilian highway using the hypercube model. **Journal of the Operational Research Society**, v. 52, p. 261-270, 2001.
- MORABITO, R.; CHIYOSHI, F.; GALVÃO, R. Non-homogeneous servers in emergency medical systems: practical applications using the hypercube queuing model. **Socio-Economic Planning Sciences**, v. 42, p. 255-270, 2008. <http://dx.doi.org/10.1016/j.seps.2007.04.002>
- OLIVEIRA, L. K. **Uma aplicação do modelo hipercubo de filas para avaliação do centro de emergência da polícia militar de Santa Catarina**. 2003. Dissertação (Mestrado em Engenharia de Produção)-Universidade Federal de Santa Catarina, Florianópolis, 2003.
- SACKS, S. R.; GRIEF, S. **Orlando Police Department uses OR/MS methodology, new software to design patrol districts**. Baltimore: OR/MS Today, 1994. p. 30-32.
- 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.
- SOUZA, R. M. **Análise da configuração de SAMU utilizando modelo hipercubo com prioridade na fila e múltiplas alternativas de localização de ambulâncias**. 2010. Tese (Doutorado em Engenharia de Produção)-Universidade Federal de São Carlos, São Carlos, 2010.
- SWERSEY, A. J. **Handbooks in OR/MS**. 6. v. Amsterdam: Elsevier Science B.V., 1994. p. 151-200.
- TAKEDA, R. A.; WIDMER, J. A.; MORABITO, R. Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queuing model. **Computers & Operations Research**, v. 34, p. 727-741, 2007. <http://dx.doi.org/10.1016/j.cor.2005.03.022>