



Universidade Federal de Minas Gerais

Programa de Pós-Graduação em Engenharia de Produção



Metodologia para Dimensionamento e Análise de Serviços de Atendimento de Emergência

Doutorando:

Pedro Marinho Sizenando Silva

Professor Orientador:

Prof. Dr. Luiz Ricardo Pinto

Banca Examinadora:

Prof. Dr. Luiz Ricardo Pinto (orientador)

Prof. Dr. Ricardo Saraiva de Camargo (DEP-UFMG)

Prof. Dra. Leise Kelli de Oliveira (ETG-UFMG)

Prof. Dr. Carlos Andrey Maia (DEE-UFMG)

Prof. Dr. José Arnaldo Barra Montevechi (IEPG-UNIFEI)

Prof. Dr. Alexandre Xavier Martins (ICEA-UFOP)

Belo Horizonte

2015

Pedro Marinho Sizenando Silva

Metodologia para Dimensionamento e Análise de Serviços de Atendimento de Emergência

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Doutor em Engenharia de Produção.

Área de Concentração: Produção e Logística

Orientador: Prof. Dr. Luiz Ricardo Pinto

Belo Horizonte

2015

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Luiz Ricardo Pinto, pelos conselhos e ajuda que vêm desde a época da graduação e que foram determinantes para a conclusão deste trabalho.

Aos professores do Programa de Pós-Graduação em Engenharia de Produção da UFMG pelos conhecimentos passados ao longo dos últimos anos.

Aos meus pais, presentes em todos os momentos de minha vida, referência de estudo, responsabilidade e família.

À minha esposa, pelo amor, carinho e paciência durante esta caminhada. Amo você.

Aos colegas do Programa de Pós-Graduação em Engenharia de Produção da UFMG pelas conversas descontraídas, incentivos e conselhos.

Aos professores colegas do Instituto Federal de Minas Gerais (IFMG) pelo apoio que sempre me deram na condução do doutorado.

“Realize suas tarefas de ordem física com toda a eficiência e dedicação, mas recorde que não é esse o único fim da vida, nem mesmo no melhor dos casos.”

da Ciência Logosófica

RESUMO

Desde a década de 1970 existem estudos que demonstram a existência de uma correlação direta entre o tempo de resposta (tempo transcorrido entre a chegada da chamada na central e a chegada da equipe médica no local do incidente) de serviços de atendimento móveis de emergência e a probabilidade de sobrevivência das vítimas envolvidas em acidentes. Torna-se evidente que qualquer ferramenta capaz de proporcionar um ganho operacional para este tipo de sistema com relação ao tempo de resposta é de grande valia para os gestores e a população em geral. Este trabalho tem como objetivo o desenvolvimento de uma metodologia baseada em otimização via simulação com o objetivo de dimensionar sistemas de atendimento móveis de emergência. A metodologia está baseada em três aspectos principais: para análise das configurações será proposto um componente baseado nas técnicas de simulação de eventos discretos; para geração dos cenários serão propostos novos procedimentos baseados nas meta-heurísticas busca tabu e *simulated annealing*; e para aceleração do processo de otimização serão utilizados metamodelos baseados em regressão múltipla para filtragem das soluções candidatas. A premissa básica a ser verificada é de que o processo de otimização via simulação pode ser acelerado e gerar melhores resultados quando integrado com o uso de metamodelos. Este procedimento nunca foi utilizado no contexto de sistemas de atendimento de emergência. A metodologia proposta resolve o problema de localização de ambulâncias a partir de bases potenciais e o problema de dimensionamento destas bases, determinando a quantidade ideal de unidades em cada. Utilizou-se dois tipos diferentes de ambulâncias: básicas, para atendimento de ocorrências menos graves e avançadas, para atendimento de ocorrências de maior gravidade. A otimização está baseada na minimização de custos de abertura das bases e aquisição das ambulâncias. As restrições estabelecidas consideram um tempo de resposta máximo para o sistema e um limite de ambulâncias alocadas em cada base. A metodologia proposta é validada utilizando dados do Serviço de Atendimento Móvel de Urgência da cidade de Belo Horizonte (Minas Gerais) e testes são feitos para identificar a qualidade dos procedimentos de otimização utilizados. Os procedimentos propostos foram comparados com o otimizador *OptQuest* da *OptTek Systems*, integrado com o simulador *Arena* da *Rockwell Automation Technologies*. Em todos os cenários analisados, a melhor configuração fornecida pela metodologia proposta superou a melhor configuração fornecida pelo *OptQuest* em termos de custos, atingindo o mesmo padrão de tempo de resposta com menos bases abertas e menos ambulâncias em operação.

Palavras-chave: Localização e Dimensionamento de Sistemas de Ambulância, Otimização via Simulação, Métodos Heurísticos, Metamodelos.

ABSTRACT

Since the 70s there are studies that shows the existence of a direct correlation between the response time (time from when the call is first dispatched until the ambulance reaches the scene) of emergency medical systems and the probability of survival for victims involved in accidents. It is clear that any tool capable of providing an operational improvement for this type of system regarding the response time is very valuable to managers and the general population. This study aims to develop a methodology based on simulation optimization with the goal of dimensioning emergency medical systems. The methodology is based on three main aspects: a module based on discrete event simulation techniques for analyzing all the configurations; new procedures based on the meta-heuristics tabu search and simulated annealing for generating different scenarios; and the use of multiple regression metamodels to act as a filter and accelerate the process of simulation optimization. The basic premise to be verified is that the simulation optimization process can be accelerated and generate better results when integrated with the use of metamodels. This procedure has never been used in the context of emergency medical systems. The proposed methodology solves the ambulance location problem and the dimensioning problem, determining the optimal amount of units in each base. Two different types of ambulances were considered: basic life support unit, to respond to less severe occurrences and advanced life support unit, to respond to more severe occurrences. The optimization module is based on minimizing costs of opening bases and acquisition of ambulances. The restrictions consider a maximum response time for the system and a limit of ambulances allocated to each base. The proposed methodology is validated using data from the Emergency Medical System of the city of Belo Horizonte (Minas Gerais) and tests were conducted to identify the quality of the optimization procedures. The proposed procedures were compared with OptQuest optimizer from OptTek Systems, integrated with the simulator Arena from Rockwell Automation Technologies. In all scenarios analyzed, the best configuration provided by the proposed methodology outperforms the best configuration provided by OptQuest in terms of cost, reaching the same level of response time with less opened bases and ambulances.

Key words: Location and Dimensioning of Ambulance Systems, Optimization for Simulation, Heuristics, Metamodels.

ABREVIATURAS

ANOVA	<i>Analysis of Variance</i>
BT	Busca Tabu
CERSAM	Centro de Referência em Saúde Mental
IBGE	Instituto Brasileiro de Geografia e Estatística
MHL	Modelo Hipercubo Latino
SA	<i>Simulated Annealing</i>
SAMU	Serviço de Atendimento Móvel de Urgência
SUS	Sistema Único de Saúde
UPA	Unidade de Pronto Atendimento
USA	Unidade de Suporte Avançado
USB	Unidade de Suporte Básico
USM	Unidade de Saúde Mental
UTM	<i>Universal Transverse Mercator</i>
VBA	<i>Visual Basic for Applications</i>

LISTA DE FIGURAS

Figura 1: Fluxograma do processo de atendimento a uma chamada	26
Figura 2: Metodologia de simulação.....	38
Figura 3: Metodologia para construção do metamodelo.....	44
Figura 4: Metodologia de otimização via simulação	49
Figura 5: Metodologia proposta de otimização via simulação.....	50
Figura 6: Componentes principais da metodologia de simulação	53
Figura 7: Número de chamadas por faixa de horário (SAMU-BH - 2009)	55
Figura 8 - Alternativas para geração dos incidentes	56
Figura 9: Exemplo de mapeamento de uma regional	57
Figura 10: Abordagens para geração dos incidentes	58
Figura 11: Algoritmo para determinar se ponto está dentro de um polígono	59
Figura 12: Exemplo do problema de ponto no polígono	60
Figura 13: Lógica de acionamento e controle das ambulâncias	61
Figura 14: Relação entre locais de saída e chegada no processo de retorno da ambulância.....	63
Figura 15: Distribuição espacial das ambulâncias (esquerda) e centros de atendimento (direita) na cidade de Belo Horizonte para o ano de 2009	70
Figura 16: Evolução do número de chamadas (out/06 - ago/11).....	71
Figura 17: Número de chamadas por dia e faixa de horário (ago/09 - nov/09)	72
Figura 18: Distribuição das chamadas por regional (ano 2009).....	73
Figura 19: Período transiente e permanente para o tempo de resposta médio	78
Figura 20: Histograma da distribuição dos tempos de resposta (SAMU-BH 2009)	80
Figura 21: Distribuição espacial das ambulâncias (esquerda) e centros de atendimento (direita) na cidade de Belo Horizonte para o ano de 2011	82
Figura 22: Distribuição das chamadas por regional (ano 2011).....	84
Figura 23: Histograma da distribuição dos tempos de resposta (SAMU-BH 2011)	85
Figura 24: Tempo de resposta médio x aumento da demanda	86
Figura 25: Tempo de resposta médio x velocidade das ambulâncias	88
Figura 26: Análise de sensibilidade com o total de unidades básicas	90
Figura 27: Análise de sensibilidade com o total de unidades avançadas.....	91
Figura 28: Análise de sensibilidade com o total de unidades.....	91
Figura 29: Pseudocódigo do método baseado na Busca Tabu	94

Figura 30: Vizinhança 1 da Busca Tabu	95
Figura 31: Vizinhança 2 da Busca Tabu	95
Figura 32: Pseudocódigo do método baseado no <i>Simulated Annealing</i>	96
Figura 33: Vizinhança do método <i>Simulated Annealing</i>	97
Figura 34: Metamodelo para unidades básicas.....	101
Figura 35: Metamodelo para unidades avançadas.....	101

LISTA DE TABELAS

Tabela 1: Variáveis de interesse	54
Tabela 2: Coleta de dados para estimar tempo de duração de cada atividade	67
Tabela 3: Quantidade média de chamadas por dia e horário (ago/09 - nov/09).....	72
Tabela 4: Distribuição dos motivos de chamada (out/06 - ago/11).....	74
Tabela 5: Distribuição dos tipos de ocorrência (out/06 - ago/11)	74
Tabela 6: Distribuição das condutas médicas (out/06 - ago/11).....	75
Tabela 7: Distribuição do tipo de ambulância despachada por tipo de ocorrência	75
Tabela 8: Distribuições de probabilidade para cada atividade do SAMU-BH.....	76
Tabela 9: Validação do número de atendimentos realizados (ano 2009)	79
Tabela 10: Quantidade média de chamadas por dia e horário (mar/11 - jun/11)	83
Tabela 11: Resultados comparativos das heurísticas.....	105
Tabela 12: Distribuição das ambulâncias em cada base	107
Tabela 13: Resultados da metodologia proposta	110

SUMÁRIO

1. INTRODUÇÃO	12
1.1 JUSTIFICATIVA E MOTIVAÇÃO	13
1.2 OBJETIVOS ESPECÍFICOS	15
1.3 MOTIVAÇÃO PARA O USO DA SIMULAÇÃO	15
1.4 CONTRIBUIÇÃO DO TRABALHO	18
1.5 ESTRUTURA DO DOCUMENTO	22
2. CARACTERIZAÇÃO DE SERVIÇOS MÓVEIS DE EMERGÊNCIA.....	23
3. REVISÃO DE LITERATURA.....	27
3.1 METODOLOGIAS DE SIMULAÇÃO	27
3.2 MÉTODOS HEURÍSTICOS	31
3.3 METAMODELOS	33
4. METODOLOGIA PROPOSTA	37
4.1 METODOLOGIA DE SIMULAÇÃO	37
4.2 METODOLOGIA PARA CONSTRUÇÃO DO METAMODELO	43
4.3 METODOLOGIA DE OTIMIZAÇÃO VIA SIMULAÇÃO	48
5. APLICAÇÃO DA METODOLOGIA PROPOSTA.....	51
5.1 MÓDULO DE SIMULAÇÃO.....	52
5.1.1 CARACTERÍSTICAS DAS CHAMADAS	54
5.1.2 GERAÇÃO DOS INCIDENTES.....	55
5.1.3 DETERMINAÇÃO SE O INCIDENTE OCORREU DENTRO DA REGIONAL	58
5.1.4 LÓGICA DE DESPACHO E CONTROLE DAS AMBULÂNCIAS	60
5.1.5 LÓGICA DE RETORNO DAS AMBULÂNCIAS	62
5.1.6 TEMPOS DE DESLOCAMENTO DAS AMBULÂNCIAS	64
5.1.7 TEMPOS DAS ATIVIDADES.....	66
5.1.8 PARÂMETROS DA SIMULAÇÃO	67
5.1.9 ESTRATÉGIAS DE VALIDAÇÃO	68
5.1.10 VALIDAÇÃO DA METODOLOGIA DE SIMULAÇÃO	69
5.1.11 ANÁLISE DE SENSIBILIDADE COM O MODELO DE SIMULAÇÃO.....	85
5.2 MÉTODOS HEURÍSTICOS PROPOSTOS	92
5.3 METAMODELO DESENVOLVIDO	98
5.4 RESULTADOS COMPUTACIONAIS	103
6. CONCLUSÕES E TRABALHOS FUTUROS	113
REFERÊNCIAS BIBLIOGRÁFICAS.....	116

1. INTRODUÇÃO

Atendimento pré-hospitalar pode ser considerado como toda e qualquer assistência realizada, direta ou indiretamente, fora do âmbito hospitalar, através dos diversos meios e métodos disponíveis, com uma resposta adequada à solicitação, a qual poderá variar de um simples conselho ou orientação médica ao envio de uma viatura de suporte básico ou avançado ao local da ocorrência, visando a manutenção da vida e/ou a minimização das sequelas (LOPES E FERNANDES, 1999). Dentro deste contexto estão inseridos os Serviços de Atendimento Móveis de Urgência/Emergência, mais conhecidos no âmbito brasileiro como SAMU.

As duas principais decisões relativas ao dimensionamento deste tipo de sistema dizem respeito à quantidade ideal e melhor posicionamento das unidades móveis ao longo do território considerado. Apesar de existirem diversos outros fatores que influenciam no desempenho do sistema (características da população, intensidade do tráfego, etc.), estes dois são os mais importantes e sobre os quais os gestores possuem maior controle e poder de decisão.

Este trabalho tem como objetivo o desenvolvimento de uma metodologia para dimensionamento e análise de serviços de atendimento móveis de emergência baseada em três aspectos fundamentais:

1. Para análise de cenários e estimativa da função a ser otimizada, desenvolveu-se um componente de simulação baseado em eventos discretos de fácil implementação;
2. Para geração dos cenários a serem simulados, desenvolveu-se algumas heurísticas específicas para o contexto de sistemas de ambulância;
3. Para a integração da ferramenta de simulação com os métodos heurísticos, utilizou-se um metamodelo baseado em regressão linear múltipla, para ser um filtro: somente os cenários que resultarem em um tempo de resposta próximo do parâmetro utilizado como restrição serão analisados pelo modelo de simulação. Como a simulação das configurações é o processo mais demorado na metodologia, procura-se desta forma otimizar o processo de busca, ampliando o espaço de soluções analisadas.

As próximas seções da introdução têm como finalidade, nesta ordem, apresentar a justificativa e motivação para o tema escolhido, os objetivos específicos do trabalho, fundamentar a escolha da simulação de eventos discretos e as principais contribuições para a área de conhecimento em que o trabalho está inserido.

1.1 JUSTIFICATIVA E MOTIVAÇÃO

Desde a década de 1970 (Mayer, 1979) existem estudos que demonstram a existência de uma correlação direta entre o tempo de resposta (tempo transcorrido entre a chegada da chamada na central e a chegada da equipe médica no local do incidente) de serviços de atendimento móveis e a probabilidade de sobrevivência das vítimas envolvidas em acidentes. O’Keeffe *et al.* (2011) fizeram um estudo com 1.161 pacientes que sofreram parada cardíaca e foram atendidos por uma ambulância no Reino Unido, e conseguiram mostrar que a redução de 1 minuto no tempo de resposta aumentava a probabilidade de sobrevivência em até 24%. Os autores também afirmam que o custo estimado para reduzir o tempo de resposta médio do sistema em 1 minuto, para o ano de 2011, era de aproximadamente 54 milhões de libras. Isto mostra a importância do bom dimensionamento deste tipo de sistema, uma vez que ganhos no tempo de resposta representam na prática investimentos financeiros bastante elevados. Sladjana *et al.* (2011) desenvolveram um estudo na Sérvia em que concluíram que as maiores chances de sobrevivência de um paciente com parada cardíaca estavam relacionadas com tempos de resposta das ambulâncias de no máximo 4 minutos. Dessa forma, puderam afirmar que quanto maior o tempo de resposta, menor a probabilidade de sobrevivência das vítimas. Na mesma linha, Pons *et al.* (2005) analisaram 34.111 ocorrências entre os meses de janeiro e dezembro de 1998 na cidade de Denver, e concluíram que tempos de resposta menores que 4 minutos aumentavam a probabilidade de sobrevivência de pacientes com risco de vida classificado como alto ou moderado; resultado este em consonância com aquele apresentado em Sladjana *et al.* (2011). Wilde (2013) desenvolveu um estudo no estado de Utah (Estados Unidos) em que foi possível mostrar que o tempo de resposta afeta significativamente a mortalidade e probabilidade da vítima ser admitida em um hospital. Nesta mesma linha, diversos outros autores apresentam estudos semelhantes (Vukmir, 2006; Blackwell e Kaufman, 2002; Sánchez-Mangas *et al.*, 2010, Osborne *et al.*, 2014).

Assim, fica evidente que qualquer ferramenta capaz de proporcionar um ganho operacional para sistemas de atendimento de emergência com relação ao tempo de resposta é de grande valia para os gestores e a população em geral. É justamente neste contexto em que estão inseridas diversas das técnicas relativas à pesquisa operacional.

O problema de análise de serviços de atendimento móveis de emergência pode ser abordado de diferentes maneiras: modelo hipercubo de filas (Iannoni, Morabito e Saydam, 2011; Takeda *et al.*, 2007; Almehdawe *et al.*, 2013), programação inteira mista (Schmid e Doerner, 2010; Maleki *et al.*, 2014), programação dinâmica (Maxwell *et al.*, 2010), otimização estocástica (Naoum-Sawaya *et al.*, 2013) simulação (Su e Shih, 2003), localização de bases (Toregas *et al.*, 1971), dentre outros.

O modelo hipercubo de filas, desenvolvido por Larson (1974), é uma ferramenta analítica e descritiva que permite calcular uma ampla variedade de medidas de desempenho que auxiliam nas decisões operacionais e de configuração do sistema de emergência. Entretanto, o hipercubo não é um modelo de otimização que determina uma configuração ótima para o sistema, mas fornece uma completa avaliação de desempenho de cada configuração sugerida (TAKEDA *et al.*, 2004).

Modelos de otimização buscam encontrar a solução ótima de um problema matemático formulado a partir de uma função objetivo e restrições. São adequados quando o problema em estudo apresenta funções diferenciáveis e pouca aleatoriedade nos dados de entrada utilizados na modelagem. Uma alternativa para esta limitação é o uso de técnicas de simulação, focando na análise do comportamento do sistema frente alterações nos parâmetros de entrada.

Em diversos trabalhos abordando sistemas de atendimento de emergência são apresentados resultados reais de ganho operacional do sistema com a aplicação das técnicas apresentadas. Su e Shih (2003) desenvolveram um modelo de simulação do sistema de ambulância da cidade de Taipei (Taiwan), e propuseram alterações na quantidade de equipes na central de atendimento e na lógica de transporte dos pacientes para os hospitais que diminuíram em 50% a probabilidade de espera de uma chamada e diminuíram a probabilidade das vítimas terem de esperar pelo despacho de uma unidade em menos de 1%. Repede e Bernardo (1994) aplicaram um modelo de localização

baseado em área de coberturas associado com um modelo de simulação para a cidade de Louisville (Kentucky) e conseguiram, através do redimensionamento do sistema uma redução de 36% no tempo de resposta. Fujiwara *et al.* (1987) utilizaram o modelo de localização de máxima cobertura proposto por Daskin (1983) e um modelo de simulação para análise das melhores soluções para analisar o sistema de atendimento móvel da cidade de Bangkok (Tailândia). Com o modelo os autores propuseram uma nova configuração para o sistema utilizando um menor número de ambulâncias, mas mantendo o mesmo padrão de desempenho.

Após esta breve apresentação da importância das técnicas de pesquisa operacional na análise de sistemas de atendimento móveis de emergência, as próximas seções apresentam os objetivos específicos e os motivos da escolha da simulação de eventos discretos, bem como buscam mostrar a carência de metodologias genéricas na área de localização de ambulâncias e também de métodos heurísticos associados com modelos de simulação para o processo de dimensionamento, fatos estes que justificam a relevância do trabalho.

1.2 OBJETIVOS ESPECÍFICOS

- i. Apresentar uma metodologia para simulação de sistemas de ambulância que possa ser facilmente implementada e aplicada em contextos diferentes.
- ii. Desenvolver métodos heurísticos que possam ser utilizados de maneira eficiente em conjunto com modelos de simulação para resolver o problema de localização de ambulâncias em uma determinada região.
- iii. Observar o comportamento do processo de otimização via simulação com o uso de metamodelos que possam funcionar como um filtro das melhores soluções.
- iv. Aplicar a metodologia desenvolvida em um sistema real.

1.3 MOTIVAÇÃO PARA O USO DA SIMULAÇÃO

A modelagem de sistemas reais de atendimento a emergências possui uma complexidade associada bastante alta, uma vez que tais sistemas são compostos por vários veículos de tipos diferentes, as solicitações por serviço ocorrem distribuídas no tempo e no espaço, existe cooperação entre veículos de áreas distintas, podem ocorrer

múltiplos despachos para atender a um mesmo chamado, o tempo médio de viagem varia de acordo com a região, o dia e o período do dia e ainda existe a possibilidade de formação de filas de espera. A simulação aparece como uma ferramenta adequada para lidar com estas questões, considerando aspectos estocásticos relativos às atividades bem como incertezas associadas ao processo de resgate.

Os serviços de atendimento médico de emergência possuem características que favorecem o uso de técnicas de modelagem, como por exemplo, a simulação. O instante de início e término de cada atividade é bem definido, sendo que as tarefas se sucedem em curtos períodos de tempo. Além disso, o número elevado de variáveis e a natureza aleatória da demanda fazem com que o problema de localização de unidades seja de natureza combinatória, com um elevado número de possíveis alternativas, fazendo com que métodos determinísticos sejam pouco atrativos.

Um sistema médico de emergência pode ser concebido como uma fila única operando sob um sistema de fila com múltiplos servidores subordinados a um controle central. A distribuição dos tempos de serviço é uma função da quantidade de ambulâncias que estão ocupadas em um determinado instante t caracterizado pela chegada de uma chamada. Além disso, a distribuição do tempo de atendimento não é a mesma para cada servidor, exceto para o caso em que todas as ambulâncias estão localizadas na mesma estação. Estas características do processo de atendimento médico de urgência dificultam a modelagem analítica deste sistema de filas, uma vez que a localização dos servidores e a distribuição espacial dos incidentes em cada localidade influenciam o tempo de deslocamento, que é um componente significativo do tempo de serviço (FITZSIMMONS, 1971).

Sorensen e Church (2010) desenvolveram em estudo recente um modelo híbrido para localização de unidades de resgate combinando os modelos de Daskin (1983) e ReVelle e Hogan (1989), sendo que a simulação computacional foi utilizada para estimar níveis de serviço agregados. Na conclusão de seu trabalho, os autores afirmam que modelos de localização de unidades de resgate são difíceis de resolver, uma vez que exigem um conjunto de simplificações e/ou aproximações feitas para facilitar a implementação de modelos de média e larga escala. Afirmam que tais técnicas, apesar de necessárias, introduzem erros nas estimativas de desempenho do sistema. Por fim, sugerem o uso de

rotinas de simulação para aumentar a confiança dos gestores nos resultados gerados por tais modelos. Além disso, comparam modelos probabilísticos de cobertura formulados como modelos de programação inteira mista com abordagens de localização que combinam rotinas de simulação com métodos heurísticos. Concluem afirmando que uma vez que a natureza aleatória das operações relativas a um sistema de emergência deve ser aproximada nos modelos de programação linear inteira, as estimativas dos padrões de desempenho fornecidas estão sujeitas a erros, e isto pode degradar o desempenho geral dos resultados. Ao contrário, rotinas de simulação devem ser capazes de fornecer uma estimativa mais precisa da performance do sistema em estudo. A mesma conclusão também havia sido abordada no trabalho de Lubicz e Mielczarek (1987).

Goldberg (2004), em seu trabalho de revisão de estudos envolvendo estratégias de modelagem para localização de unidades de resgate, sugere o uso de teoria das filas ou simulação associadas a algum método heurístico de busca, uma vez que proporcionam uma maior utilidade aos gestores para aplicações reais de planejamento de sistemas de emergência.

Além dos aspectos apresentados, segue-se alguns motivos pontuais que testemunham a favor do uso das técnicas de simulação no presente contexto: na simulação, entidades possuem características que determinam seu caminho ao longo da rede de operações, exatamente da mesma maneira como pacientes possuem características individuais que determinam seu caminho através das atividades inerentes ao atendimento hospitalar (BRAILSFORD, 2007); modelos de simulação enfatizam a representação direta da estrutura e lógica de um sistema, ao invés de abstrair o sistema na forma de um modelo estritamente matemático (STANDRIDGE, 1999); diferentes alternativas podem ser testadas sem a preocupação que consequências negativas possam atrapalhar ou danificar as operações diárias (STANDRIDGE, 1999); pode-se definir e avaliar qualquer medida de desempenho que seja de interesse, incluindo aquelas que são particulares de um determinado sistema (STANDRIDGE, 1999); pode ser útil no auxílio à organização do processo de tomada de decisão da empresa e identificação de potenciais áreas problemas (LOWERY, 1996); sistemas na área de saúde são definidos por atividades e procedimentos caracterizados muitas vezes por um alto grau de incerteza e variabilidade, sendo adequado o uso de uma abordagem estocástica para tratamento e

análise dos eventos de interesse (BRAILSFORD, 2007); é capaz de lidar de maneira eficaz com a complexidade inerente aos processos que são característicos das organizações de saúde (BRAILSFORD, 2007); permite ao programador associar às entidades todas as características necessárias, fazendo com que as distribuições de tempos de serviço possam ser dependentes das mesmas e qualquer distribuição empírica ou paramétrica pode ser utilizada para modelar a duração dos processos (BRAILSFORD, 2007); fornece uma descrição mais detalhada do sistema (SINGER & DONOSO, 2008); permite uma análise dinâmica dos efeitos causados por mudanças no funcionamento do sistema (BERCHI *et al.*, 2010).

1.4 CONTRIBUIÇÃO DO TRABALHO

A contribuição principal do trabalho diz respeito à utilização de metamodelos para guiar os métodos heurísticos no procedimento de otimização via simulação no contexto de sistemas de atendimento de emergência. Um metamodelo pode ser definido como um modelo de uma função com entradas e saídas (*input/output*) denotado por $Y(x)$, onde x representa um vetor k -dimensional dos k parâmetros de entrada, também denominados de fatores (Kleijnen e Van Beers, 2004). Heurísticas são técnicas utilizadas para resolver mais rapidamente problemas de otimização quando métodos exatos são muito demorados ou impossíveis de serem utilizados.

A premissa a ser verificada é que a utilização dos metamodelos pode gerar resultados melhores do que somente a utilização de métodos heurísticos dentro do contexto de otimização via simulação. O procedimento proposto tem como objetivo gerar soluções de qualidade (com relação ao custo de operação do sistema e tempo de resposta) para as questões de alocação e dimensionamento de ambulâncias nas bases de operação.

Além da metodologia, o trabalho pretende contribuir no desenvolvimento de uma metodologia simplificada de simulação baseada em eventos discretos para análise e estimativa dos principais parâmetros de desempenho de sistemas de emergência e na proposição de métodos heurísticos específicos para geração de configurações candidatas a serem analisadas via simulação.

Os modelos de simulação são utilizados para analisar sistemas de emergência desde a década de 1960 (Bell and Allen, 1969; Savas, 1969). Desde então, diversos estudos foram desenvolvidos abordando alguma cidade ou região específica (Kock e Weigl, 2003; Wu e Hwang, 2009; Henderson e Mason, 1999; Su e Shih, 2003). Gunal e Pidd (2010) afirmam que o número de artigos envolvendo modelos de simulação na área de saúde teve um crescimento acelerado desde 2004.

Dentro deste contexto, Fone *et al.* (2003) apresentam um vasto estudo de revisão bibliográfica de aplicações da pesquisa operacional na área de saúde, concluindo que a qualidade dos modelos aumentou de 1980 até 1999. Entretanto, os autores salientam que a maioria dos modelos estudados analisa aspectos específicos de sistemas maiores, como por exemplo, departamentos dentro de um hospital. Concluem apontando para a falta de modelos mais genéricos que possuam uma maior aplicabilidade. Na mesma linha, Gunal e Pidd (2010) afirmam que apesar do crescimento no número de trabalhos na área desde 2004, existe uma falta de procedimentos genéricos para analisar sistemas no contexto da saúde. Eles argumentam que os modelos são bastante específicos para cada caso, sendo que sua reusabilidade é extremamente difícil. Em consonância com o trabalho de Gunal e Pidd (2010), Young *et al.* (2009) sugerem o uso de metamodelos como uma estratégia para melhorar a adoção de técnicas da pesquisa operacional na saúde, de maneira que os gestores possam adaptar modelos para suas necessidades específicas.

Diversos trabalhos propuseram abordagens de simulação para análise de sistemas de atendimento de emergência (Su e Shih, 2003, Garcia, 2006, Berchi *et al.*, 2010, Rauner *et al.*, 2012, dentre outros). Entretanto, todos se valeram de simplificações que interferem na construção de estimativas do tempo de resposta, como por exemplo, a desconsideração de atividades relevantes do sistema (e.g. manutenção das ambulâncias e reposição de equipamentos). Maiores detalhes dos trabalhos referenciados são apresentados na seção 3 de revisão. Um dos objetivos do presente trabalho é justamente suprir esta lacuna na área de modelagem de sistemas móveis de emergência, que carece de metodologias genéricas que possam favorecer o desenvolvimento de modelos possíveis de serem implementados de maneira rápida e fácil, além de reutilizados independente da localidade. Acredita-se que a metodologia proposta considera de maneira simples (em termos de implementação) e eficiente os processos inerentes ao

atendimento de emergências, gerando estimativas estatisticamente confiáveis para os principais parâmetros de desempenho deste tipo de sistema.

Dos vários trabalhos desenvolvidos na área, alguns procuram integrar modelos de localização com alguma rotina de simulação, utilizando desta última para analisar os melhores cenários gerados pelo modelo matemático (Nogueira Junior, 2011; Aringhieri, 2010). Dos trabalhos analisados, a grande maioria (Swoveland *et al.*, 1973, Fujiwara *et al.*, 1987, Aringhieri *et al.*, 2007, Sorensen e Church, 2010, dentre outros) não abordou de maneira integrada o uso de modelos de simulação com métodos heurísticos para guiar o processo de busca de uma solução de qualidade. Somente Zhen *et al.* (2014) propuseram um método baseado em otimização via simulação de maneira integrada, utilizando da meta-heurística algoritmo genético.

O presente trabalho busca ir além ao incorporar o uso de metamodelos ao processo de otimização via simulação. Em todos os outros trabalhos, primeiro o modelo é rodado com algum parâmetro estocástico sendo substituído por um valor médio. Segundo, o modelo de simulação é utilizado para avaliar as soluções propostas em um contexto mais dinâmico e realista, onde os parâmetros estocásticos são representados por distribuições de probabilidade. Desta maneira, os modelos de otimização são utilizados como uma aproximação inicial, e suas soluções devem ser refinadas pelo uso de rotina de simulação.

Mesmo os trabalhos que propuseram fazer tal integração (Silva e Pinto, 2010), utilizaram de softwares específicos que, por sua vez, fazem uso de algoritmos extremamente genéricos baseados em técnicas da literatura de otimização determinística e meta-heurísticas (Fu, 2002). Como exemplo, podemos citar o pacote de otimização OptQuest da OptTek Systems (integrado com o simulador Arena da Rockwell Automation Technologies) que faz uso das estratégias de busca por dispersão, busca tabu e redes neurais; e o SimRunner da ProModel Corporation, que se utiliza de algoritmos evolucionários e genéticos. O trabalho de Fu (2002) ainda destaca que para sistemas complexos, o fator principal a ser considerado é a modelagem da natureza estocástica das atividades e processos, sendo que os algoritmos atualmente implementados não conseguem lidar de maneira adequada com esta questão.

Diante dessa constatação, propõem-se integrar as etapas de simulação e otimização, utilizando para isso de métodos heurísticos específicos para geração das configurações candidatas que considerem as características inerentes deste tipo de sistema bem como as estimativas dos parâmetros de desempenho fornecidas pelo modelo de simulação (e.g. tempo de resposta, taxa de utilização das unidades). Maiores detalhes dos métodos propostos são apresentados na seção 4.2.

Banks *et al.* (2005) afirmam que para sistemas com incerteza, o uso da otimização pode ser dificultado por diversos motivos como, por exemplo, o número elevado de variáveis para modelagem e a falta de conhecimento sobre a estrutura da função a ser otimizada. O processo de otimização para simulação acrescenta uma dificuldade adicional, uma vez que o desempenho de uma configuração específica não pode ser avaliado de maneira exata, mas necessita ser estimado. Desta maneira, pode não ser possível determinar de maneira conclusiva se uma configuração é melhor que outra, dificultando o uso de algoritmos de otimização que se movem de acordo com direções de melhora. A princípio, a variância correspondente a determinada configuração pode ser reduzida com o aumento no número de replicações ou no tempo de cada corrida. A contrapartida é o fato de cada vez menos alternativas poderem ser analisadas, já que o tempo para análise de cada uma fica cada vez maior. É justamente neste contexto em que está inserido o uso de metamodelos para aumentar a eficiência do processo de otimização para simulação, conforme proposto no presente trabalho.

O uso de metamodelos para reduzir as limitações do consumo elevado de tempo do processo de simulação computacional foi proposto inicialmente por Blanning (1975). Tradicionalmente, as duas abordagens mais utilizadas para construção de metamodelos são os modelos de regressão e métodos baseados em superfície de resposta (Fishwick, 1989). Desde então, estes têm sido largamente utilizados como auxílio ao processo de otimização para simulação (Zaytoon e Villerman-Lecolier, 1994; Lei *et al.*, 2007; Barton, 2009; Kleijnen, 2009a).

No presente trabalho, a ideia básica está em utilizar os metamodelos para filtragem das configurações candidatas geradas pelos métodos heurísticos. Os metamodelos produzem uma estimativa do tempo de resposta do sistema dada uma configuração qualquer em tempo computacional extremamente reduzido. Somente as soluções que tiverem um

tempo de resposta próximo do parâmetro utilizado como restrição é que serão analisadas via simulação, ou seja, as configurações com maior potencial de gerarem boas soluções serão refinadas via simulação. Desta maneira, reduz-se o número de configurações analisadas pelo modelo de simulação, que é o processo mais demorado dentro da metodologia proposta, e torna-se possível analisar um maior número de soluções candidatas, aumentando consequentemente a probabilidade de encontrar uma configuração de boa qualidade. Cabe ressaltar que a eficiência do método está diretamente relacionada com a qualidade das estimativas geradas pelo metamodelo, conforme discutido mais adiante em seção específica.

Uma busca por trabalhos semelhantes nos principais periódicos da área de Pesquisa Operacional, inclusive específicos da área de saúde, até o começo do ano de 2015, não retornou nenhuma publicação igual ao proposto no presente trabalho. Foram analisados, dentre outros, os periódicos *European Journal of Operational Research*, *Management Science*, *Health Care Management Science*, *Journal of Heuristics*, *Journal of Simulation*, *Simulation Modelling Practice and Theory*, *Location Science*, *Computers & Operations Research*, *Operations Research Letters*, *Annals of Operations Research*, *Operations Research*, *Transportation Research* e *INFORMS Journal on Computing*.

1.5 ESTRUTURA DO DOCUMENTO

A seção 2 do trabalho apresenta uma descrição geral do funcionamento de sistemas móveis de emergência no contexto nacional, a seção 3 uma revisão de literatura dos principais trabalhos envolvendo simulação de sistemas de ambulância, bem como o uso de heurísticas e metamodelos neste contexto, a seção 4 descreve todos os aspectos da metodologia proposta, a seção 5 os resultados computacionais considerando a aplicação da metodologia em um contexto real e a seção 6 as conclusões do trabalho bem como sugestões para trabalhos futuros.

2. CARACTERIZAÇÃO DE SERVIÇOS MÓVEIS DE EMERGÊNCIA

O serviço de atendimento móvel de urgência no Brasil é inspirado no 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).

A grande maioria dos serviços hoje em operação no território nacional atua com dois tipos principais de unidades: básica e avançada. As unidades de suporte básico (USB) e as unidades de suporte avançado (USA) são constituídas por uma tripulação exclusiva, formada por equipes compostas por auxiliar ou técnico de enfermagem, motorista-socorrista, médico e enfermeiro que possuem experiência profissional no atendimento de urgências e emergências, que se diferenciam pelo atendimento específico e competência técnica dos profissionais lotados em ambas as unidades, conforme avaliação do grau de gravidade do caso (LOPES; FERNANDES, 1999).

A primeira é acionada para atender casos de gravidade leve e média ou então em auxílio a uma unidade avançada, por estar mais próxima do local do incidente, sendo composta por motorista socorrista e auxiliar de enfermagem. A segunda só é empenhada em casos graves, por serem em menor quantidade, compostas por motorista, enfermeiro e médico. Alguns sistemas ainda atuam com uma unidade de saúde mental (USM), exclusiva para atendimento de casos psiquiátricos mais graves.

As unidades pré-hospitalares (básicas e avançadas) instalam-se em Unidades Básicas Distritais de Saúde, que se deslocam a partir de determinação do médico regulador, lotado em uma central de regulação, para prestar o atendimento determinado. Estas têm competência para atender pacientes traumatizados ou com patologias clínicas (LOPES; FERNANDES, 1999).

A Central de Regulação Médica no Brasil assume uma função coordenadora e disponibilizadora de atendimento pré-hospitalar à população local, sendo de

competência deste serviço, assumida pelo médico regulador e/ou coordenador, podendo ser de abrangência municipal ou estadual, micro ou macrorregional. Dispõe de linhas privadas pertencentes ao sistema 192 para os usuários, além de linhas externas de apoio e de linhas internas disponibilizadas nos hospitais e centros de referência (LOPES; FERNANDES, 1999).

Quanto ao médico regulador, este é uma autoridade sanitária pública, que por delegação do Gestor do SUS (Secretário Municipal ou Estadual de Saúde), ordena e coordena o uso de todos os recursos envolvidos no atendimento de saúde às urgências, sejam elas civis, militares, públicas ou privadas, com perfil voltado às experiências na área hospitalar em nível de urgência, além da formação própria para o serviço pré-hospitalar, de modo a detectar as necessidades de atendimento de uma maneira rápida, estabelecendo prioridades, baseadas no grau de comprometimento dos sinais vitais, capazes de comprometer a qualidade ou expectativa de vida (ALMOYNA; NITSCHKE, 1999).

Toda a solicitação é atendida por técnicos auxiliares (ou teledigifonistas) e, uma vez anotada e triada, é repassada ao médico regulador para decisão de conduta. Caso a solicitação implique em despacho de viaturas, o regulador repassa a ficha ao radioperador (ou despachante) que, em seguida, procede ao despacho da mesma. Solicitações internas, provenientes das próprias unidades de saúde, são atendidas pelos técnicos auxiliares que, nestes casos, solicitam a presença do médico que atendeu o paciente, repassando a ligação ao médico regulador, que regula o fluxo conforme seu julgamento. Sempre que há necessidade de transporte do paciente/vítima até o hospital, o médico regulador entra em contato com a equipe médica do hospital que, naquele momento, tem condições de atendimento e garante o acesso do paciente àquele centro hospitalar. Desta forma, tanto para as USA quanto para as USB, o médico regulador gerencia o fluxo de circulação entre o campo pré-hospitalar e o hospitalar (LOPES; FERNANDES, 1999).

Competirá, portanto, ao médico regulador a determinação da inexistência de risco imediato à vida, pois, uma vez determinado este risco, o atendimento competirá às USA. Na ausência de risco imediato à vida, o atendimento competirá às USB. A Figura 1 apresenta um resumo das principais etapas que fazem parte do processo de resgate

conduzido por unidades móveis, desde a instalação do problema à saúde e sua identificação até o retorno da ambulância à sua base de operação.

Ao final do processo de entrega do paciente no hospital, pode haver a necessidade de repor algum equipamento na unidade móvel que fez o atendimento. Dependendo do sistema, esta atividade acontece no próprio hospital ou então na central de regulação. Durante o trajeto até a central, a ambulância fica indisponível para fazer um novo atendimento. Após a reposição, se houver chamada em espera, a ambulância se dirige até o local do incidente, caso contrário retorna para sua base de origem. Além da reposição de equipamentos, pode acontecer da unidade móvel ter sofrido alguma avaria ou necessitar de manutenção, tendo de se deslocar até a oficina conveniada para realização dos serviços de reparo. Durante este tempo, a unidade fica novamente indisponível. Todas estas atividades foram consideradas na simulação.

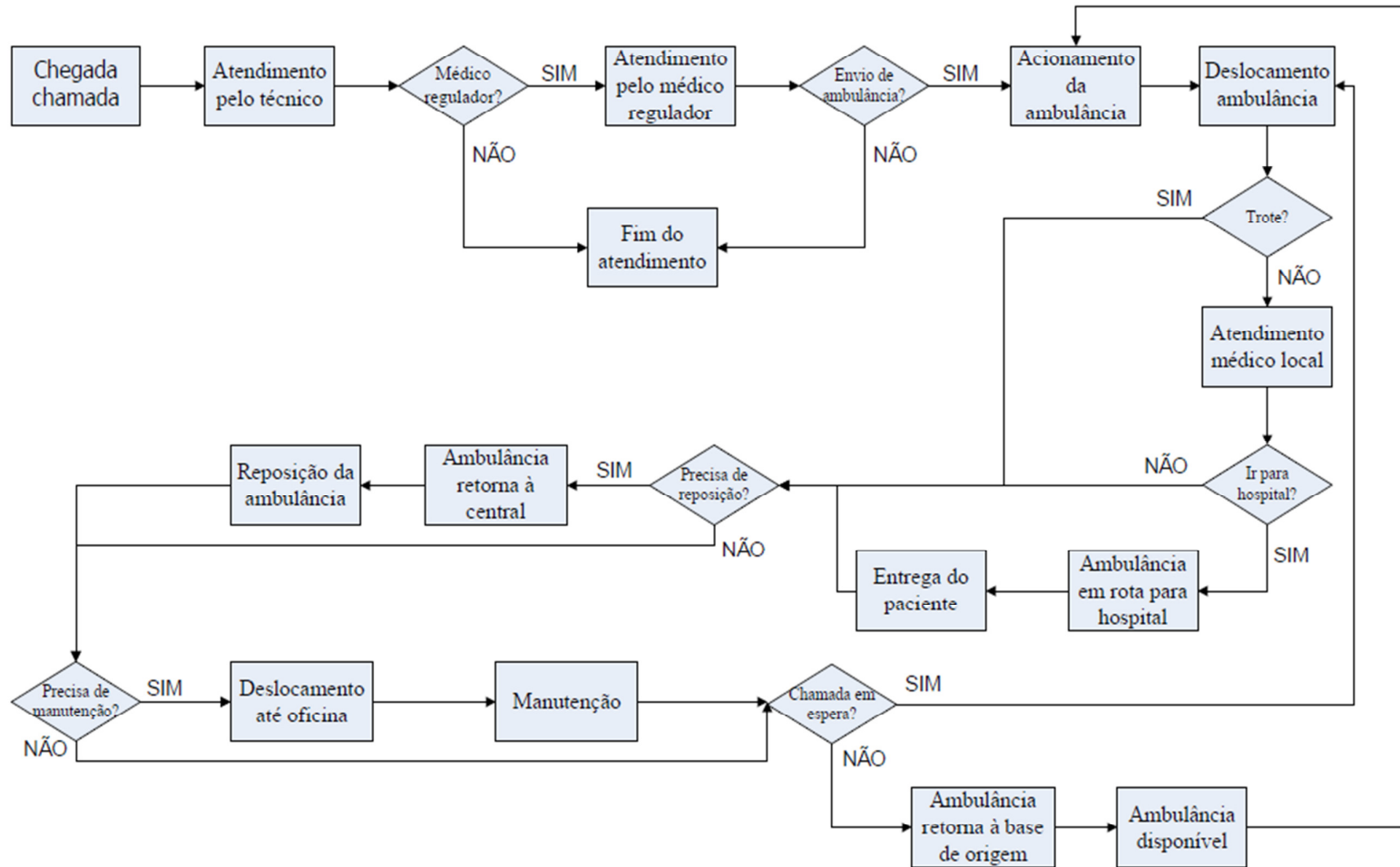


Figura 1: Fluxograma do processo de atendimento a uma chamada
 Fonte: Silva (2010)

3. REVISÃO DE LITERATURA

O capítulo de revisão da literatura está dividido em três seções. A primeira apresenta uma revisão sobre modelos de simulação voltados para análise de sistemas de emergência, buscando-se evidenciar a falta de modelos mais abrangentes que possam ser facilmente adaptados para novos contextos. Nesta seção mostra-se também que a maioria dos trabalhos que utilizaram métodos de otimização e simulação em conjunto, não fizeram de maneira integrada. A seção 3.2 apresenta uma revisão sobre métodos heurísticos aplicados no contexto de localização de unidades de resgate, focando naqueles mais utilizados. A seção 3.3 apresenta os conceitos básicos relacionados com a construção de metamodelos e sua aplicação no processo de otimização via simulação.

3.1 METODOLOGIAS DE SIMULAÇÃO

Modelos de simulação são utilizados para análise de sistemas de emergência desde a década de 1960 (Bell e Allen, 1969; Savas, 1969; Fitzsimmons, 1971). Desde então, diversos estudos foram conduzidos procurando focar em uma região ou localidade específica. Henderson e Mason (1999) desenvolveram um modelo de simulação para refinar as soluções propostas por um modelo de filas, no serviço de ambulância da cidade de Auckland (Nova Zelândia). Kock e Weigl (2003) utilizaram da simulação para analisar a logística do sistema de emergência na Áustria, comparando políticas de centralização e descentralização. Su e Shih (2003) desenvolveram estudo semelhante na cidade de Taipei (Taiwan).

Mais recentemente, Wu e Hwang (2009) implementaram um modelo de simulação para estimar a necessidade de investimento na expansão da frota de ambulâncias e analisar diferentes estratégias de despacho durante grandes eventos, como concertos ou maratonas. Nesta mesma linha, Rauner *et al.* (2012) sugerem a adoção de técnicas de simulação em um estudo para melhorar o nível do atendimento de emergência em situações com número elevado de vítimas. Aboueljinane *et al.* (2014) utilizaram técnicas de simulação discreta para analisar o SAMU da cidade de Val-Der-Marne (França), em que testes foram conduzidos variando a quantidade de recursos, realocação das unidades existentes e redução dos tempos de operação do sistema. Os autores concluem que o reposicionamento de parte das unidades em locais potenciais de

operação aumentou o percentual de chamadas atendidas dentro de 20 minutos em 4,8%. Quando reduzido o tempo de regulação em 20%, o percentual das chamadas atendidas no mesmo padrão aumentou 5,2%.

Outros trabalhos focam na análise de diferentes políticas de operação dos sistemas, como por exemplo, centralização ou descentralização (Lee, 2012), políticas de envio dos pacientes aos hospitais (Nafarrate *et al.*, 2011), posicionamento dinâmico das ambulâncias nas bases ao longo dos dias (Su e Liang, 2002). No contexto nacional, trabalhos semelhantes também foram desenvolvidos nos últimos anos: Garcia (2006) desenvolveu um modelo de simulação integrando as regiões de Niterói e São Gonçalo, buscando identificar a configuração mais adequada com relação à quantidade de ambulâncias em operação; Silva e Pinto (2010) trabalharam na mesma linha com a análise do SAMU de Belo Horizonte, analisando os impactos no tempo de resposta do sistema frente ao aumento na demanda pelo serviço e mudanças na lógica de operação.

De maneira geral, a aplicação da modelagem via simulação ou utilizando de outras técnicas da pesquisa operacional na área de saúde (Morohosi e Furuta, 2013) têm sido extensamente discutidas na literatura. Este fato é motivado principalmente pelo envelhecimento observado na população mundial nos últimos anos, principalmente nos centros urbanos (OPAS, 2009), com conseqüente aumento na necessidade por serviços de saúde, como por exemplo, os sistemas de ambulância.

Muitos autores discutem ainda as principais barreiras que dificultam o uso destas técnicas no contexto da saúde (Kuljis *et al.* 2007; Young *et al.* 2009). Entretanto, pode-se destacar ainda alguns fatores que podem ser vistos como oportunidades de desenvolvimento na área de modelagem e simulação, como por exemplo, a pressão constante por melhores serviços, a disponibilidade reduzida de recursos e a necessidade de avaliar o impacto de possíveis mudanças antes de sua real implementação.

Dentro desta linha, poucos trabalhos se propuseram a desenvolver metodologias que pudessem ser reutilizadas para condução de estudos e implementação de modelos na área de simulação de serviços de atendimento móveis de emergência. Na maior parte das vezes o foco é no desenvolvimento de modelos específicos, considerando

particularidades da região em estudo ou na reutilização de modelos já desenvolvidos (principalmente no caso de otimização).

A partir da década de 1980, é possível identificar trabalhos que se propuseram ao desenvolvimento de modelos e abordagens mais genéricas. Hill *et al.* (1984) tentaram tal abordagem para análise do sistema de emergência da cidade de Boston. Apesar de chegarem a resultados que conseguiram reduzir o tempo de resposta com uma melhor distribuição das unidades, o modelo proposto apresenta uma série de simplificações que dificultam sua adoção como metodologia. Os tempos de preparação da equipe, tempo em cena e viagem de volta foram estimados como valores médios de todos os dados coletados em campo, sem diferenciação da região de origem do incidente, dia da semana e horário. Para o tempo de deslocamento das unidades foi desenvolvido um modelo específico para a cidade de Boston. Com relação ao processo de geração das chamadas, a cidade foi dividida em regiões menores e o incidente gerado sempre no ponto central, desconsiderando outras possibilidades. Foram desconsideradas também características específicas de cada chamada, como por exemplo, o tipo da ocorrência. Atividades importantes como a reposição de equipamentos e manutenção foram desconsideradas.

Nesta mesma linha, Lubicz e Mielczarek (1987) apresentaram um modelo que pudesse ser aplicado no sistema polonês de atendimento a emergências. Os autores conseguiram mostrar que o simples aumento no número de unidades não possui grande impacto no tempo de resposta, mas sim na diminuição da taxa de utilização das unidades. Os autores incorporaram no modelo, diferentes padrões de chegada das chamadas nos horários de pico, além do uso de distribuições de probabilidade para o tempo em cena considerando o tipo de ocorrência e dia da semana. Entretanto, a abordagem utilizada possui algumas limitações: consideração de diferentes tipos de ambulância, consideração das atividades de reposição e manutenção e de todas as atividades concernentes à central de regulação.

Goldberg *et al.* (1990) apresentaram uma metodologia aplicada na cidade de Tucson (Arizona), com foco nas etapas de validação do modelo de simulação. Apesar de concluírem pela validação do estudo, pode-se destacar algumas de suas limitações: as ambulâncias só saem da própria base para atendimento a um chamado, e somente

quando estão ociosas; os autores optaram pela divisão do território em zonas, sendo que o tempo de deslocamento foi aproximado pelo ponto central, e não até o local exato do incidente; este mesmo tempo foi calculado com base em um modelo de regressão linear múltiplo com diversos parâmetros estimados, agregando maior incerteza aos resultados e não foram consideradas as atividades de manutenção, reposição e as relativas à central de regulação. Em anos mais recentes, Su e Shih (2003), Garcia (2006), Singer e Donoso (2007), Berchi *et al.* (2010), Rauner *et al.* (2012), dentre outros, também apresentaram contribuições para a modelagem de sistemas de atendimento de emergência via simulação.

Ainda neste contexto, alguns trabalhos procuram integrar o processo de simulação com alguma técnica ou modelo de otimização, buscando identificar a configuração ideal do sistema de acordo com parâmetros relacionados com o tempo de resposta ou taxa de utilização das unidades, por exemplo. Tal procedimento foi aplicado com sucesso em diferentes contextos, como sequenciamento do tipo *flow shop* (Allaoui e Artiba, 2004), gerenciamento de unidades médicas (Angelis *et al.*, 2003), minimização da duração de projetos (Zhang e Li, 2004), otimização de estoques com demanda estocástica (Arreola-Risa *et al.*, 2011), dentre outros.

No contexto específico de sistemas de atendimento de emergência, dos trabalhos analisados, somente Zhen *et al.* (2014) apresentaram um método integrado de simulação e otimização para dimensionamento e localização das unidades de resgate. No estudo, um simulador baseado em eventos discretos foi desenvolvido para avaliar soluções candidatas em cenário aleatório, com incertezas associadas aos tempos de deslocamento, chegadas das chamadas na central e tempos de serviço. Com base nas respostas geradas, o componente de otimização da estrutura proposta baseado na meta-heurística algoritmo genético procura por soluções alternativas que são também analisadas pelo simulador. A metodologia proposta foi aplicada na cidade de Xangai, analisando o impacto nas políticas de alocação de unidades frente a alterações nos parâmetros do modelo, como quantidade de hospitais, de ambulâncias, de bases e nas taxas de chegada das chamadas na central.

Outro trabalhos utilizaram modelos de otimização para uma aproximação inicial, sendo as soluções refinadas pelo uso de alguma rotina de simulação. Swoveland *et al.* (1973)

utilizaram uma metodologia de otimização via simulação com o objetivo de avaliar o sistema de atendimento de emergência da região metropolitana de Vancouver (Canadá). As saídas da simulação foram utilizadas para estimar a função objetivo de um procedimento de otimização baseado em *branch-and-bound*. As soluções finais foram verificadas via simulação, tendo sido possível diminuir o tempo de resposta do sistema em 0,3 minutos. Fujiwara *et al.* (1987), na mesma linha, utilizaram da abordagem de otimização via simulação na cidade de Bangkok (Tailândia). Os autores utilizaram o modelo de máxima cobertura esperada proposto por Daskin (1983), analisando via simulação as melhores soluções propostas. Conseguiram melhorar o nível de serviço do sistema utilizando um número menor de ambulâncias.

O trabalho de Aringhieri *et al.* (2007) utilizou de modelos de programação linear inteira para analisar o serviço de emergência da cidade de Milano (Itália). Novamente, as melhores soluções foram analisadas via modelo de simulação baseada em agente, apresentando resultados que indicaram possíveis melhorias no sistema. Sorensen e Church (2010) desenvolveram em estudo recente um modelo híbrido para localização de unidades de resgate combinando os modelos de Daskin (1983) e ReVelle e Hogan (1989), sendo que a simulação computacional foi utilizada para estimar níveis de serviço agregados. Nessa perspectiva, percebe-se que os trabalhos analisados que propuseram aplicar a metodologia de otimização via simulação em serviços de atendimento de emergência não o fizeram de maneira integrada. A simulação foi utilizada para estimar algum parâmetro do modelo de otimização ou para analisar as melhores soluções propostas por este.

3.2 MÉTODOS HEURÍSTICOS

Segundo Brotcorne *et al.* (2003), os primeiros modelos desenvolvidos para localização de unidades de resgate abordavam o problema estático e determinístico, ou seja, desconsideravam considerações estocásticas relativas aos processos e atividades características deste tipo de modelagem. Tais modelos foram concebidos para utilização no estágio de planejamento. Entretanto, o congestionamento em serviços de atendimento de emergência pode causar a indisponibilidade de um servidor, mesmo que esteja localizado a menos de uma distância padrão, motivando o desenvolvimento de modelos de localização probabilísticos.

Em situações do mundo real a facilidade que atende a demanda de uma localidade pode não estar disponível no momento da chegada de uma nova chamada. Uma solução para este problema é estimar o tempo em que a facilidade não está disponível. Modelos probabilísticos maximizam a cobertura das diversas regiões, dispondo de um número finito (e muitas vezes incapaz de cobrir toda a demanda) de facilidades com uma confiança α estabelecida (FIGUEIREDO *et al.*, 2004).

Em contrapartida ao modelo estático, na localização de veículos de emergência, decisões de realocação devem ser feitas periodicamente de maneira a não permitir que certas áreas de atendimento fiquem descobertas. Este fato foi considerado inicialmente no trabalho de Kolesar e Walker (1974), na modelagem de um sistema de realocação de unidades do corpo de bombeiros.

O problema de realocação de ambulância é mais difícil de ser tratado, uma vez que necessita ser resolvido mais frequentemente, em pequenos intervalos de tempo. Com o desenvolvimento de métodos heurísticos eficientes e sistemas computacionais mais eficazes, torna-se possível a resolução rápida em tempo real deste tipo de problema (BROTCORNE *et al.*, 2003). Para uma revisão dos principais modelos de localização, sugere-se os trabalhos de Brotcorne *et al.* (2003), ReVelle e Eiselt (2005) e Arabani e Farahani (2012).

Ao longo dos anos, diversas heurísticas foram propostas para resolução dos problemas estático, determinístico, probabilístico e dinâmico de localização. Adenso-Díaz e Rodríguez (1997) utilizaram da busca tabu para resolução do modelo de Church e ReVelle (1974) aplicado na província de León (Espanha). A heurística proposta começa a partir de uma solução inicial S_0 (um vetor binário x_i^0 com exatamente p 1's indicando em quais localidades existe base instalada) para a qual é avaliada a população atendida por um padrão de tempo T_i . Para qualquer solução $S_i = \langle x_1^i, \dots, x_n^i \rangle$, a vizinhança ∂_i é definida como: $\partial_i = \{ \langle x_1, \dots, x_n \rangle \subseteq \{0,1\}^n : \exists j, k | x_j^i + x_j^k = 1, x_j^i + x_j = 1, x_k^i + x_k = 1, x_l^i = x_l \forall l \neq j, k \}$ ou seja, duas regiões com diferentes alocações são escolhidas e seus valores trocados. O elemento da vizinhança que fornece maior cobertura é selecionado como S_{i+1} , sendo o processo repetido até atingir um determinado critério de parada. O algoritmo ainda mantém uma lista dos movimentos de troca mais recentes de maneira a evitar o processo de ciclagem. Gendreau *et al.* (1997) também utilizaram dos

princípios da busca tabu para resolução de um problema de localização de dupla cobertura, com dados simulados para validação. A solução inicial é obtida pela relaxação linear do modelo. Dada uma solução, uma nova é gerada transferindo um conjunto de veículos entre pontos potenciais de localização de bases, movendo para um vértice onde o número máximo de veículos permitido ainda não foi atingido. Os movimentos são feitos de maneira a recuperar a viabilidade das restrições primária e secundária de cobertura, nesta ordem. Uma tentativa é feita para melhorar a função objetivo realizando movimentos não-tabu. Quando a melhor solução encontrada não é melhorada depois de n iterações, a heurística passa para uma etapa de diversificação. Os mesmos autores expandiram a metodologia para resolução de um problema de alocação dinâmica de ambulâncias na cidade de Montreal. A nova abordagem considera o cálculo prévio das decisões de realocação caso a ambulância i seja alocada à próxima chamada. Desta maneira, quando a chamada acontece, uma unidade é alocada de acordo com as regras já estabelecidas, e o cenário de realocação já computado é aplicado. Para maiores detalhes, sugere-se o trabalho de Gendreau *et al.* (2001).

O procedimento de busca tabu parece ser o mais utilizado para localização de unidades de resgate, como pode ser verificado também nos trabalhos de Doerner *et al.* (2005), Rajagopalan *et al.* (2007), Rajagopalan *et al.* (2008) e Rajagopalan e Saydam (2009). Sajjadian (2009) propôs um novo método para realocação de ambulâncias ao longo do dia utilizando da meta-heurística *simulated annealing*. O método se baseia em duas etapas: a primeira define as regiões mais prováveis da cidade para um acidente acontecer e a segunda direciona as ambulâncias disponíveis para tais regiões. Na literatura existem ainda trabalhos envolvendo as meta-heurísticas algoritmo genético (Iannoni *et al.*, 2009; Sasaki *et al.*, 2010; Amponsah *et al.*, 2011), *variable neighborhood search* (Schmid e Doerner, 2010; Schmid, 2011), heurísticas gulosas (Iannoni *et al.*, 2011; Karaman, 2008) e algumas heurísticas específicas (McAleer e Naqvi, 1994; Brotcorne *et al.*, 2002).

3.3 METAMODELOS

As técnicas de simulação são largamente utilizadas para o estudo e avaliação de sistemas complexos. Devido à natureza estocástica e complexa da maioria dos sistemas reais, os modelos desenvolvidos são geralmente difíceis de implementar e demandam

uma grande quantidade de tempo para sua execução. Metamodelos são aproximações matemáticas de modelos de simulação utilizados principalmente para diminuir a carga computacional (com relação ao tempo de execução) de tais modelos. No processo de otimização via simulação geralmente a forma explícita da função a ser otimizada é desconhecida. Desta maneira, metamodelos são utilizados para gerar aproximações rápidas em substituição ao processo normalmente demorado de simulação.

Um metamodelo pode ser definido como um modelo de uma função com entradas e saídas (input/output) denotado por $Y(x)$, onde x representa um vetor k -dimensional dos k parâmetros de entrada, também denominados de fatores (Kleijnen e Van Beers, 2004). O uso de metamodelos para reduzir as limitações do consumo elevado de tempo do processo de simulação computacional foi proposto inicialmente por Blanning (1975). Seu uso envolve basicamente três passos: (i) seleção do planejamento de experimentos (*experimental design*) para geração dos dados, (ii) seleção do modelo para representar os dados e (iii) ajuste dos dados observados utilizando do modelo escolhido. Existem diversas alternativas para abordar cada passo.

O planejamento pode ser feito através de experimentos fatoriais ou amostragem utilizando do modelo hipercubo latino. A construção dos metamodelos pode ser feita através do modelo de Kriging que gera previsões a partir de uma função linear ponderada (Kleijnen e Van Beers, 2005; Kleijnen, 2009a), redes neurais artificiais (Fonseca e Navarrese, 2002; Fonseca e Navarrese, 2003), modelos de regressão (Dos Santos e Dos Santos, 2009; Jalal *et al.* 2013), superfícies de resposta (Merkuryeva, 2005) dentre outros. A etapa de ajuste pode fazer uso, por exemplo, do método dos mínimos quadrados para modelos de regressão. Trabalhos de revisão de metamodelos podem ser encontrados em Barton (2009) e Kleijnen (2009a). Li *et al.* (2010) propuseram a incorporação de metamodelos em Sistemas de Suporte à Decisão para problemas analisados via simulação estocástica, e conduziram uma comparação sistemática para analisar as vantagens e desvantagens de várias técnicas para construção de metamodelos, como redes neurais artificiais e modelos de Kriging.

Ao longo dos anos, alguns autores procuraram apresentar diferentes estratégias para construção dos metamodelos. Lei *et al.* (2007) apresentaram um método baseado em metamodelos para representar modelos de simulação reutilizáveis, focando na discussão

dos aspectos relacionados à sua implementação e especificação; Kleijnen e Beers (2004) propuseram um novo método para geração de estimativas para construção de metamodelos baseados no modelo determinístico de Kriging; Neddermeijer e Oortmarsen (2000) desenvolveram uma estrutura para modelos de simulação e otimização estocástica utilizando a metodologia de superfície de resposta.

Com relação às aplicações, o uso de metamodelos já foi testado em diferentes contextos. Sartori e Smith (1997) apresentaram um método baseado em regressão múltipla para estudos de investimento de capital em que múltiplos fatores estão sujeitos a fontes de incerteza. O método é ilustrado e validado em dois casos reais. No mesmo contexto, Chaveesuk e Smith (2005) investigaram a aplicação do modelo de Kriging como uma alternativa à regressão polinomial e superfícies de resposta para análise de investimentos de capital.

No ambiente industrial, Durieux e Pierreval (2004) analisaram o impacto na utilização média de um sistema de manufatura com máquinas paralelas a partir de alterações nos parâmetros de operação. Para isso, desenvolveram um modelo de simulação estocástico e construíram um metamodelo de regressão para as análises. Como conclusão, afirmaram que fatores como a velocidade de movimentação de materiais não possuíam grande impacto nas medidas de desempenho, ao contrário do esperado pelos gestores.

Na área logística, Winch *et al.* (2012) propuseram uma abordagem híbrida combinando metamodelos baseados em simulação e métodos de otimização para análise de problemas de logística reversa. A abordagem foi aplicada em sistemas reais visando minimização de custos e desperdícios na rede de distribuição. No contexto da saúde, Kilmer *et al.* (1997) descrevem um modelo de simulação discreta aplicado a um departamento de emergência e o desenvolvimento de um metamodelo da simulação proposta, através do uso de redes neurais artificiais. Ambos são comparados com relação à média e variância do tempo de espera do paciente no setor. Pode-se citar ainda o trabalho de Testi *et al.* (2010) que apresentaram uma estrutura integrada das diferentes partes que compõem um sistema de saúde para decisões táticas e operacionais através do uso de metamodelos, Yi *et al.* (2010) desenvolveram um metamodelo paramétrico para predizer de maneira dinâmica a capacidade de vários hospitais frente a um desastre natural; Cimellaro *et al.* (2011) que utilizaram a abordagem via metamodelo para

estimar a capacidade de atendimento de um hospital, dentre outros. No contexto de sistemas de atendimento móveis de emergência, Swoveland *et al.* (1973) utilizaram da simulação para ajustar os parâmetros de um metamodelo que estima o tempo de resposta das ambulâncias. Este tempo foi otimizado através de um algoritmo *branch-and-bound*.

4. METODOLOGIA PROPOSTA

As próximas seções apresentam as metodologias que foram utilizadas, respectivamente, para construção do componente de simulação, do metamodelo e do módulo de otimização via simulação.

4.1 METODOLOGIA DE SIMULAÇÃO

A metodologia utilizada para construção do módulo de simulação baseou-se naquela apresentada no livro de Banks *et al.* (2005), conforme Figura 2. A primeira etapa da metodologia diz respeito à definição clara de qual problema será analisado. Nesta etapa, é de extrema importância a participação do responsável pela simulação bem como dos gestores do sistema em estudo.

A segunda etapa consiste na definição de objetivos e no planejamento geral do plano de ação. Os propósitos e objetivos do estudo devem ser claramente definidos. Nesta etapa, avalia-se se a simulação é realmente a ferramenta indicada para responder os objetivos propostos. Em caso positivo, o planejamento geral deve conter os cenários que serão analisados bem como qual método será utilizado para avaliar a efetividade das alternativas testadas. Deve incluir também a quantidade de pessoas envolvidas, o custo estimado do estudo, o número de dias necessários para atingir cada fase do plano e os resultados esperados.

A terceira fase envolve a modelagem conceitual do sistema em estudo. De acordo com Freitas Filho (2008), esta etapa consiste em traçar um esboço do sistema, de forma gráfica (fluxograma, por exemplo) ou algorítmica (pseudocódigo), definindo componentes, descrevendo as variáveis e interações lógicas que constituem o sistema. Recomenda-se começar por um modelo mais simples e crescendo aos poucos até atingir uma complexidade suficiente para atender os objetivos propostos anteriormente. Law (2007) afirma que a etapa de criação do modelo conceitual é o aspecto mais importante de um estudo de simulação. Existem diversas ferramentas e técnicas que auxiliam nesta etapa, como o Diagrama de Ciclos de Atividades (*Activity Cycle Diagram*) e a ferramenta integrada para modelagem de funções IDEF (*Integration DEFinition*).

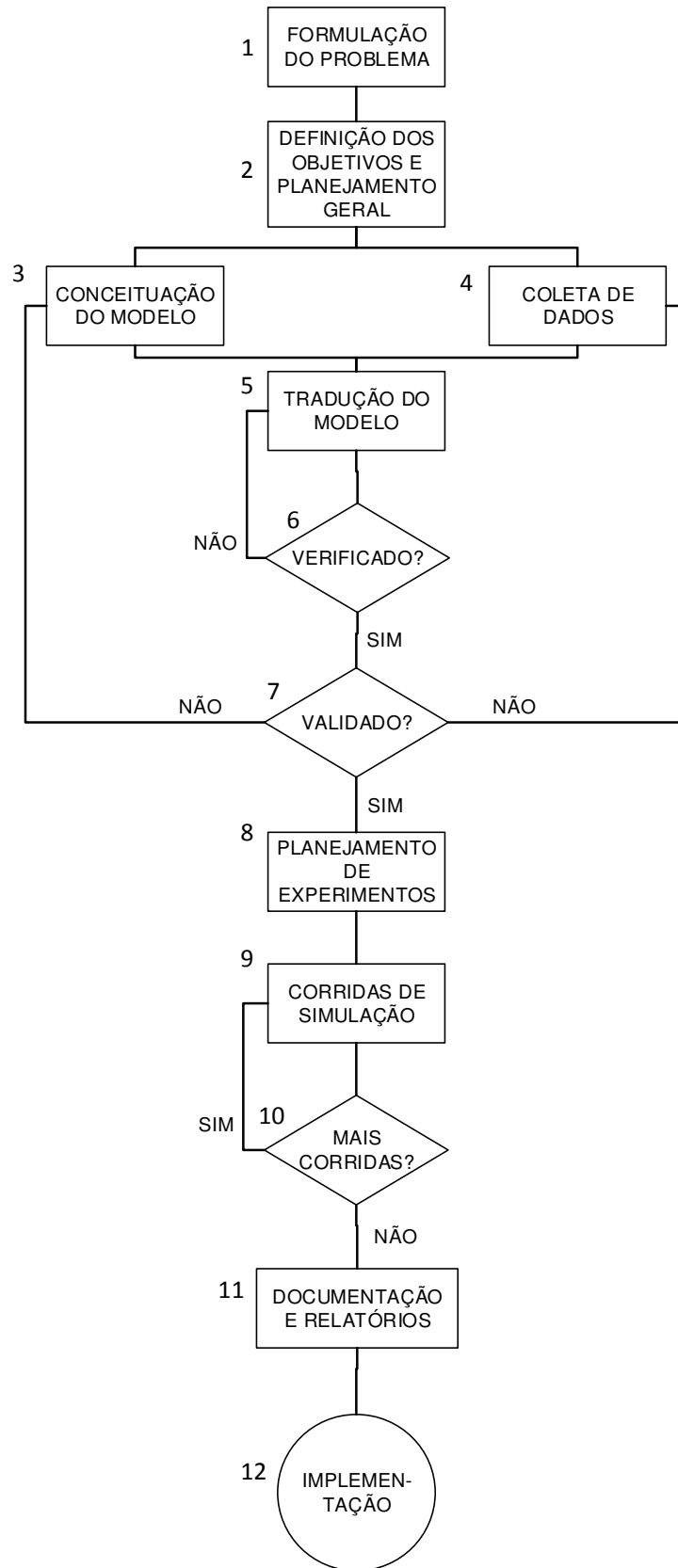


Figura 2: Metodologia de simulação
 Fonte: Banks *et al.* (2005)

De maneira integrada com a construção do modelo conceitual, acontece a atividade de coleta de dados. Se a complexidade do modelo for alterada (fase de modelagem conceitual), os dados necessários também podem sofrer alterações. As etapas desta análise de dados dizem respeito à amostragem, tratamento e inferência. Com relação à coleta de dados, Vincent (1998) sugere que o tamanho da amostra deve estar entre 100 e 200 observações. Amostras com menos de 100 observações podem comprometer a identificação do melhor modelo probabilístico, e amostras com mais de 200 observações não trazem ganhos significativos ao estudo. Além disso, é importante observar se os dados sofrem alguma influência dependendo das condições de operação, como por exemplo dia da semana ou hora do dia.

Feita a amostragem, os dados coletados devem ser analisados estatisticamente para identificar possíveis falhas ou valores discrepantes. Nesta etapa, sugere-se o uso dos quartis da amostra e construção de gráficos *boxplot* para identificação de possíveis valores atípicos. Seja Q_1 e Q_3 o primeiro e terceiro quartis da amostra, respectivamente, e α um parâmetro utilizado para determinar a rigidez na identificação dos valores discrepantes. Para um dado amostrado x_i qualquer, temos que este será considerado um valor discrepante se uma das duas inequações forem satisfeitas.

$$x_i \leq Q_1 - \alpha(Q_3 - Q_1) \quad (1)$$

$$x_i \geq Q_3 + \alpha(Q_3 - Q_1) \quad (2)$$

Os valores mais usuais para o parâmetro α é de 1,5 (valor discrepante moderado) ou 3 (valor discrepante extremo) (Chwif e Medina, 2010). Cabe ressaltar que nem sempre deve-se retirar todos os valores discrepantes identificados, pois às vezes trata-se de um valor que não pode ser desprezado ou de uma característica do próprio processo em estudo. Nestes casos, sugere-se a opinião dos gestores e/ou operadores do sistema ou o aumento no número n de dados coletados.

Uma vez amostrados e tratados, os dados coletados devem ser analisados buscando identificar se existe alguma distribuição de probabilidade que os representem a um nível de significância α , ou seja, realizar um teste de aderência do tipo: $H_0 \rightarrow$ a variável aleatória X segue a distribuição sob hipótese e $H_1 \rightarrow$ a variável aleatória X não segue a distribuição sob hipótese.

Para esta análise, existem testes baseados na comparação das frequências observada e esperada da variável de interesse, conforme equação 3.

$$\sum_{i=1}^k \frac{(fo_i - fe_i)^2}{fe_i} \quad (3)$$

Como exemplo, pode-se citar o Teste Qui-Quadrado ou o Teste Kolmogorov-Smirnov (K-S). Neste tipo de teste, deve-se analisar o *square error* (erro quadrático), que segundo Freitas Filho (2008), é obtido a partir do somatório das diferenças entre as frequências relativas observadas (fo) e as frequências relativas esperadas (fe) tomadas ao quadrado, para todos os intervalos do histograma representativo da amostra, sendo o primeiro indicador da qualidade do ajuste realizado. Além deste parâmetro, alguns *softwares* estatísticos fornecem ainda o *p-value* (p-valor ou nível descritivo), parâmetro que representa o menor nível de significância que pode ser assumido para se rejeitar a hipótese de aderência (Chwif e Medina, 2010), ou seja:

- se $p\text{-value} \leq \alpha$, então H_0 é rejeitada ao nível de significância α ;
- se $p\text{-value} > \alpha$, então H_0 não é rejeitada ao nível de significância α .

A quinta etapa da metodologia consiste em codificar o modelo em uma linguagem de simulação apropriada. É possível utilizar a programação a partir de alguma linguagem específica de simulação ou utilizar *softwares* de simulação, como Arena ou ProModel. As linguagens de simulação são extremamente flexíveis, enquanto os *softwares* específicos podem reduzir consideravelmente o tempo de implementação. A escolha deverá se basear no tempo disponível para condução do projeto, no usuário final do sistema desenvolvido e nos objetivos definidos inicialmente.

As etapas seis e sete dizem respeito, respectivamente, à verificação e validação do modelo de simulação implementado. A verificação diz respeito à implementação computacional do modelo, garantindo que as funções e procedimentos programados estejam funcionando corretamente. Nesta etapa, pode-se utilizar alguns procedimentos para facilitar o processo de verificação, como a implementação modular, utilização de depuradores, simulação manual para sistemas mais simples e animação gráfica. Caso algum problema seja encontrado, retorna-se para o passo 5 de tradução do modelo.

A validação de modelos de simulação diz respeito à sua representatividade com relação ao sistema real. De acordo com Pidd (1998), se o modelo se comportar de acordo com o mundo real sob as mesmas condições, então ele é válido. A qualidade e a validade de um modelo de simulação são medidas pela proximidade entre os resultados obtidos pelo modelo e aqueles originados do sistema real. Uma vez que pressupostos e simplificações sobre o comportamento do sistema real são freqüentes no desenvolvimento do modelo, qualquer tomada de decisão com base em seus resultados deve ser precedida de uma avaliação de sua qualidade e representatividade.

A validação de modelos de simulação é um assunto já discutido por diversos autores, como Robinson (1997) e Sargent (2007). Suas técnicas vão desde métodos quantitativos (teste *t*, *chi-square test*) até procedimentos mais qualitativos (*Teste de Turing*, percepção dos operadores do sistema). Caso o modelo não possa ser validado, retorna-se para os passos 3 e 4 de modelagem conceitual e coleta de dados, respectivamente.

A etapa 8 da metodologia consiste em determinar as alternativas a serem simuladas. Nesta fase, para cada configuração analisada, define-se os períodos transiente e permanente, a duração da simulação e o número de replicações. Sistemas de simulação podem ser classificados em terminais e não-terminais. Os primeiros apresentam condições iniciais fixas (condições estas que o sistema volta a assumir ao início de cada rodada de simulação) e um evento que determina um fim natural para o processo de simulação (isto é, pode-se programar o fim de uma rodada de simulação) (FREITAS FILHO, 2008).

Para sistemas não terminais, é necessário identificar os períodos transiente e permanente. O período transiente é caracterizado por uma maior variabilidade das variáveis de resposta, que sofrem uma influência maior das condições iniciais do sistema. Uma vez que as análises estatísticas recaem sobre o desempenho do sistema quando o mesmo se encontra em regime permanente, é preciso que se reduza a influência da fase transiente na simulação do modelo.

Definido os períodos transiente e permanente do sistema, torna-se necessária a definição do número de replicações exigidas para se atingir um grau de precisão desejado para as variáveis de resposta. Por lidar com parâmetros de entrada estocásticos, a abordagem

mais comum e de fácil implementação é o uso de intervalos de confiança, conforme equação 4,

$$\left(\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right) \quad (4)$$

onde n representa o número de replicações da simulação inicial, \bar{x} a média da amostra, s o desvio padrão da amostra, α a confiança desejada e $t_{n-1, \alpha/2}$ o $(1 - \alpha/2)$ percentil da distribuição t de Student com $n-1$ graus de liberdade. A precisão h do intervalo é dada pela equação 5.

$$t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \quad (5)$$

A metodologia para determinação do número de replicações, apresentada em Chwif e Medina (2010) se baseia na utilização de uma amostra piloto. O modelo desenvolvido deve ser replicado um número n de vezes e construído o intervalo de confiança para o parâmetro de desempenho escolhido. Se a precisão h do intervalo obtido ainda não for menor do que a desejada h^* , então, o número de replicações necessário pode ser obtido por:

$$n^* = \left\lceil n \left(\frac{h}{h^*} \right)^2 \right\rceil \quad (6)$$

A etapa 9 consiste na execução de corridas de simulação para os cenários levantados anteriormente e sua posterior análise. Esta etapa é importante para estimar as medidas de desempenho do sistema em estudo. Como os dados de entrada geralmente apresentam uma alta variabilidade associada, é importante que a técnica utilizada para estimação dos parâmetros de desempenho leve em consideração esta incerteza, como por exemplo a utilização de intervalos de confiança. Esta análise vai ser utilizada pelo analista para determinar se novas corridas de simulação são necessárias (etapa 10). Além disso, testes estatísticos de inferência devem ser usados para determinar se uma configuração é melhor do que outra com relação aos parâmetros de desempenho analisados (ALTIOK e MELAMED, 2007).

Na etapa 11 da metodologia é feita a documentação do programa e do progresso do projeto. De acordo com Chwif e Medina (2007), documentar um projeto envolve

armazenar, catalogar e descrever de maneira funcional a especificação do projeto, as reuniões realizadas e suas atas, os e-mails e comunicações trocadas ao longo do projeto, o modelo conceitual, com as hipóteses e simplificações assumidas, o modelo de entrada de dados, o modelo computacional com uma descrição detalhada das suas variáveis, atributos, rotinas etc., a verificação e validação realizadas, os cenários analisados e seus resultados computacionais, as apresentações realizadas para a equipe, clientes, etc. e os relatórios parciais e final do projeto.

A etapa 12 consiste na implementação do estudo desenvolvido. Esta fase depende diretamente do envolvimento do usuário final do sistema nas fases discutidas anteriormente, de maneira que o mesmo consiga entender a operação do modelo de simulação, os dados de saída gerados e as vantagens de sua implementação.

4.2 METODOLOGIA PARA CONSTRUÇÃO DO METAMODELO

A construção do metamodelo baseou-se nos passos da metodologia apresentada na Figura 3, adaptada do trabalho de Dos Santos (2002).

Na primeira etapa, de formulação do problema, determina-se qual o objetivo da aplicação do metamodelo, quais variáveis serão utilizadas para modelagem (variável dependente e variáveis independentes) e a região experimental. O objetivo principal de um metamodelo é estimar ou aproximar a superfície de resposta da variável em estudo. De acordo com Dellino e Meloni (2011), este procedimento pode ser utilizado para entender como a superfície de resposta se comporta em diferentes pontos da região experimental, estimar como a resposta se altera ao redor de uma solução específica (com pequenas variações dos parâmetros de entrada) ou encontrar (estimar) a configuração ótima dos dados de entrada.

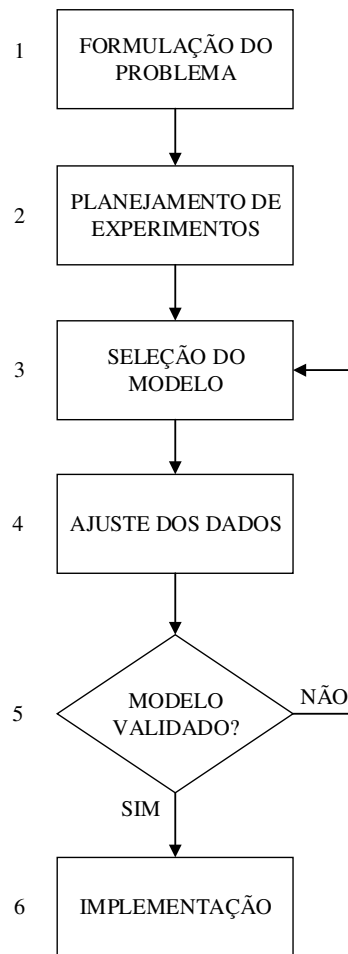


Figura 3: Metodologia para construção do metamodelo
 Fonte: Adaptado de Dos Santos (2002)

De acordo com Yaesoubi (2006), modelos computacionais de simulação que visam representar um sistema real, geralmente consistem em um grande número de parâmetros de entrada (ou fatores, na linguagem de planejamento de experimentos). As saídas de um modelo de simulação (respostas) podem ser exploradas utilizando experimentos planejados. Neste contexto, a seleção (*factor screening*) das variáveis independentes cumpre um papel importante na construção de metamodelos. Sua lógica envolve experimentos aonde muitos fatores são examinados com o objetivo de identificar quais têm uma significância maior para explicar a variável de resposta do sistema. Nesta etapa, sugere-se como referência os trabalhos de Yaesoubi (2006) e Kleijnen (2009b).

A etapa 2 consiste no planejamento de experimentos (*design of experiments*) para geração dos dados que serão utilizados na construção do metamodelo. Uma escolha cuidadosa dos pontos a serem analisados é essencial para predizer a resposta do sistema

considerando cenários ainda não analisados (PISTONE e VICARIO, 2010). Nesta etapa, abordagens mais comuns são aquelas baseadas em experimentos fatoriais ou amostragem utilizando do modelo hipercubo latino (MHL).

Experimentos fatoriais completos consistem em dois ou mais fatores, cada um com valores possíveis discretos (ou níveis), cujas unidades experimentais assumem todas as possíveis combinações destes níveis. O planejamento fatorial é indicado para a fase inicial do procedimento experimental quando há necessidade de se definir os fatores mais importantes e estudar os efeitos sobre a variável resposta escolhida (GALDÁMEZ, 2002). Esta abordagem é baseada no modelo de análise de variância e permite estimativas de vários efeitos e interações entre os fatores (Meckesheimer, 2001). Entretanto, não é uma abordagem adequada quando considerado um número grande de variáveis. Para lidar com esta limitação, pode-se utilizar o planejamento de experimentos através do modelo hipercubo latino.

O MHL foi proposto por McKay, Beckman e Conover (1979) como uma extensão da amostragem estratificada. Este método, de acordo com Salas e Shin (1999), reduz o número de conjuntos de parâmetros necessários para os experimentos e, conseqüentemente, reduz o número de respostas geradas bem como o custo computacional. O conceito básico desse método está em gerar números randômicos de uma variável, dentro de seus limites, de uma maneira estratificada, de forma que a variabilidade como um todo do parâmetro possa ser razoavelmente delineada por uma amostra de tamanho limitado. Fang *et al.* (2006), em suas comparações, observaram que a amostragem por hipercubo latino é mais vantajosa em relação a outros métodos porque o número de simulações pode ser reduzido consideravelmente para alcançar o mesmo nível de precisão. Uma abordagem comparativa dos métodos de geração dos cenários pode ser encontrada no trabalho de Kleijnen (2005).

Definido o conjunto de pontos da amostra, torna-se necessária a escolha do modelo para representação dos dados gerados. A etapa 3 consiste na escolha deste modelo, utilizado para estimar a variável dependente do sistema. Dos Santos (2002) afirma que uma vez que existem diversos tipos de metamodelos disponíveis, é preciso saber como se fazer uma escolha entre eles, apresentando para isso alguns critérios: a capacidade de captar a essência do problema a partir da forma do metamodelo, a facilidade de cálculo da

função de estimação f e a existência de software para calcular o metamodelo, caracterizar seu ajuste e usá-lo para previsão. A seguir uma breve descrição de alguns tipos mais usuais encontrados na literatura.

Modelos de regressão polinomiais têm sido largamente utilizados para construção de metamodelos (Durieux e Pierreval, 2004; Dos Santos e Dos Santos, 2009; Chi, 2013) devido principalmente à facilidade de construção dos estimadores. Tais modelos são baseados no ajuste de uma função polinomial às variáveis de resposta do modelo de simulação. Este tipo de metamodelo pode ser utilizado para explicar o comportamento do modelo de simulação e predição das variáveis de resposta para valores de entrada que ainda não foram simulados (Kleijnen, 2009a). Entretanto, tais modelos geralmente não produzem estimativas satisfatórias fora da região experimental, nem para valores extremos da variável dependente (Dos Santos, 2002).

De acordo com Kleijnen (2009a), metamodelos de Kriging são ajustados geralmente a dados obtidos de grandes experimentos. A idéia básica está na predição do valor de uma função em um determinado ponto a partir do cálculo de uma média ponderada dos valores conhecidos desta função na vizinhança do ponto. Uma revisão de modelos de Kriging pode ser encontrada no trabalho de Kleijnen (2009a). Esta abordagem já foi utilizada em vários contextos, como apresentado nos trabalhos de Biles *et al.* (2007) e Zarekifar *et al.* (2009).

Outras abordagens envolvem redes neurais artificiais (Fonseca *et al.*, 2003), funções base radiais (Ji e Kim, 2013) e MARS (*multivariate adaptive regression splines* – Alizadeh e Xia, 2014). Estudos comparativos entre diferentes metamodelos podem ser encontrados nos trabalhos de Barton e Meckesheimer (2006) e Li *et al.* (2010).

A etapa 4 da metodologia consiste no ajuste dos dados gerados (etapa 2) utilizando o metamodelo escolhido (etapa 3). A estimação de um metamodelo consiste em utilizar técnicas estatísticas para estimar os parâmetros desconhecidos do metamodelo hipotético (Dos Santos, 2002). Esta estimação deve ser feita utilizando métodos estatísticos, como o método dos mínimos quadrados, através da determinação do mínimo global para a soma dos quadrados dos erros (ou resíduos), comparando os valores preditos com os observados.

O metamodelo gerado deve ser analisado com relação à sua validade (etapa 5) para poder ser implementado (etapa 6) buscando atender aos objetivos propostos inicialmente. O processo de validação de metamodelos tem como objetivo avaliar a capacidade geral de previsão com relação ao modelo computacional de simulação. Este processo resume-se basicamente em três passos: a verificação da adequação do metamodelo com relação às hipóteses assumidas, a validação do metamodelo com respeito ao modelo de simulação (se as estimativas estão suficientemente próximas daquelas geradas pela simulação, tendo em conta o objetivo principal da aplicação) e a validação do metamodelo com relação ao problema real.

As hipóteses relacionadas à distribuição dos dados de entrada para o ajuste do metamodelo dizem respeito à variância constante e a normalidade dos erros. Utiliza-se geralmente a análise gráfica para verificação de tais hipóteses, através do gráfico de resíduos versus variáveis independentes, resíduos versus valores ajustados e gráfico de probabilidade normal dos resíduos. Para mais detalhes, sugere-se o livro de Harrell (2001).

O método preferencial para validar um metamodelo em relação ao modelo de simulação, é através da coleta de novos resultados de simulação. Uma alternativa razoável, quando o conjunto de resultados é suficientemente grande, é dividir os resultados em dois conjuntos. O primeiro, denominado conjunto de construção do modelo, é utilizado para desenvolver o metamodelo. O segundo conjunto de resultados, denominado conjunto de validação, é utilizado para avaliar a razoabilidade e a capacidade preditiva do metamodelo. Este procedimento de validação é frequentemente denominado de validação cruzada (DOS SANTOS, 2002).

Outras métricas podem também ser utilizadas para auxiliar no processo de análise da qualidade das previsões, como o erro absoluto máximo (*maximum absolute error* – MAE), erro absoluto médio (*average absolute error* – AAE) e o coeficiente de determinação (*coefficient of determination* – R^2).

Para maiores detalhes do processo de validação de metamodelos, sugere-se os trabalhos de Kleijnen *et al.* (1998), Iooss *et al.* (2010) e Hamad (2011).

4.3 METODOLOGIA DE OTIMIZAÇÃO VIA SIMULAÇÃO

O campo de pesquisa na área de otimização para simulação tem como foco o uso da simulação para projeto e otimização de sistemas. O uso da simulação para estimativa da performance de um sistema geralmente requer um tempo computacional elevado, sendo que a determinação da configuração ótima é um problema difícil de tratar devido à necessidade de avaliar a performance do sistema para diferentes configurações.

O problema de otimização via simulação é um problema de otimização em que a função objetivo, restrições ou ambas são respostas que somente podem ser avaliadas através da simulação computacional. Desta maneira, estas funções são apenas funções implícitas e parâmetros de decisão do sistema (AZADIVAR, 1992).

Existem diversos métodos sugeridos na literatura que podem ser utilizados para geração de cenários no processo de otimização via simulação. Podemos citar, como exemplos, otimização estocástica, metodologia de superfície de resposta, métodos estatísticos, método do gradiente e métodos heurísticos. Estes últimos, conforme mencionado, estão presentes em diversos pacotes de simulação integrados com otimizadores, como o *OptQuest* e *SimRunner*. Os algoritmos heurísticos de busca fornecem resultados rápidos e confiáveis para uma extensa variedade de problemas, e serão utilizados na metodologia proposta.

A Figura 4 apresenta a metodologia clássica de otimização via simulação (Fu, 2002). O processo começa com a simulação da configuração inicial (θ^0) do sistema. As medidas de desempenho do sistema geradas pelo modelo de simulação são utilizadas pelo método heurístico para geração de uma nova solução candidata (θ^{i+1}). O processo se repete até que algum critério de parada seja satisfeito, quando o modelo retorna a melhor solução encontrada até o momento (θ^*).

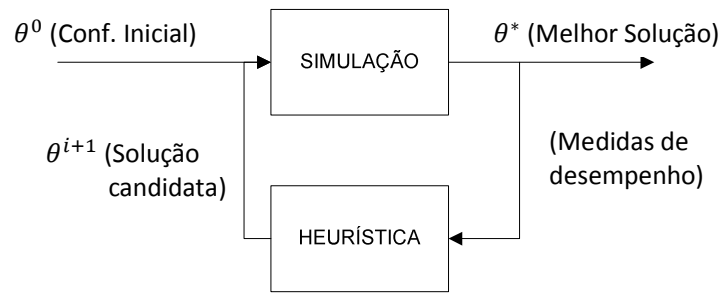


Figura 4: Metodologia de otimização via simulação
Fonte: Fu (2002)

O diferencial na metodologia proposta no presente trabalho está no uso de metamodelos para avaliação das soluções candidatas geradas pelo método heurístico (Figura 5). A etapa que consome maior tempo computacional é o processo de simulação das alternativas. A ideia se baseia em fazer uma aproximação do tempo de resposta da solução candidata utilizando do metamodelo, cujo tempo de processamento é muito menor. Antes de ser simulado, o cenário é analisado pelo metamodelo, que funciona como um filtro. Somente as soluções que ficarem a uma distância máxima α com relação ao valor padrão estipulado para a variável de resposta (VR) do sistema é que serão simuladas ($|VR_{padr\tilde{a}o} - VR_{metamodelo}| \leq \alpha$). Aqueles valores retornados pelo metamodelo que ficarem muito distantes do padrão não serão analisados via simulação ($|VR_{padr\tilde{a}o} - VR_{metamodelo}| > \alpha$), retornando para o processo de geração de um novo cenário.

De maneira semelhante, o processo começa com a simulação da configuração inicial do sistema (θ^0). O componente de simulação irá gerar um vetor de variáveis de resposta $\delta(\theta^i)$ com base nos parâmetros de entrada fornecidos. A metodologia verifica se algum dos critérios de parada foi atingido. Em caso positivo, retorna a melhor solução encontrada até o momento (θ^*). Caso contrário, o vetor $\delta(\theta^i)$ será utilizado pelo procedimento heurístico para gerar uma nova configuração. Os critérios de parada podem utilizar, por exemplo, um tempo máximo de máquina ou um número limite de iterações sem melhora no valor da função objetivo. A heurística irá gerar um novo vetor de entrada (θ^{i+1}) a ser avaliado pelo metamodelo ajustado. A metodologia verifica se o valor estimado para a variável de resposta a partir do metamodelo (VR_m) está próximo do valor estipulado como restrição para esta mesma variável (VR_p). Em caso afirmativo, a solução é refinada pelo componente de simulação. Em caso negativo, a

metodologia entende que o cenário sugerido está longe da região de viabilidade do problema e não necessita ser refinado via simulação. O procedimento retorna para a heurística que irá gerar uma nova configuração candidata.

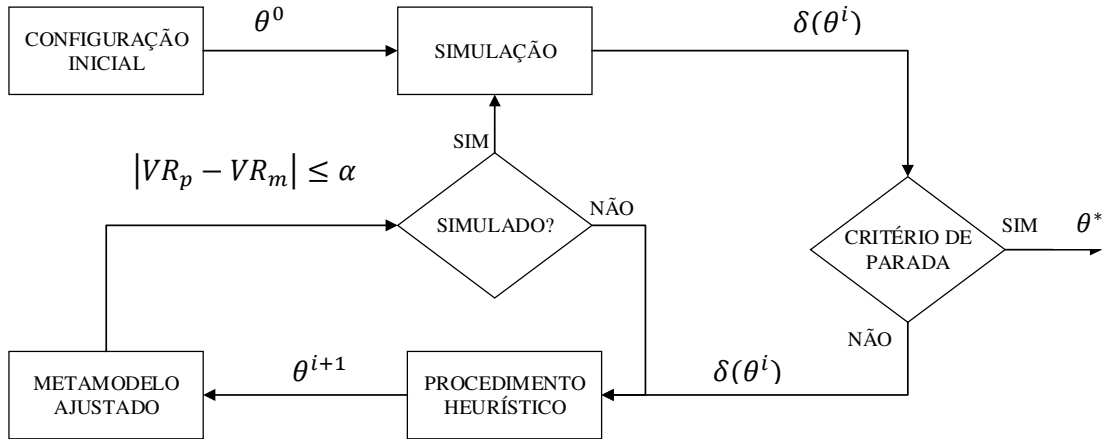


Figura 5: Metodologia proposta de otimização via simulação
 Fonte: Elaborado pelo autor (2014)

5. APLICAÇÃO DA METODOLOGIA PROPOSTA

A metodologia apresentada na seção 4 foi validada utilizando dados do ano de 2009 relativos ao Serviço de Atendimento Móvel de Urgência da cidade de Belo Horizonte. A escolha deste período foi devido a dois motivos principais. Primeiro, existe uma dificuldade grande de acesso à base de dados do sistema bem como autorização para acompanhamento das atividades do serviço para coleta de dados e tempos relevantes das operações. Segundo, no ano de 2009 foi desenvolvido um trabalho (Silva, 2010) voltado para a área de simulação de processos relativo ao mesmo Serviço de Atendimento, onde foi possível realizar uma extensa coleta e validação de dados. Como o objetivo do presente trabalho é o de validar a metodologia, independente do ano de operação do sistema, optou-se por trabalhar com os dados de 2009 que possuem uma qualidade e validação asseguradas. Para maiores detalhes, sugere-se o trabalho de Silva (2010).

A modelagem voltada para otimização do sistema via simulação é baseada na proposta de Fu (2002). A função objetivo (7) procura direcionar a busca para determinação da melhor configuração.

$$\text{Min}_{\theta \in \Theta} K(\theta) = E[K(\theta, \omega)] \quad (7)$$

O índice θ representa o conjunto das variáveis de entrada, ω é a replicação, $K(\theta, \omega)$ é a medida de desempenho e $E[K(\theta, \omega)]$ é o valor esperado da medida de desempenho.

Seja B_{bk} o conjunto de características de cada base de ambulância, em que B_{b1} é uma variável booleana que indica se uma base será ativada ou não. Seja Q_u o número total de ambulâncias tipo u ($1 = \text{básica}$, $2 = \text{avançada}$), β_b o custo de ativação de cada base b , η_u o custo de uma ambulância do tipo u , T_{bu} o número de ambulâncias do tipo u na base b , TR_u o tempo de resposta das unidades do tipo u , TU_u a taxa de utilização das unidades do tipo u e n_b o número total de bases .

Para determinar a melhor configuração para o sistema considerando diversos pontos potenciais de localização de bases, pode-se utilizar um modelo de otimização via simulação dado por (8) a (16), conforme apresentado a seguir (Pinto *et al.*, 2015).

$$\text{Min } Z = \sum_{b=1}^{n_b} \beta_b B_{b1} + \sum_{u=1}^2 \eta_u Q_u \quad (8)$$

$$Q_u - \sum_{b=1}^{n_b} T_{bu} \geq 0 \quad \forall u \quad (9)$$

$$TR_u \leq TR_u^* \quad \forall u \quad (10)$$

$$TU_u \leq TU_u^* \quad \forall u \quad (11)$$

$$T_{bu} \geq T_{bu}^- \quad \forall b, \forall u \quad (12)$$

$$Q_u \geq Q_u^- \quad \forall u \quad (13)$$

$$T_{bu} \leq T_{bu}^+ * B_{b1} \quad \forall b, \forall u \quad (14)$$

$$Q_u \leq Q_u^+ \quad \forall u \quad (15)$$

$$B_{b1} \in \{0,1\} \quad \forall b \quad (16)$$

onde B_{b1} , Q_u e T_{bu} são os controles e TR_u e TU_u são as respostas. A restrição (9) garante que o número de unidades u em cada base não ultrapasse o número total de unidades permitidas no sistema; as restrições (10) e (11) fornecem limites superiores para o tempo de resposta e taxa de utilização para cada tipo de unidade u , respectivamente e as restrições (12) a (16) estabelecem os limites inferiores e superiores para os controles.

A seção 5.1 apresenta maiores detalhes da módulo de simulação, a seção 5.2 as heurísticas propostas e a seção 5.3 o metamodelo desenvolvido.

5.1 MÓDULO DE SIMULAÇÃO

A metodologia de simulação proposta se baseia em três processos principais, que apesar de poderem ser desmembrados em atividades mais complexas, são bastante semelhantes entre sistemas de atendimento de emergência: geração das chamadas, despacho das ambulâncias e operações relativas às unidades móveis (Figura 6).

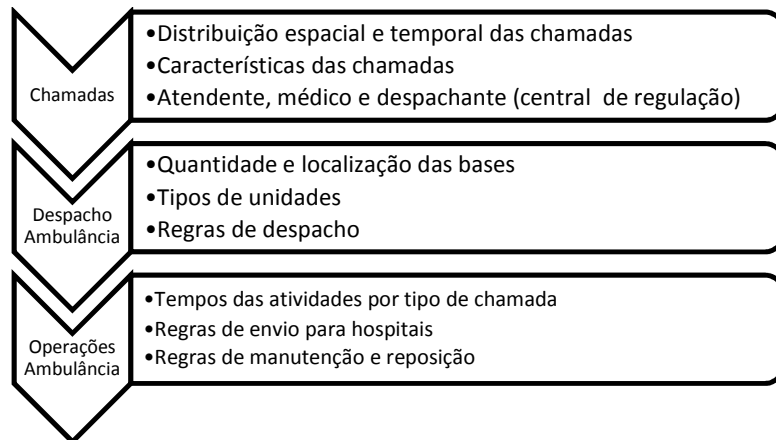


Figura 6: Componentes principais da metodologia de simulação

Fonte: Elaborado pelo autor (2014)

As características de cada parâmetro da metodologia serão detalhadas nas subseções a seguir e maiores detalhes podem ser obtidos no trabalho de Pinto *et al.* (2015).

Um dos aspectos inovadores da metodologia, que simplifica bastante sua implementação, é a modelagem de todo o sistema de atendimento através das coordenadas UTM (*Universal Transverse Mercator*), de maneira a determinar o ponto de geração do incidente, a localização das bases e hospitais e os tempos de deslocamento das ambulâncias. Os sistemas de projeções cartográficas são analisados pelo tipo de superfície de projeção adotada e grau de deformação. Na projeção UTM, utiliza-se um cilindro tangente à superfície da Terra como superfície de projeção, sendo os meridianos e paralelos representados por retas perpendiculares. O sistema UTM resulta na composição de 60 fusos distintos que representam a superfície da Terra.

No início da simulação, todas as entradas relativas às características do sistema são lidas. A seguir, o processo de simulação tem início e todas as unidades móveis são geradas, definindo sua localização e tipo (básica, avançada ou de saúde mental). As chamadas são geradas baseadas em taxas específicas para cada faixa de horário e suas características de acordo com estimativas definidas na fase de coleta de dados. As chamadas são recebidas por teledigifonistas que fazem a triagem inicial baseada nos atributos específicos das chamadas. Aquelas que necessitam de atendimento são atendidas pelos médicos reguladores e, caso seja constatada a necessidade de envio de uma ambulância, a chamada é transferida para um dos despachantes. Após o despacho, as unidades se deslocam até o local do incidente, realizam o atendimento, transportam o

paciente para o hospital mais próximo e, se for o caso, realizam a reposição de equipamentos e/ou manutenção. Estas duas atividades são realizadas sempre ao final do ciclo de atendimento. Ao término da simulação, estatísticas básicas são geradas para estimativa das variáveis de interesse, conforme Tabela 1.

Tabela 1: Variáveis de interesse

Variável	Descrição
TR	Tempo de resposta médio do sistema (min)
TR_u	Tempo de resposta médio para unidades do tipo u (min)
TU_u	Taxa de utilização das ambulâncias do tipo u (%)
PT	Percentual de chamadas com $TR \leq TR^*$ (%)
PT_u	Percentual de chamadas do tipo u com $TR \leq TR^*$ (%)

Fonte: Elaborado pelo autor (2014)

5.1.1 CARACTERÍSTICAS DAS CHAMADAS

As chamadas possuem diversas características sazonais relacionadas com taxas de chegada, distribuição espacial, natureza da ocorrência, tipo de ocorrência e tipo de ambulância necessária para atendimento. Estas informações são tratadas em quatro etapas. Primeiro as chamadas são classificadas de acordo com sua natureza (chamada médica ou outro motivo, como por exemplo, trote ou engano), utilizando-se de dados históricos. Após, as chamadas médicas são classificadas de acordo com o tipo de ocorrência. Na literatura existem basicamente quatro classes distintas de emergência (Allen *et al.*, 1973; Siddharthan *et al.* 1996; NHS, 2012), que foram agrupadas em clínica, trauma, psiquiátrica e gineco-obstétrica. Esta informação é utilizada para encaminhamento da vítima aos hospitais, uma vez que nem todos os centros médicos conseguem atender todos os tipos de ocorrência. Terceiro, define-se a conduta médica a ser adotada (envio de ambulância ou orientação pelo telefone). Por fim, caso seja identificada a necessidade de envio de uma ambulância, define-se o tipo de unidade mais apropriada (básica, avançada ou de saúde mental). Estas etapas são realizadas pelos teledigifonistas, médicos reguladores e despachantes na central de regulação.

5.1.2 GERAÇÃO DOS INCIDENTES

Serviços de atendimento de emergência lidam com uma variabilidade grande no processo de requisição dos serviços, tanto ao longo da semana como ao longo do dia. Diversos estudos já foram realizados de maneira a estimar taxas de chegada para sistemas de atendimento de emergência. Mais recentemente, Matteson *et al.* (2011) propuseram uma metodologia utilizando uma estrutura dividida em dia da semana e hora do dia, aplicada com sucesso na cidade de Toronto (Canadá). A suposição de que o processo de chegada acontece de acordo com um processo de Poisson não-homogêneo é largamente utilizada. Cabe ressaltar que dentro de cada faixa de horário considerada, o processo é homogêneo, mas o mesmo varia ao longo das 24 horas do dia.

Para caracterizar esta aleatoriedade, a Figura 7 mostra a quantidade de chamadas por faixa de horário (considerando dias úteis e final de semana) para a cidade de Belo Horizonte ao longo do ano de 2009. Pode-se observar que para uma modelagem adequada do processo de chegada das chamadas, é importante considerar as faixas de horário bem como o dia da semana. A metodologia proposta discretiza estas faixas em 168 intervalos (24 intervalos para cada dia e 7 dias na semana).

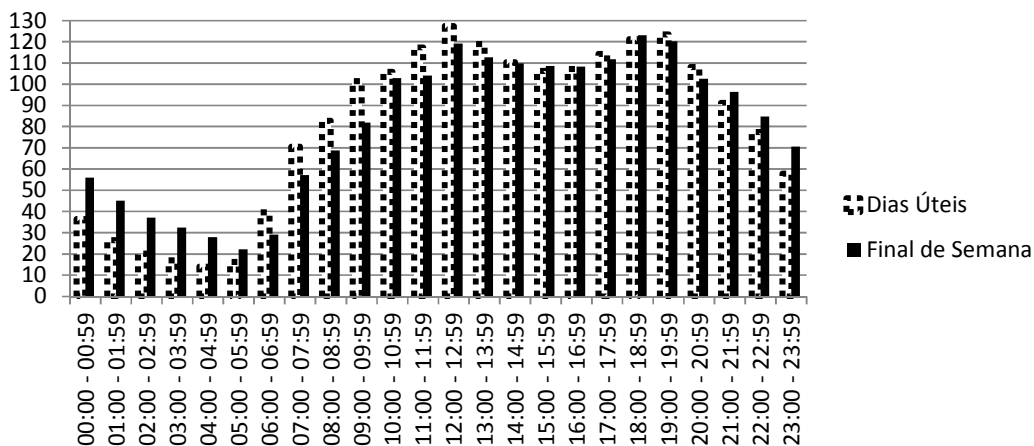


Figura 7: Número de chamadas por faixa de horário (SAMU-BH - 2009)

Fonte: Adaptado de Prodabel (2009)

Na implementação computacional da metodologia, para determinar a melhor maneira de geração dos incidentes, foram testadas duas abordagens diferentes, utilizando o SAMU da cidade de Belo Horizonte para escolha da mais adequada. A primeira divide o território considerado em áreas menores (Figura 8), utilizando o ponto central da área como ponto de origem da chamada. Foram testadas divisões do território em 1 área

(ocorrências sempre no mesmo ponto central da cidade – Figura 8a), 9 áreas (número de regionais – Figura 8b), 74 áreas (número de unidades de planejamento de acordo com divisão da Prefeitura – Figura 8c) e 377 áreas (número de bairros da cidade onde foram observadas pelo menos uma ocorrência – Figura 8d).

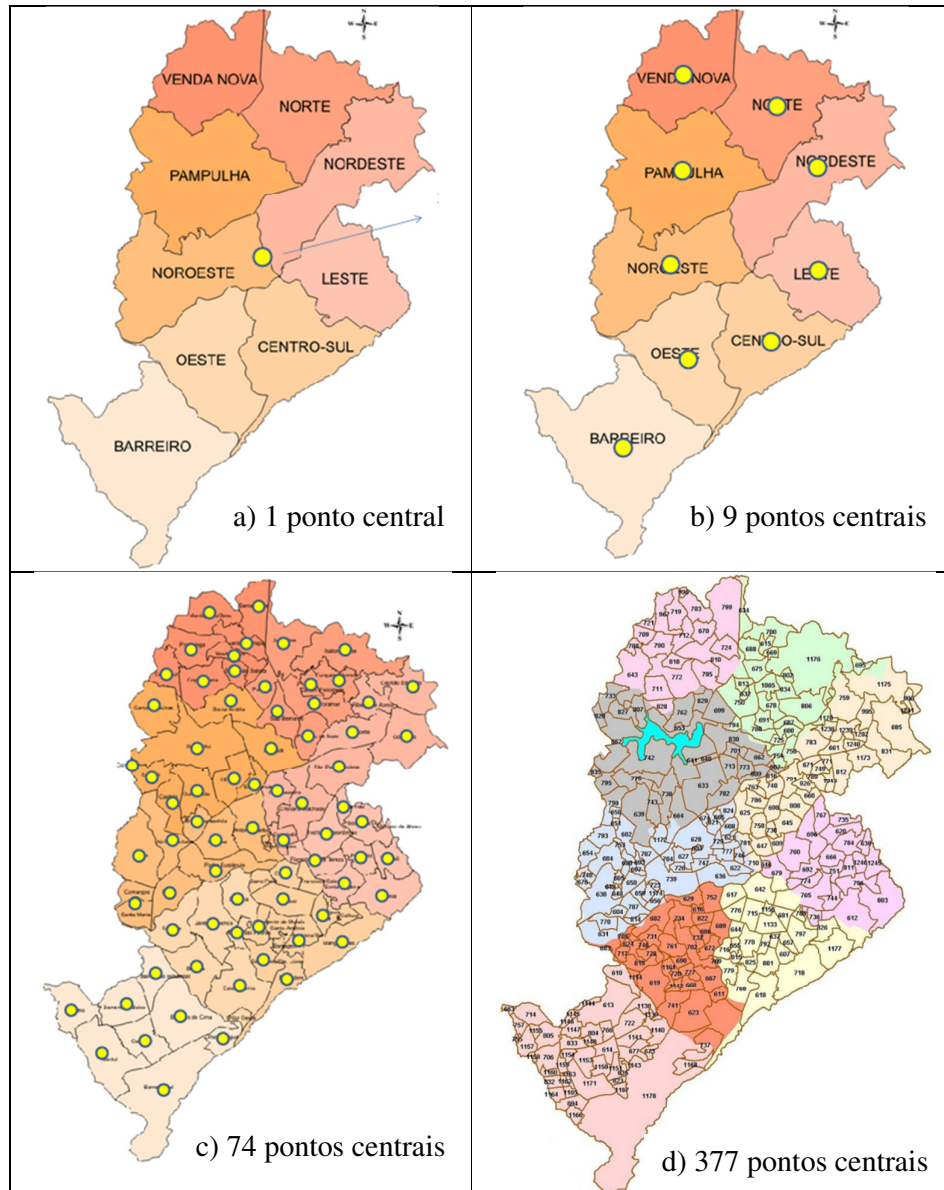


Figura 8 - Alternativas para geração dos incidentes

Fonte: Adaptado de Nogueira Junior (2011)

A outra abordagem utiliza o mapeamento do território através do sistema UTM de coordenadas, de acordo com valores retirados de aplicativo desenvolvido pela Prodabel e disponível no site da prefeitura de Belo Horizonte. O contorno de cada regional (9 para a cidade de Belo Horizonte) foi aproximado por um polígono (linha vermelha na

Figura 9), buscando considerar da melhor maneira possível os limites da região. Para sorteio das coordenadas do incidente, utilizaram-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 (linhas mais grossas da Figura 9).

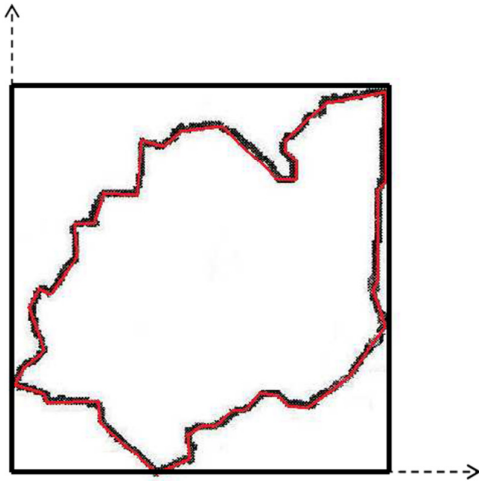


Figura 9: Exemplo de mapeamento de uma regional
Fonte: Silva (2010)

Uma rotina em VBA (*Visual Basic for Applications*) foi implementada para verificar se o ponto sorteado se encontra dentro ou na fronteira do polígono traçado para representar aquela região (detalhes na seção 5.1.3). Caso não esteja, um novo ponto é sorteado e verificado até que a condição seja satisfeita. Desta maneira, qualquer ponto dentro da regional é um candidato para geração de um incidente, todos com igual probabilidade. O SAMU de BH foi simulado considerando estas duas abordagens e o valor do tempo de resposta médio do sistema foi comparado com o tempo de resposta médio “real” (estimativa de tempos coletada pelo pesquisador na central de regulação, sendo igual a 19,92 minutos para o ano de 2009). A Figura 10 apresenta a comparação feita.

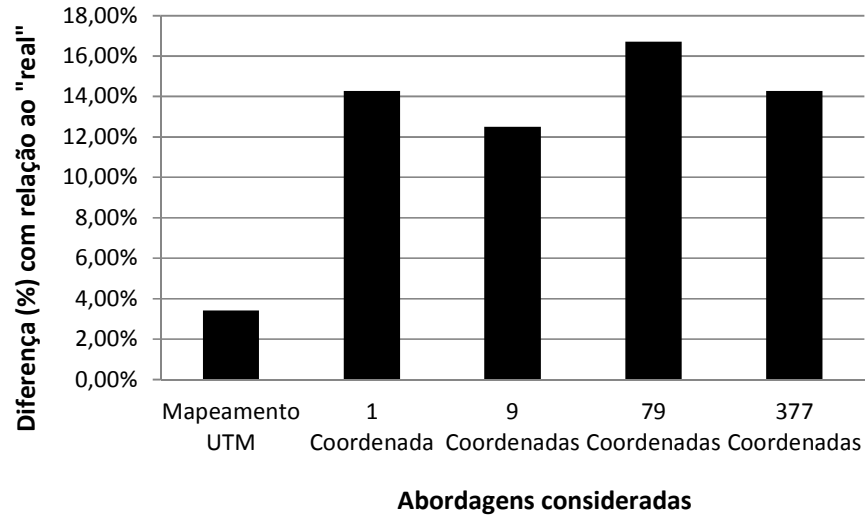


Figura 10: Abordagens para geração dos incidentes
 Fonte: Elaborado pelo autor (2014)

Pode-se observar que a segunda abordagem baseada no mapeamento das regionais via coordenadas UTM apresentou resultado significativamente mais próximo do “real”, de maneira que foi a abordagem escolhida para representar o processo de geração dos incidentes.

5.1.3 DETERMINAÇÃO SE O INCIDENTE OCORREU DENTRO DA REGIONAL

Este problema pode ser formulado da seguinte maneira: dado um polígono simples $P = p_1p_2 \dots p_n$ e um ponto p , verificar se o ponto é interior, exterior ou está na fronteira do polígono.

Segundo Figueiredo e Carvalho (1991), o problema faz sentido, já que um polígono fechado e simples separa o plano em duas regiões conexas e abertas, uma limitada e a outra não, ambas tendo o polígono como fronteira.

Uma solução para o problema consiste em considerar uma semirreta L partindo de p e determinar seus pontos de interseção com o polígono. Se p coincidir com um destes pontos de interseção, conclui-se que ele está na fronteira de P . Senão, basta contar quantas vezes a semirreta atravessa o polígono. No infinito, L se encontra na região

exterior. Logo, se o número de cruzamentos for ímpar, o ponto p é interior, caso contrário, p é exterior (FIGUEIREDO E CARVALHO, 1991).

No algoritmo utilizado, Figueiredo e Carvalho (1991) consideram a semirreta horizontal $L = \{(x_0, y_0) + t(1,0) \mid t \geq 0\}$. A Figura 11 apresenta o algoritmo utilizado.

```

N = 0
for i = 1 to n do
  if  $y_i \neq y_{i+1}$  then
     $(x,y) =$  a interseção de  $p_i, p_{i+1}$  com L
    if  $x = x_0$  then
       $p_0$  é fronteira: PARE
    else if  $x > x_0$  and  $y > \min\{x_i, x_{i+1}\}$  then
       $\nexists$  y mínimo
       $N = N + 1$ 
    end if
  else if  $p \in$  ao lado horizontal  $p_i, p_{i+1}$  then
     $p$  é da fronteira: PARE
  end if
end for
if  $N \bmod 2 \neq 0$  then
   $p$  é interior a P
else
   $p$  é exterior
end

```

Figura 11: Algoritmo para determinar se ponto está dentro de um polígono
Fonte: Adaptado de Figueiredo e Carvalho (1991)

Nele, os pontos de interseção de L com o polígono são obtidos calculando o ponto de interseção de L com cada lado do polígono. Quando um ponto de interseção é um vértice, não é necessariamente verdade que L passe do interior para o exterior (ou vice-versa) naquele ponto. No caso da Figura 12, por exemplo, o ponto a não deve ser contado como ponto de interseção, enquanto b deve. Problemas surgem também quando L contém um lado do polígono. Uma forma prática de solucionar estes conflitos é contar a interseção de um lado $p_i p_{i+1}$ com L somente se ela ocorrer em um ponto que não seja de ordenada mínima em $p_i p_{i+1}$.

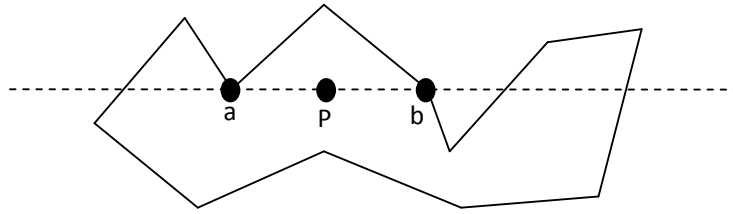


Figura 12: Exemplo do problema de ponto no polígono
 Fonte: Figueiredo e Carvalho (1991)

Esse algoritmo funciona para qualquer polígono simples de n lados e tem complexidade computacional $O(n)$.

5.1.4 LÓGICA DE DESPACHO E CONTROLE DAS AMBULÂNCIAS

No processo de despacho estático, a ambulância se desloca para o local do incidente da base mais próxima que possui disponível uma unidade do tipo requisitado. Depois do atendimento, a ambulância retorna para base e aguarda um novo chamado. No despacho semi-dinâmico, a alocação das unidades pode ocorrer logo após a entrega do paciente no hospital ou término do atendimento no local do incidente. As ambulâncias não necessitam retornar à base de operação antes do próximo despacho. No processo dinâmico, as ambulâncias podem ser realocadas após o despacho para atender um chamado mais urgente, ou seja, elas ficam disponíveis mesmo após o empenho da unidade.

Billhardt *et al.* (2014) desenvolveram um modelo dinâmico de coordenação de ambulâncias, envolvendo um mecanismo para alocação das unidades e outro para redistribuição a partir de dados históricos do sistema, sendo aplicado na cidade de Madri (Espanha). Os autores afirmam que o modelo desenvolvido é mais adequado para situações em que vários pacientes precisam ser atendidos ao mesmo tempo. Em média, as duas abordagens desenvolvidas conseguiram reduzir o tempo de resposta do sistema em quase 16%. Lim *et al.* (2011) fizeram uma revisão de políticas de despacho dinâmicas comparando vantagens e desvantagens de cada tipo. O despacho dinâmico necessita de menos unidades por ser mais eficiente, mas se o número de realocações for pequeno, o desempenho do sistema fica bastante próximo quando comparado com o despacho semi-dinâmico, ou seja, este último pode ser utilizado como uma boa

aproximação para o caso dinâmico, além da facilidade de modelagem não havendo necessidade de utilização de sistemas de informação geográficos.

Para a metodologia proposta, considere que cada ambulância possui um índice específico armazenado em um vetor $I[]$, com número de posições igual a n . O modelo possui uma matriz de controle ($matriz[][]$) com informações de cada ambulância. Considere que a linha é referente às ambulâncias e a coluna (na posição c) informa se a ambulância é do tipo básica ou avançada, ou seja, a ambulância i será básica se $matriz[i][c] = 1$ ou avançada se $matriz[i][c] = 2$. A Figura 13 apresenta os passos da lógica de controle.

```

1  for i = 1 to n
2      V[i] = M
3  next i
4  if tipo_unid = usb then
5      for i = 1 to n
6          if matriz[i][c] = 1 then
7              if ( $h_i^i > t_c$ ) then
8                   $V[i] = (h_i^i - t_c) + f_c \times d_e/v_m$ 
9              else
10                  $V[i] = f_c \times d_e/v_m$ 
11             end if
12         end if
13     next i
14 else
15     for i = 1 to n
16         if matriz[i][c] = 2 then
17             if ( $h_i^i > t_c$ ) then
18                  $V[i] = (h_i^i - t_c) + f_c \times d_e/v_m$ 
19             else
20                  $V[i] = f_c \times d_e/v_m$ 
21             end if
22         end if
23     next i
24 end if
25 sort V[], I[]
26 return j ← I[1]

```

Figura 13: Lógica de acionamento e controle das ambulâncias

Fonte: Elaborado pelo autor (2014)

Seja j o índice da unidade a ser alocada a uma determinada chamada, n o número total de ambulâncias, h_i^i o instante de tempo em que a unidade i vai estar disponível de acordo com o tempo da simulação, t_c o tempo corrente da simulação, f_c o fator de correção da distância euclidiana, d_e a distância euclidiana, v_m a velocidade média, $V[]$ um vetor para armazenar o tempo de resposta de cada unidade, $tipo_unid$ um atributo de cada chamada que determina a necessidade por ambulância (básica ou avançada) e M um número muito grande.

O vetor $V[]$ é inicializado (passos 1 a 3) e o modelo verifica qual o tipo de unidade necessária (passo 4). Nos passos 6 e 16, a lógica de controle verifica o tipo de cada ambulância para identificar aquelas que são do tipo requisitado. Para estas, calcula-se o tempo de resposta considerando a hora em que a unidade vai estar disponível. Se a unidade estiver em atendimento, o tempo de resposta considera o tempo até a ambulância ficar disponível mais o tempo de deslocamento até o incidente (passos 7 e 8 / 17 e 18). Se a unidade já estiver disponível, o tempo de resposta é composto apenas pelo tempo de deslocamento (passo 10 / 20). Ao final das n iterações, o modelo ordena o vetor tempo de resposta de maneira crescente e aloca para a chamada a ambulância na posição 1 (passos 25 e 26). As ambulâncias são geradas com base no parâmetro Q_u e sua distribuição é dada por T_{ub} .

5.1.5 LÓGICA DE RETORNO DAS AMBULÂNCIAS

Uma vez disponível, caso não esteja em sua base de operação, a ambulância deve fazer a viagem de retorno, ficando indisponível durante este tempo. Para aproximar ainda mais a metodologia do caso dinâmico, uma verificação periódica da necessidade de empenho da unidade foi considerada durante este deslocamento.

Seja (x_1^0, y_1^0) as coordenadas de liberação da ambulância (podendo ser referentes ao local do incidente, a algum centro de atendimento médico, a central de regulação ou a oficina de manutenção) e (x_2^0, y_2^0) as coordenadas da base de origem da unidade móvel. O trajeto compreendido entre estes dois pontos pode ser dividido em segmentos menores de maneira que, a cada segmento percorrido o modelo atualiza as coordenadas e o horário de liberação da ambulância e verifica se existe alguma chamada alocada a ela. Em caso afirmativo, interrompe o deslocamento de retorno à base e se desloca para

o local do incidente. Em caso contrário, continua deslocando até o próximo segmento, fazendo a mesma verificação. O processo continua até que a ambulância tenha atingido sua base de origem. Considere também que o número total de segmentos seja igual a n . Tem-se as seguintes possíveis situações para a posição de saída (S) e chegada de uma unidade móvel (Figura 14).

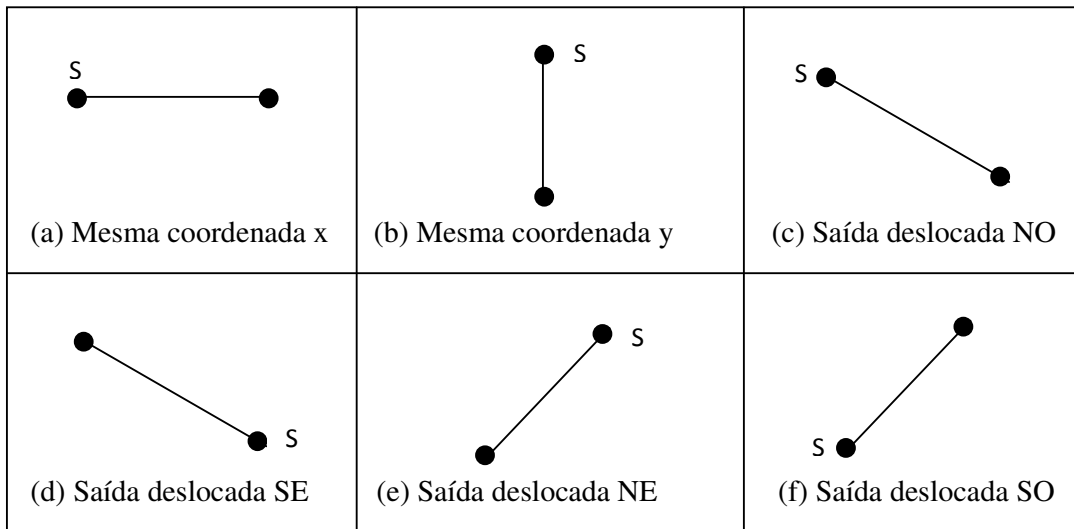


Figura 14: Relação entre locais de saída e chegada no processo de retorno da ambulância
Fonte: Elaborado pelo autor (2014)

Para o caso (a), a coordenada x se mantém a mesma e a coordenada y varia de acordo com o número de segmentos n utilizados. A hora de liberação da ambulância é calculada utilizando o tempo de deslocamento total entre a saída e chegada dividido por n :

$$h_l = \frac{\sqrt{(x_1^0 - x_2^0)^2 + (y_1^0 - y_2^0)^2}}{v_m} \times \frac{1}{n} \quad (17)$$

A mesma lógica vale para a situação em (b), com a diferença de que a coordenada y permanece a mesma. Para as situações de (c) a (f), as coordenadas de liberação em cada segmento são atualizadas da seguinte maneira:

$$x_1^i = x_1^{i-1} + \left(\frac{x_2^0 - x_1^0}{n} \right), \forall i = 1, \dots, n - 1 \quad (18)$$

$$y_1^i = y_1^{i-1} + \left(\frac{y_2^0 - y_1^0}{n} \right), \forall i = 1, \dots, n - 1 \quad (19)$$

5.1.6 TEMPOS DE DESLOCAMENTO DAS AMBULÂNCIAS

Diversos autores já apresentaram estudos e modelos para estimativa dos tempos de deslocamento de unidades móveis (Kolesar e Blum, 1973; Budge *et al.*, 2010). Entretanto, a grande maioria envolve estimativa de um número elevado de parâmetros, aumentando a incerteza relativa à análise dos resultados gerados. Hausner (1975) afirma que o tempo de deslocamento é melhor estimado como uma função linear da distância, ou seja, $t_{ij} = b_0 + b_1 D_{ij}$, onde t_{ij} é o tempo de deslocamento entre i e j , D_{ij} a distância entre i e j , b_0 e b_1 os coeficientes da regressão a serem determinados. A metodologia proposta, por seu caráter de simplicidade e maior abrangência de aplicação, segue este mesmo raciocínio.

Os tempos de deslocamento relativos às unidades móveis dizem respeito às viagens até o incidente, os hospitais, central de regulação para reposição dos equipamentos, oficina para manutenção ou retorno à base de operação. A metodologia proposta utiliza de duas informações básicas para estimativa do tempo gasto com estas atividades. A primeira é a distância euclidiana entre origem e destino e a segunda é a velocidade média desenvolvida pela ambulância. Na validação do modelo, utilizou-se da distribuição de probabilidades triangular para representar este segundo parâmetro, de maneira a melhor considerar as variações da velocidade decorrentes de fatores como, por exemplo, intensidade do tráfego.

Considere que a origem tenha coordenadas UTM (x_A, y_A) , o destino tenha coordenadas (x_B, y_B) , os parâmetros da distribuição triangular representativa da velocidade média no intervalo de tempo $[t_1, t_2]$ sejam dados por a (mínimo), b (moda) e c (máximo) e o fator de correção da distância euclidiana relativo à região s seja F_s . O tempo de deslocamento é dado por:

$$t_d = \frac{\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}}{f(v_t | a, b, c)} \times F_s \quad (20)$$

O uso da distância euclidiana, na maioria das aplicações reais de transportes, constitui-se em uma abstração útil para os cálculos e estruturação dos modelos. A razão está 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). Como o objetivo do trabalho é desenvolver uma metodologia que possa ser replicada sem maiores dificuldades independente do contexto, tal abordagem facilita este processo por não fazer uso de sistemas de informação geográfica (Derekenaris *et al.*, 2001; Huang e Pan, 2007).

A outra informação, a velocidade média das unidades, sofre influência de diversos fatores, como por exemplo, o horário do dia relacionado com a intensidade do tráfego, condições das vias de trânsito, períodos de chuva, etc. Espera-se que as unidades móveis possam desenvolver velocidades maiores nos períodos de madrugada, por exemplo. A metodologia busca considerar o funcionamento do sistema em condições normais de operação, ou seja, sofrendo influência apenas da intensidade do tráfego ao longo do dia. A ideia proposta é desenvolver uma estrutura que possa considerar estas variações ao longo do dia (Pinto *et al.*, 2015).

Observando a distribuição das chamadas no caso do SAMU de Belo Horizonte (Figura 7), sugere-se o uso de quatro faixas de horário (manhã, tarde, noite e madrugada) para definição das velocidades médias, que considerem os horários de pico e os horários de menor intensidade do tráfego.

Além disso, existe um critério base para definição de limites de velocidade em cada tipo de via. De acordo com o Código Brasileiro de Trânsito, o trânsito é composto por vias de circulação e essas por sua vez são divididas em urbanas e rurais, em especial pela presença e ausência, respectivamente, de edificações em sua extensão.

As vias urbanas são divididas em quatro subgrupos com os seguintes limites de velocidade padrão:

- 80 km/h – Via de Trânsito Rápido – vias com diversas faixas, sem semáforos, sem trânsito de pedestres e com grande extensão.

- 60 km/h – Via Arterial – avenidas com semáforos, cruzamentos e grande fluxo de trânsito, que ligam regiões de uma cidade.
- 40 km/h – Via Coletora – ruas que permitem o acesso e saída das vias arteriais, normalmente com semáforos e que permitem a circulação dentro de uma região da cidade.
- 30 km/h – Via Local – ruas de pequeno porte, com cruzamentos sem semáforo, pouco fluxo de trânsito e utilizadas normalmente para circulação local.

Para definir os limites de velocidade de uma via, as autoridades competentes a classificam de acordo com os tipos descritos acima, depois consultam o artigo 62 do Código Brasileiro de Trânsito, que determina as velocidades padrão por tipo de via, e então ajustam esta velocidade a dados particulares da via, como as condições naturais, fluxo e características de trânsito que por ela circulam.

O tempo de viagem será computado sempre que uma unidade de atendimento se deslocar para o local do incidente, para sua base, para a central de regulação, para o hospital ou para a oficina de manutenção. Hospitais possuem uma capacidade definida de atendimento para receber cada tipo de emergência. Caso esteja em ocupação máxima, a ambulância será direcionada para outra unidade. Em sistemas integrados, esta verificação é feita pelo médico regulador na central, evitando desperdícios de tempo. O redirecionamento de ambulâncias causa impacto no tempo de transporte (Pham *et al.*, 2006), tendo sido estudado por diversos autores (Ramirez-Nafarrate *et al.*, 2014; Ramirez-Nafarrate e Hafizoglu, 2012) utilizando das técnicas de simulação. Geralmente modelos representativos de sistemas móveis de atendimento de emergência não levam em consideração este fenômeno, sendo mais pertinente para modelos de departamentos de emergência em hospitais (Lin *et al.*, 2012) objetivando analisar o tempo de espera das admissões.

5.1.7 TEMPOS DAS ATIVIDADES

A coleta de dados relativa aos tempos das diversas atividades relativas ao atendimento a uma chamada de emergência deve levar em consideração as possíveis influências do dia, horário e características da ocorrência. Como exemplo, pode-se citar que o tempo que o médico regulador gasta na central de regulação para chamadas básicas, avançadas

ou simplesmente orientação por telefone é bastante diferente, devendo ser discretizado para cada caso. A Tabela 2 apresenta uma sugestão de padrões de coleta para cada atividade componente do sistema. As atividades de deslocamento e chegada das chamadas já foram discutidas anteriormente.

Tabela 2: Coleta de dados para estimar tempo de duração de cada atividade

Atividades	Divisão Sugerida
Teledigfonista	chamadas médicas e não médicas
Médico regulador	chamadas básicas, avançadas e orientação por telefone
Despachante	estimativa única
Atendimento local	chamadas avançadas e básicas
Reposição	estimativa única
Manutenção	estimativa única
Entrega do paciente	chamadas avançadas e básicas

Fonte: Elaborado pelo autor (2014)

A coleta e análise dos dados foram feitas de acordo com a metodologia proposta (seção 4.1) e serão apresentadas com maiores detalhes na seção 5.1.10 de validação do componente de simulação, utilizando do SAMU de Belo Horizonte.

Para o modelo de simulação, é necessário ainda definir o percentual de chamadas que necessita enviar a vítima para o hospital, o padrão de reposição para cada tipo de ambulância e o percentual de chamadas que são trotes (a ambulância desloca até o local do incidente, mas não realiza atendimento, ficando imediatamente disponível para atender um novo chamado).

5.1.8 PARÂMETROS DA SIMULAÇÃO

Os sistemas de atendimento de emergência podem ser caracterizados como sistemas não terminais, ou seja, não se sabe à priori sobre o momento em que a simulação deve ser terminada.

Para determinação do período destas duas fases, a metodologia proposta sugere a utilização do tempo de resposta médio, por ser este o parâmetro utilizado pelos gestores para avaliação do desempenho deste tipo de sistema. Para tanto, o modelo deve ser executado utilizando diferentes períodos de simulação, armazenando para cada rodada o

tempo utilizado e o valor do tempo de resposta obtido. Seja t_0 o primeiro período de tempo utilizado, e $t_0 + \Delta$ os períodos subsequentes até $t_0 + n$. Considere também para cada simulação feita, o parâmetro de resposta associado p_{t_0} até p_{t_0+n} . Ao final da análise, tem-se um conjunto de pares ordenados do tipo $(t_{0+i}, p_{t_{0+i}})$. A determinação do período transiente pode ser feita por análise gráfica. Através da observação do comportamento do sistema, determina-se o tempo de aquecimento (período inicial durante o qual não são coletadas estatísticas) de acordo com o instante em que a curva do gráfico não aparenta mais alguma tendência clara de crescimento ou diminuição, mas apenas oscila entre valores limites máximos e mínimos. Sugere-se utilizar o valor de $\Delta = 1$ hora e o valor de n determinado de acordo com os experimentos feitos.

5.1.9 ESTRATÉGIAS DE VALIDAÇÃO

Para o sistema em estudo, a metodologia proposta sugere uma combinação entre abordagens quantitativas e qualitativas para o processo de validação do módulo de simulação.

Para análise estatística comparativa entre os resultados da simulação e do sistema real, sugere-se a utilização do teste *Intervalo de Confiança para a Diferença das Médias com Amostras Pareadas* (Freitas Filho, 2008) sobre as variáveis de interesse (Law, 2007). O teste consiste na construção do intervalo de confiança para as diferenças dos valores médios da variável de resposta no sistema real (X_1) e no modelo de simulação (X_2) e na análise dos valores limites deste intervalo. De acordo com Chwif e Medina (2010), considerando um intervalo de confiança $[\theta_1, \theta_2]$ para a média das diferenças, temos as seguintes situações:

- $\theta_1 < 0$ e $\theta_2 > 0 \Rightarrow$ não é possível tirar nenhuma conclusão sobre a diferença entre as médias das alternativas;
- $\theta_1 > 0$ e $\theta_2 > 0 \Rightarrow$ a média da alternativa 1 (X_1) é a maior;
- $\theta_1 < 0$ e $\theta_2 < 0 \Rightarrow$ a média da alternativa 2 (X_2) é maior.

Os valores de X_1 e X_2 podem ser relativos ao tempo de resposta do sistema, a taxa de utilização das unidades, ao número de atendimentos em um determinado período de tempo, dentre outros, dependendo da disponibilidade de dados ou facilidade de coleta dos mesmos.

Além disso, a visão dos operadores do sistema (médicos e atendentes) sobre os parâmetros de desempenho gerados pelo modelo de simulação é de extrema importância. A distribuição do tempo de resposta (via histograma) pode ser utilizada, por exemplo, para identificar uma possível concentração de valores muito pequenos ou grandes que não estejam de acordo com a realidade.

5.1.10 VALIDAÇÃO DA METODOLOGIA DE SIMULAÇÃO

A metodologia para simulação descrita nas seções anteriores foi aplicada em um serviço de atendimento móvel de emergência existente de maneira a validar as etapas apresentadas. O serviço escolhido foi o Serviço de Atendimento Móvel de Urgência da cidade de Belo Horizonte, pela facilidade no acesso e coleta dos dados. Considerando que o autor desenvolveu um extenso trabalho (Silva, 2010) de coleta e análise dos dados do referido serviço para o ano de 2009, optou-se por validar a metodologia utilizando destes dados. As informações apresentadas a seguir referem-se à configuração do sistema para o ano de 2009. Na seção 5.4 com os testes computacionais será considerada a configuração 2011 do sistema, com o modelo de simulação já validado.

O modelo computacional foi implementado no *software* Arena 9.00 da *Rockwell Automation Technologies*. De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE), a cidade de Belo Horizonte possui 2.375.444 habitantes (2009) espalhados em nove regionais administrativas (Venda Nova, Norte, Pampulha, Nordeste, Noroeste, Leste, Centro-Sul, Oeste e Barreiro). O SAMU na cidade trabalha 24 horas por dia com 20 bases de operação, todas com um número fixo de unidades (Figura 15). Ao todo são 17 ambulâncias básicas, 5 avançadas e 1 ambulância de saúde mental. O sistema trabalha em conjunto com 21 centros de atendimento (Figura 12), divididos em Unidades de Pronto Atendimento (atende ocorrências de caráter clínico de média gravidade), Hospitais (ocorrências de caráter clínico de alta gravidade, trauma e gineco-obstétrico) e Centros de Referência em Saúde Mental (ocorrências de caráter psiquiátrico).

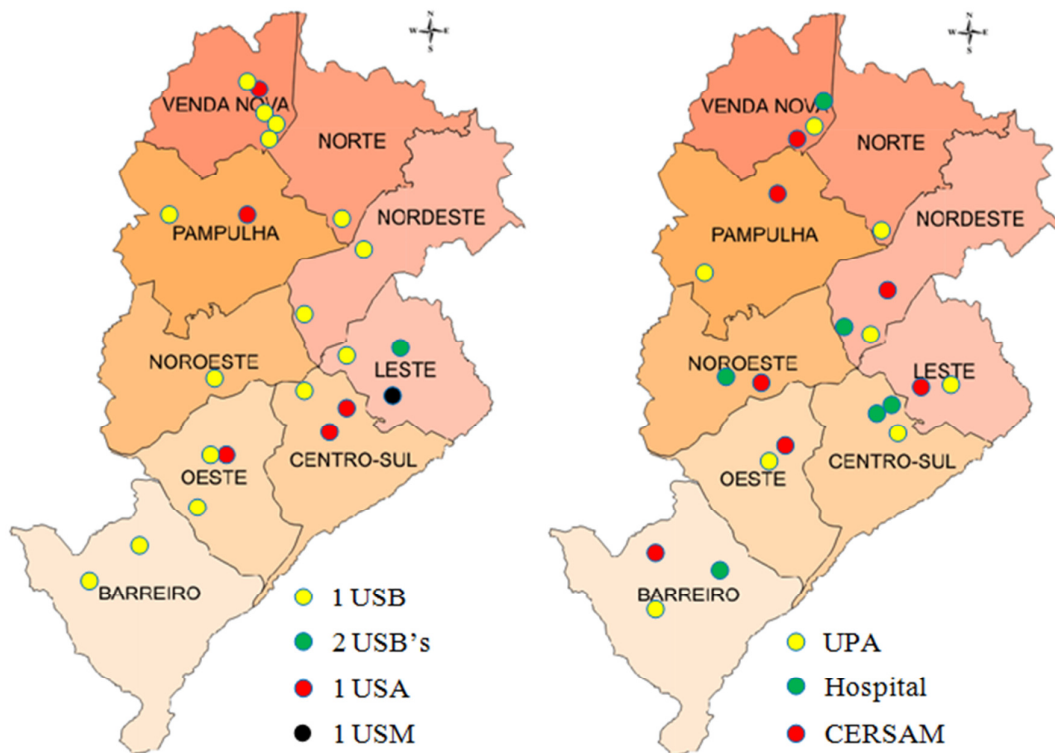


Figura 15: Distribuição espacial das ambulâncias (esquerda) e centros de atendimento (direita) na cidade de Belo Horizonte para o ano de 2009
 Fonte: Elaborado pelo autor (2014)

A central de regulação (regional Centro Sul) opera com 6 teledigifonistas (atendentes que fazem o primeiro contato com o solicitante), 5 médicos reguladores (decidem sobre o envio ou não de ambulância) 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).

Para levantamento e estimativa dos *inputs* do modelo, utilizou-se o banco de dados do SAMU-BH com informações de outubro de 2006 até dezembro de 2009. Cada chamada para a central possui um registro específico com todas as suas características, independente de ter gerado um acionamento de ambulância ou simplesmente uma orientação por telefone. A Figura 16 apresenta a evolução do número de chamadas ao longo do período analisado no banco de dados. Pode-se perceber que existe uma tendência não linear de crescimento ao longo dos anos. Vale ressaltar que o mês de agosto de 2011 não possui todos os dados completos, por isso a queda no gráfico em relação ao mês anterior.

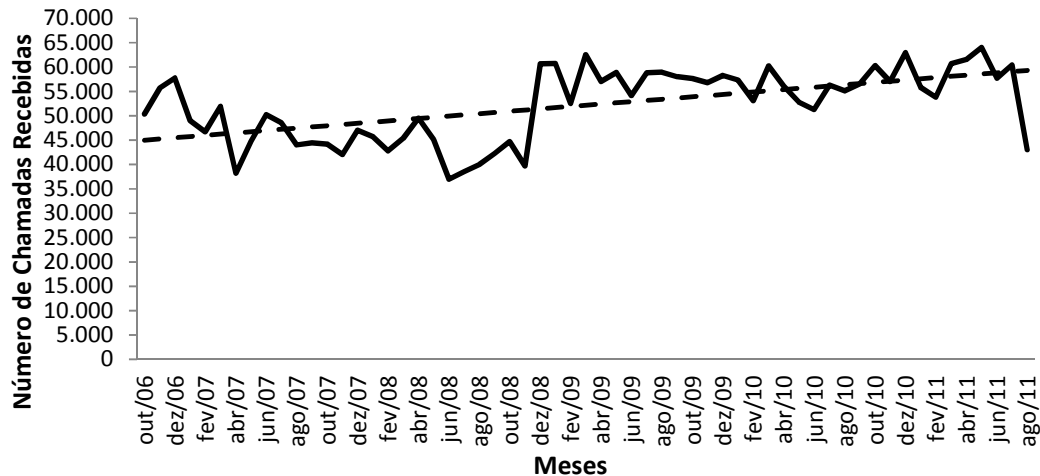


Figura 16: Evolução do número de chamadas (out/06 - ago/11)

Fonte: Adaptado de Prodabel (2011)

Para coletar os dados de acordo com a metodologia apresentada, foi extraída a quantidade média de chamadas por dia e horário considerando os meses de agosto de 2009 a novembro de 2009. Estes meses foram escolhidos para representar uma situação de operação “normal” do sistema, desconsiderando os meses de férias quando o fluxo de pessoas na cidade sofre uma alteração. O resultado da análise está apresentado na Tabela 3, sendo estes os dados utilizados no modelo de simulação.

Pela Figura 17, pode-se perceber que existe um comportamento semanal para a taxa de chegada. Em dias úteis, existem dois picos nos horários de 12:00 horas e 18:00 horas, sendo que aos finais de semana a quantidade de chamadas no período da madrugada ultrapassa a mesma quantidade em dias úteis. A figura demonstra a importância da análise discretizada do processo de chegada de chamadas, de maneira a não subestimar horários de pico e os padrões de comportamento diferentes para dias úteis e finais de semana.

Tabela 3: Quantidade média de chamadas por dia e horário (ago/09 - nov/09)

Horário / Dias	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
00:00 - 00:59	41,45	32,21	30,73	38,87	38,95	50,51	61,31
01:00 - 01:59	32,85	23,97	24,69	26,65	24,11	37,28	52,75
02:00 - 02:59	22,08	17,83	18,62	18,99	23,17	33,64	40,67
03:00 - 03:59	20,85	16,27	15,32	15,94	16,64	27,98	36,95
04:00 - 04:59	18,88	13,56	10,14	14,31	13,29	23,24	32,65
05:00 - 05:59	17,30	17,92	14,14	14,76	17,37	21,97	22,49
06:00 - 06:59	36,88	39,48	39,16	40,71	41,87	30,17	28,03
07:00 - 07:59	72,96	77,86	69,67	66,76	65,30	59,28	55,12
08:00 - 08:59	92,14	83,51	81,11	77,77	79,39	72,31	65,36
09:00 - 09:59	114,67	109,94	97,04	94,44	90,95	86,00	77,70
10:00 - 10:59	116,71	109,78	102,75	100,90	98,77	106,95	98,83
11:00 - 11:59	127,67	121,41	117,72	106,23	113,23	106,21	101,82
12:00 - 12:59	129,22	132,17	129,56	128,01	117,83	119,89	118,44
13:00 - 13:59	122,84	116,71	122,03	114,97	117,82	117,27	107,87
14:00 - 14:59	109,77	106,24	109,32	116,87	109,31	110,93	108,73
15:00 - 15:59	106,87	102,60	107,47	113,21	100,82	108,12	108,94
16:00 - 16:59	112,56	99,76	108,83	112,16	102,44	105,79	110,63
17:00 - 17:59	118,12	111,91	112,26	119,94	108,64	109,26	114,29
18:00 - 18:59	130,92	119,29	116,68	121,15	117,73	121,58	124,49
19:00 - 19:59	127,08	117,28	124,55	123,82	124,50	125,70	114,73
20:00 - 20:59	108,44	101,45	109,13	109,88	111,59	106,42	98,83
21:00 - 21:59	91,63	86,91	89,33	91,78	96,39	104,40	88,20
22:00 - 22:59	71,66	72,24	77,71	78,85	87,27	91,42	78,00
23:00 - 23:59	52,26	51,38	54,85	60,07	70,53	80,48	60,49

Fonte: Adaptado de Prodabel (2009)

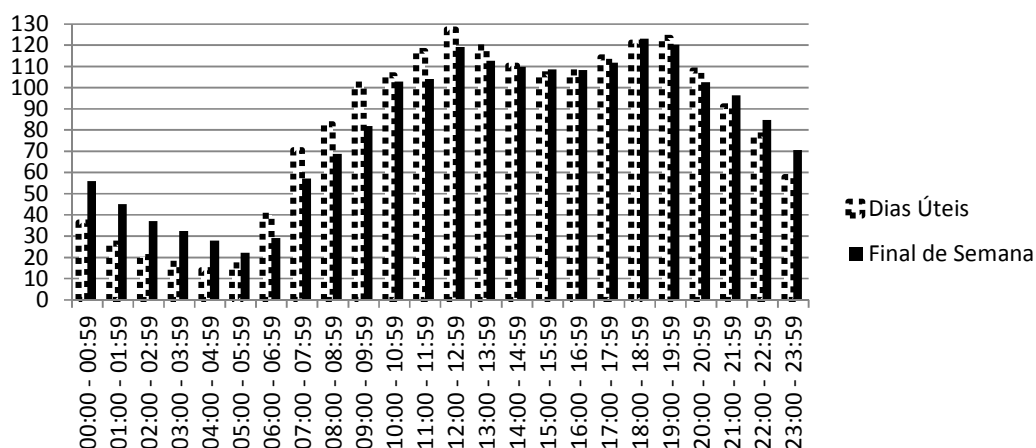


Figura 17: Número de chamadas por dia e faixa de horário (ago/09 - nov/09)

Fonte: Adaptado de Prodabel (2009)

A partir do banco de dados também foi possível identificar a distribuição das chamadas ao longo das nove regionais consideradas, conforme apresentado na Figura 18.

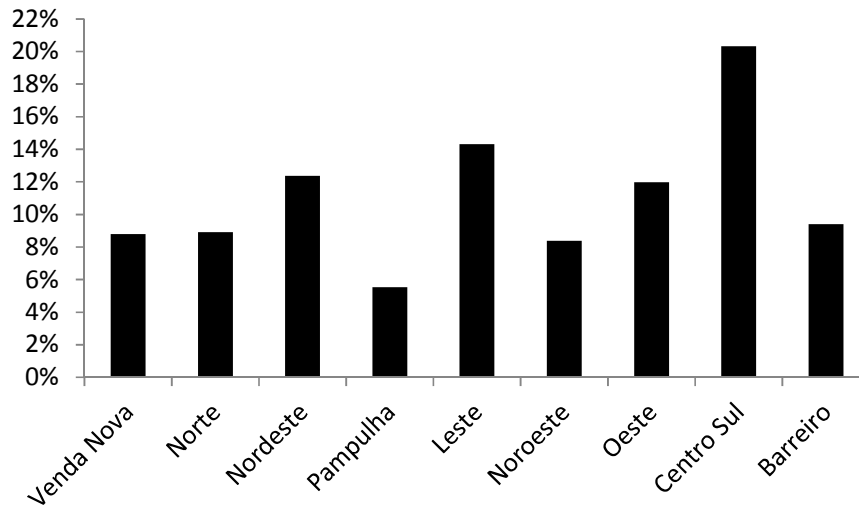


Figura 18: Distribuição das chamadas por regional (ano 2009)
Fonte: Adaptado de Prodabel (2009)

A região centro sul é a que apresenta a maior quantidade de chamadas, como previsto, uma vez que concentra um fluxo elevado de veículos (carros e ônibus) e pessoas devido à sua localização e serviços fornecidos na região. A regional com a frequência relativa de chamadas mais baixa é a Pampulha.

Para geração dos incidentes, foi feito o mapeamento de cada regional de acordo com o sistema UTM de coordenadas, buscando aproximar o contorno da região por um polígono de n lados.

Com relação às características das chamadas, foi possível extrair do banco de dados informações relativas ao motivo, tipo de ocorrência, conduta médica adotada e tipo de ambulância empenhada. Para o ano de 2009, apenas 31,93% das chamadas resultaram em atendimento pelo médico regulador (Tabela 4), sendo que um grande percentual das chamadas é decorrente de trote, ligação interrompida ou ligação solicitando alguma informação. A proporção das chamadas de engano, chamadas internas e transporte inter-hospitalar apresentaram um discreto aumento do ano de 2006 até 2011.

Tabela 4: Distribuição dos motivos de chamada (out/06 - ago/11)

Motivo da Chamada	2006	2007	2008	2009	2010	2011
Atendimento	26,47%	29,46%	32,28%	31,42%	32,22%	29,86%
Informação	13,38%	12,98%	12,94%	11,72%	13,67%	14,35%
Ligação Interrompida	25,61%	21,99%	22,74%	19,69%	17,41%	16,44%
Informação em saúde	2,30%	2,89%	2,56%	3,25%	1,91%	0,99%
Trote	21,06%	20,33%	17,84%	20,24%	18,24%	20,71%
Abuso	0,29%	0,35%	0,29%	0,33%	0,16%	0,23%
Engano	3,90%	4,01%	3,86%	3,51%	4,06%	4,25%
Chamada Interna	6,69%	7,58%	7,02%	9,32%	11,68%	12,45%
Transporte Inter-hospitalar	0,30%	0,41%	0,47%	0,51%	0,64%	0,73%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fonte: Adaptado de Prodabel (2011)

Uma vez definido o motivo da chamada, caso esta seja transferida para o médico regulador, cabe a este profissional identificar o tipo de ocorrência e sua gravidade, de maneira a escolher a unidade mais adequada para envio. A Tabela 5 apresenta a evolução do tipo de ocorrência. As chamadas de transporte intermunicipal e “outros” apresentaram um aumento discreto de 2006 para 2011, sendo que a ocorrência clínica é a responsável pela maior porcentagem em todos os anos considerados, seguida pela ocorrência de trauma, psiquiátrico e gineco-obstétrico.

Tabela 5: Distribuição dos tipos de ocorrência (out/06 - ago/11)

Tipo de Ocorrência	2006	2007	2008	2009	2010	2011
Trauma	26,04%	24,53%	23,87%	23,06%	22,65%	22,84%
Clínico	46,45%	47,80%	47,97%	47,15%	45,30%	43,34%
Psiquiátrico	6,63%	6,45%	6,57%	6,56%	6,64%	6,69%
Gineco-obstétrico	2,92%	2,82%	2,73%	2,39%	2,08%	1,93%
Transporte Intermunicipal	1,68%	1,82%	2,04%	2,18%	2,43%	2,73%
Outros	16,28%	16,56%	16,82%	18,66%	20,90%	22,47%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fonte: Adaptado de Prodabel (2011)

Cabe ao médico regulador determinar o procedimento a ser adotado para cada chamada dependendo das informações que ele coleta via telefone com o solicitante. Este procedimento é denominado de conduta médica e os percentuais de cada possibilidade estão apresentados na Tabela 6. Para todos os anos, a maior proporção é de envio de ambulância, seguido da orientação de buscar ajudar por meios próprios, orientação por telefone, outros e repasse da chamada para outras instituições. O único tipo de conduta

que apresenta um crescimento ao longo dos anos é a orientação por telefone. O repasse da chamada apresenta decréscimo.

Tabela 6: Distribuição das condutas médicas (out/06 - ago/11)

Conduta Médica	2006	2007	2008	2009	2010	2011
Orientação por telefone	13,59%	16,26%	17,00%	18,38%	21,93%	22,27%
Recursos próprios	27,58%	27,25%	28,13%	30,48%	26,88%	29,09%
Envio de ambulância	43,22%	44,54%	44,47%	39,87%	40,96%	38,35%
Outros	14,28%	11,24%	9,94%	10,60%	9,59%	9,88%
Repasse da chamada	1,33%	0,71%	0,45%	0,67%	0,63%	0,41%
Total	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Fonte: Adaptado de Prodabel (2011)

Determinado o tipo de ocorrência e identificada a necessidade de envio de uma ambulância, o próximo passo é determinar se será enviada uma unidade básica, avançada ou de saúde mental. Foram extraídos do banco de dados os percentuais de envio de cada tipo de ambulância discretizados por tipo de ocorrência, conforme apresentado na Tabela 7. Pode-se observar que a grande maioria das ocorrências é atendida por uma unidade básica, e um percentual muito pequeno por unidades de saúde mental. Uma ocorrência de caráter psiquiátrico pode ser atendida por qualquer tipo de ambulância, somente nos casos mais graves é que se utiliza a ambulância específica de saúde mental.

Tabela 7: Distribuição do tipo de ambulância despachada por tipo de ocorrência

Tipo de Ambulância	Tipo de Ocorrência - ano 2011			
	Trauma	Clínico	Psiquiátrico	Gineco
USB	91,68%	92,14%	98,90%	92,56%
USA	8,12%	7,48%	0,98%	6,99%
USM	0,20%	0,38%	0,12%	0,45%

Fonte: Adaptado de Prodabel (2011)

Também foram obtidas, a partir do banco de dados, estimativas para o número de empenhos de ambulância cancelados e a proporção de atendimentos locais que necessitam do transporte da vítima para algum centro de atendimento. Os cancelamentos, para uso no modelo computacional desenvolvido, foram definidos como a chegada da unidade de resgate no local do incidente e a constatação de que não existe nenhuma pessoa necessitando de cuidados. Desta maneira, para uma chamada

cancelada, considera-se o tempo de deslocamento, mas não o tempo de atendimento local. As estimativas foram de 5,21% para as chamadas canceladas e 90% para as ocorrências necessitando transporte da vítima para algum centro de atendimento.

Para estimativa dos tempos das diversas operações que fazem parte do SAMU-BH (desde o atendimento pelos técnicos até a entrega do paciente em um dos centros de atendimento) foram utilizados dados coletados diretamente na central de regulação, priorizando diferentes dias da semana e diferentes horários ao longo do dia. Os dados foram ajustados através do teste de aderência de Kolmogorov-Smirnov, baseado na comparação das probabilidades acumuladas das distribuições teórica e observada. Os resultados estão apresentados na Tabela 8. Para maiores informações sugere-se consulta ao trabalho de Silva (2010).

Tabela 8: Distribuições de probabilidade para cada atividade do SAMU-BH

Atividade	Distribuição	Média	Desvio Padrão	Unidade
Despacho	Gama	22,55	11,16	seg
Teledigifonista Chamada Médica	Weibull	110,94	49,01	seg
Teledigifonista Chamada Não Médica	Weibull	28,53	18,71	seg
Médico Regulador Não Envia	Weibull	102,57	70,51	seg
Médico Regulador Envia USB	Gama	66,49	26,12	seg
Médico Regulador Envia USA	Beta	96,76	60,02	seg
Atendimento Local USB	Weibull	18,41	8,38	min
Atendimento Local USA	Beta	19,32	5,61	min
Reposição	Triangular	21,67	3,12	min
Manutenção	Erlang	2,22	1,57	hr
Informações para Despachante	Triangular	55,53	10,21	seg
Regulação Secundária	Triangular	16,80	4,49	seg

Fonte: Elaborado pelo autor (2014)

Para a atividade de reposição foi considerado o padrão de 1 reposição a cada 24 horas para as unidades básicas e 2 reposições diárias para as unidades avançadas. Para a atividade de manutenção, utilizou-se uma taxa média de 3 ambulâncias por dia, sendo que a escolha da ambulância é feita pelo modelo de maneira aleatória. A atividade de regulação secundária acontece quando a equipe da ambulância já chegou ao local do incidente e transmite novas informações para o médico regulador na central, de maneira que este possa decidir sobre os procedimentos a serem realizados e qual centro de atendimento irá receber a vítima.

O fator de correção da distância euclidiana foi estimado de maneira independente para cada uma das 9 regionais da cidade de Belo Horizonte, de maneira a melhor considerar as particularidades de cada região. A estimativa foi realizada com o uso do aplicativo *Google Earth* e utilizando 50 pontos (coordenadas UTM) em cada regional, considerando localidades em diferentes proximidades e priorizando o uso de vias principais para o deslocamento.

A determinação da velocidade média considerou a intensidade do tráfego nos diferentes horários do dia. Utilizou-se a distribuição triangular para estimativa das velocidades, sendo que a construção dos parâmetros baseou-se na percepção dos motoristas das unidades de resgate. As faixas de horário utilizadas foram de 12:00PM – 6:00AM, 06:00AM – 11:00AM, 11:00AM – 07:00PM e 07:00PM – 11:00PM.

Para determinação do período transiente do sistema, utilizou-se a análise do tempo de resposta médio, por ser este o parâmetro utilizado pelos gestores para avaliação do desempenho do sistema. Para tanto, o modelo foi executado em um tempo de simulação variando de uma até trezentas horas, conforme apresentado na Figura 19, armazenando para cada rodada o valor da variável de análise no eixo vertical e o tempo de simulação no eixo horizontal. A determinação do período transiente foi feita por análise gráfica. Através da observação do comportamento do sistema, definiu-se um tempo de aquecimento do modelo de 160 horas, pois a partir deste instante a curva não aparenta mais alguma tendência clara de crescimento ou diminuição, mas apenas oscila entre os valores limites máximos e mínimos.

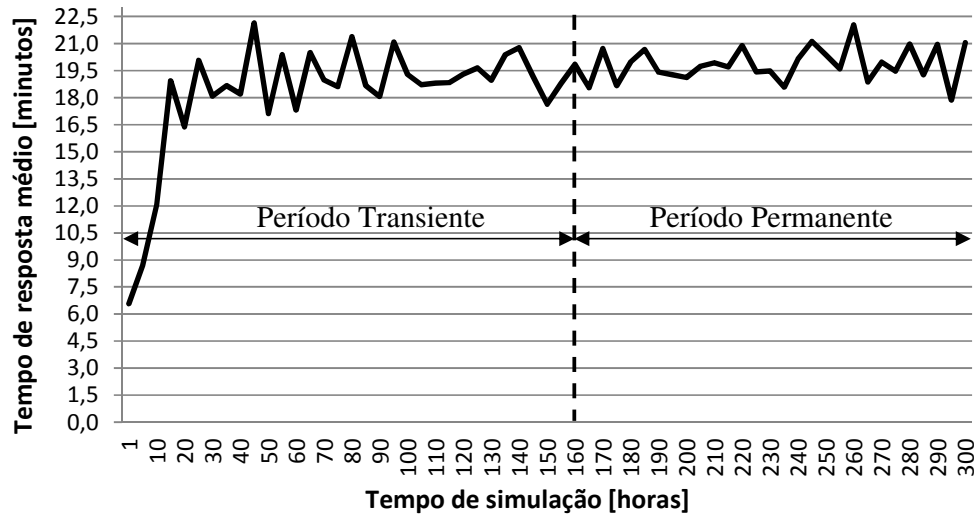


Figura 19: Período transiente e permanente para o tempo de resposta médio
 Fonte: Elaborado pelo autor (2014)

O SAMU-BH foi simulado de acordo com a metodologia proposta e dados apresentados anteriormente. Para validação da metodologia, além do acompanhamento direto da gerência do SAMU nas etapas de modelagem e coleta de dados, utilizou-se outras duas informações: o tempo de resposta do sistema e o total de atendimentos realizados em cada mês. O tempo de resposta foi utilizado por ser o parâmetro de desempenho mais importante na análise deste tipo de sistema. O total de atendimentos foi utilizado para aumentar a confiança do modelo buscando levar em consideração as atividades dentro e fora da central de regulação. Para o tempo de resposta, construiu-se um intervalo de confiança de 95% com a média de 10 replicações, conforme apresentado a seguir:

$$I. C_{TR} = [18,25 ; 20,14] \text{ minutos}$$

A validação do número de replicações deu-se pela análise da precisão do intervalo de confiança construído. Para o ano de 2009, chegou-se a uma precisão de 0,95, considerada aceitável pelos gestores do sistema e adequada considerando os objetivos do trabalho.

Considerando que os dados coletados na central de regulação para estimativa dos tempos das diversas atividades, foi possível construir uma estimativa para o tempo de resposta médio real do sistema, cujo valor foi de 19,92 minutos. Este valor possui uma diferença percentual para o ponto médio do intervalo de confiança construído de 3,6%.

Pode-se observar que os valores real e simulado estão bastante próximos. Com relação ao total de atendimentos realizados, a Tabela 9 apresenta os resultados encontrados, sendo que a coluna “Real” diz respeito aos valores retirados diretamente do banco de dados e a coluna “Simulado” a média das dez replicações.

Tabela 9: Validação do número de atendimentos realizados (ano 2009)

Mês	Real	Simulado	Diferença (%)
jan/09	7.266	7.296	-0,41%
fev/09	6.752	6.731	0,31%
mar/09	7.503	7.469	0,45%
abr/09	7.281	7.207	1,02%
mai/09	7.528	7.438	1,20%
jun/09	6.934	7.005	-1,02%
jul/09	7.417	7.484	-0,90%
ago/09	7.429	7.240	2,54%
set/09	7.175	7.099	1,06%
out/09	7.298	7.328	-0,41%
nov/09	7.308	7.129	2,45%
dez/09	7.590	7.369	2,91%
Média	7.290	7.233	1,22%

Fonte: Elaborado pelo autor (2014)

Podemos observar que a diferença média em valores absolutos entre o real e o simulado é de apenas 1,22%.

Além dessas informações, foi possível construir um histograma de frequência com as variáveis de saída da simulação (Figura 20) para a distribuição dos tempos de resposta, para verificar se o comportamento está de acordo com a percepção dos operadores do sistema. Pode-se observar que a frequência relativa dos valores extremos (tempos de resposta menores que 5 minutos ou maiores que 40 minutos) é bastante reduzida. Aproximadamente 42% de todos os atendimentos possuem tempo de resposta entre 10 e 20 minutos, fato este confirmado pela experiência e estimativa empírica dos operadores do sistema.

Com base nas análises feitas, na percepção do especialista e gestores do sistema sobre as respostas do modelo de simulação desenvolvido e nos objetivos definidos para o estudo (observar o comportamento de serviços de atendimento de emergência com

relação ao tempo de resposta), concluiu-se pela validação efetiva da metodologia proposta.

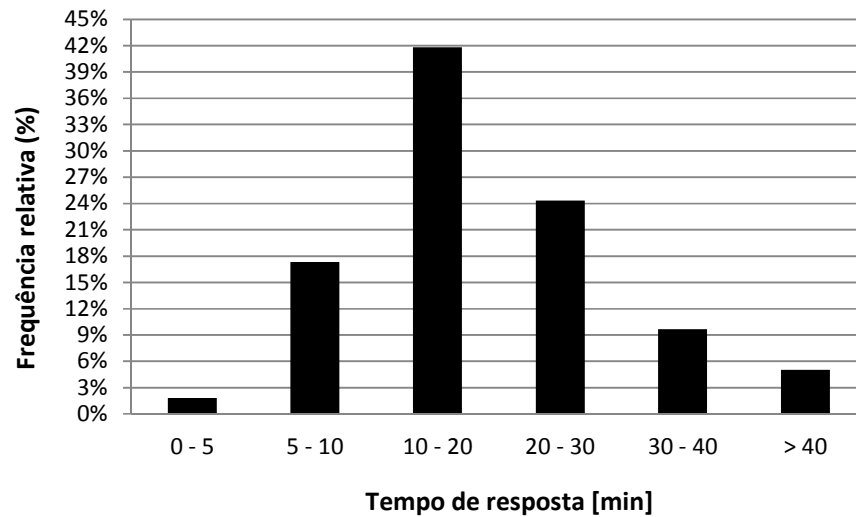


Figura 20: Histograma da distribuição dos tempos de resposta (SAMU-BH 2009)
Fonte: Elaborado pelo autor (2014)

Com relação à taxa de utilização das ambulâncias, das 10 replicações realizadas foi possível obter uma média de 53.37% para as unidades básicas e 12.15% 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. O trabalho de Garcia (2006) apresenta uma taxa de utilização de 49.2% para as ambulâncias do serviço de atendimento da cidade de Niterói. O valor mais baixo encontrado para as unidades avançadas se refere à quantidade reduzida de chamadas de maior gravidade e na política de priorização dos médicos reguladores sobre as unidades básicas. Os valores encontrados para o SAMU de BH estão de acordo com a percepção dos operadores.

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.

Aplicação da metodologia SAMU-BH (2011)

Uma vez validada a metodologia, a mesma foi aplicada considerando a configuração do sistema de atendimento móvel de urgência de Belo Horizonte para o ano de 2011. A definição do ano se deu pela dificuldade no acesso às informações do sistema, não só na coleta em campo mas também aquelas armazenadas no banco de dados. As proporções relativas a cada característica das chamadas (motivo, tipo de ocorrência, conduta médica e tipo de ambulância) são aquelas apresentadas nas tabelas anteriores, mas considerando o ano de 2011. Os tempos das diversas operações e demais parâmetros continuaram os mesmos, com exceção da distribuição das ambulâncias ao longo do território considerado. A nova distribuição está apresentada na Figura 21.

Os centros de atendimento também não sofreram alteração. Pode-se observar que houve mudança na quantidade de ambulâncias em algumas regionais: na Centro-Sul houve um acréscimo de 2 unidades básicas e diminuição de 1 unidade avançada, nas regionais Leste, Nordeste e Venda Nova houve diminuição de 1 unidade básica, na regional Noroeste houve acréscimo de 2 unidades básicas e na regional Oeste houve acréscimo de 1 unidade básica. As demais mantiveram com a mesma configuração. Quando comparado com o ano de 2009, pode-se observar que houve uma melhor distribuição das unidades ao longo da cidade. O total de ambulâncias básicas se manteve em 17, 1 ambulância de saúde mental e as ambulâncias avançadas aumentaram para 6 (em 2009 eram 5 unidades).

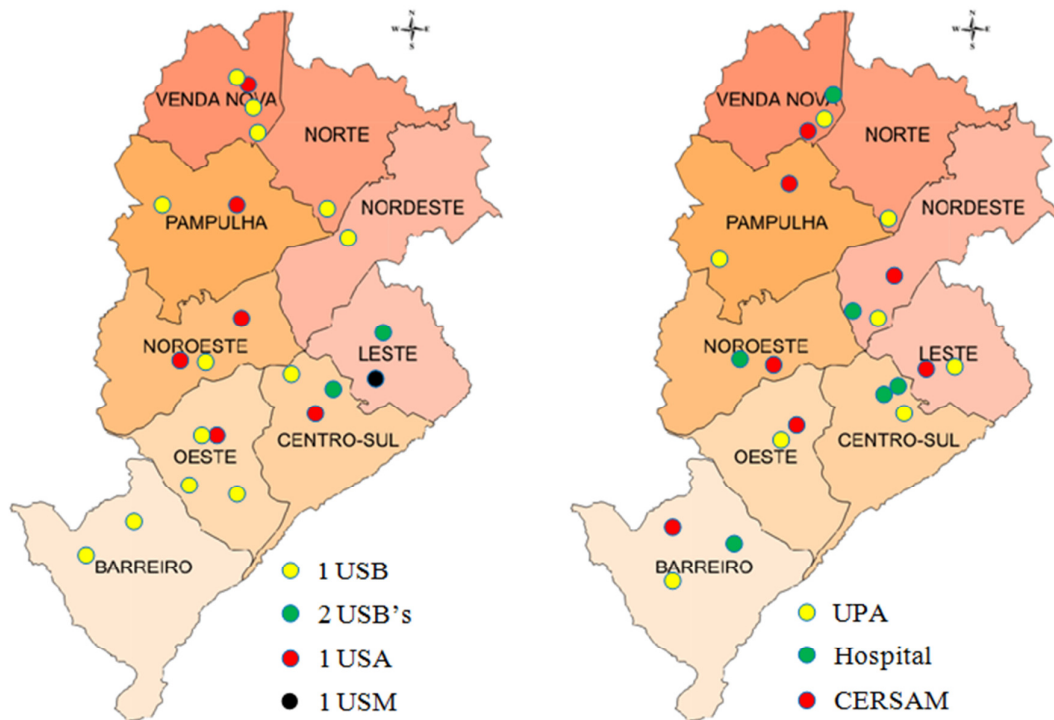


Figura 21: Distribuição espacial das ambulâncias (esquerda) e centros de atendimento (direita) na cidade de Belo Horizonte para o ano de 2011
 Fonte: Elaborado pelo autor (2014)

Outra informação que foi atualizada no modelo diz respeito ao padrão de chegada das chamadas, que sofreu um leve aumento conforme apresentado na Tabela 10.

Tabela 10: Quantidade média de chamadas por dia e horário (mar/11 - jun/11)

Horários / Dias	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
00:00 - 00:59	43,06	32,89	32,11	34,50	36,24	46,71	55,35
01:00 - 01:59	28,53	22,17	21,83	21,72	26,59	38,00	44,94
02:00 - 02:59	20,59	16,50	14,94	16,67	16,29	28,82	38,71
03:00 - 03:59	19,94	13,39	13,33	16,72	15,29	24,12	30,88
04:00 - 04:59	15,06	11,28	9,94	15,61	12,47	22,18	28,47
05:00 - 05:59	14,88	13,78	11,39	15,44	14,41	22,18	28,88
06:00 - 06:59	43,53	35,61	35,17	35,72	40,06	33,47	30,59
07:00 - 07:59	71,65	68,50	72,78	66,00	71,12	59,71	48,47
08:00 - 08:59	90,41	80,33	76,22	73,11	82,88	78,18	57,18
09:00 - 09:59	95,41	89,78	83,17	84,11	87,41	82,00	69,53
10:00 - 10:59	114,53	113,00	104,50	105,61	108,47	104,82	93,82
11:00 - 11:59	124,41	131,33	117,83	119,11	114,88	118,82	101,12
12:00 - 12:59	148,53	154,39	141,67	146,67	150,41	142,12	134,41
13:00 - 13:59	130,24	127,83	133,67	137,17	139,59	138,88	120,29
14:00 - 14:59	131,53	122,33	128,33	124,83	133,47	129,76	118,53
15:00 - 15:59	129,29	125,11	132,06	124,17	117,94	122,88	119,82
16:00 - 16:59	124,35	113,56	112,56	118,28	130,47	112,24	126,65
17:00 - 17:59	129,59	122,06	131,72	125,28	134,35	117,47	122,24
18:00 - 18:59	119,76	114,00	111,72	121,67	117,18	111,65	112,47
19:00 - 19:59	119,12	120,61	108,11	124,22	119,65	118,35	123,76
20:00 - 20:59	97,94	96,33	98,44	108,11	107,18	99,71	106,88
21:00 - 21:59	80,18	81,67	87,22	86,06	91,12	93,94	92,76
22:00 - 22:59	71,94	71,78	74,17	72,00	77,71	90,65	79,06
23:00 - 23:59	49,71	54,78	50,00	57,94	72,06	80,88	55,76

Fonte: Adaptado de Prodabel (2011)

Por fim, atualizou-se também a distribuição das chamadas em cada regional, conforme apresentado na Figura 22.

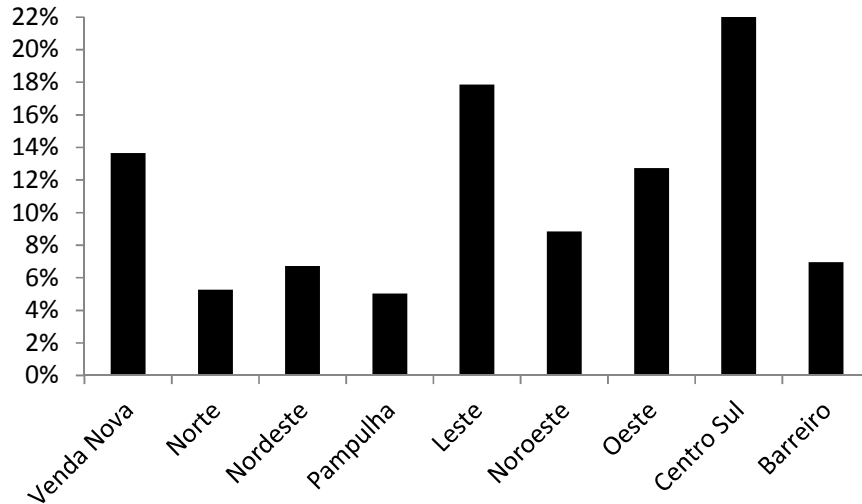


Figura 22: Distribuição das chamadas por regional (ano 2011)
 Fonte: Adaptado de Prodabel (2011)

Observa-se um crescimento no percentual de participação das regionais Venda Nova, Leste, Noroeste, Oeste e Centro Sul; e uma diminuição nas demais.

Para o tempo de resposta, construiu-se novamente um intervalo de confiança de 95% com a média das 10 replicações, conforme apresentado a seguir:

$$I.C._{TR} = [19,33 ; 22,11] \text{ minutos}$$

Para comparação com o cenário anterior (SAMU-BH 2009), construiu-se um intervalo de confiança $[\theta_1, \theta_2]$ (95%) para as diferenças das médias em cada replicação ($y_i - y_j$) com $i \in$ cenário 2009 e $j \in$ cenário 2011 ($\forall(i, j) \in \{1, 2, \dots, 10\}$). Os valores encontrados foram $\theta_1 = -2,780$ e $\theta_2 = -0,269$; $I.C. = [-2,780 ; -0,269]$ minutos. Como $\theta_2 < 0$, podemos afirmar com 95% de confiança que o $TR_{2011} > TR_{2009}$. Considerando os pontos médios dos dois intervalos, temos uma diferença percentual de 7,94%. Como o número de unidades básicas se manteve constante e a taxa de chegada das chamadas sofreu um pequeno aumento, era de se esperar que o tempo de resposta encontrado para o cenário 2011 fosse realmente maior do que aquele encontrado para o cenário 2009. A Figura 23 apresenta o comportamento do tempo de resposta para diferentes faixas. As barras em preto representam o comportamento para o ano de 2009 e as barras sem preenchimento para o ano de 2011.

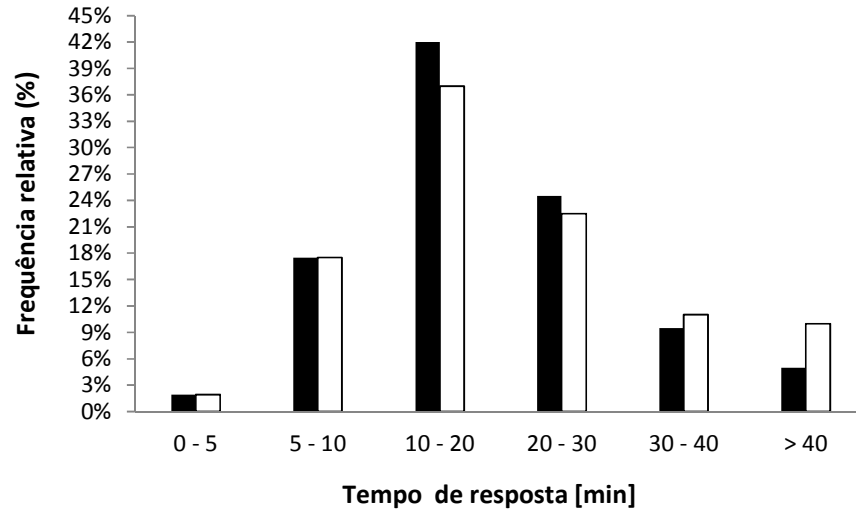


Figura 23: Histograma da distribuição dos tempos de resposta (SAMU-BH 2011)
 Fonte: Elaborado pelo autor (2014)

Pode-se perceber que a frequência relativa nas faixas entre 10-20 minutos e 20-30 minutos diminuiu quando comparado com o ano de 2009 e aumentou nas faixas entre 30-40 e > 40 minutos. Este cenário fortalece a motivação no uso de técnicas de otimização e simulação para dimensionamento de sistemas de emergência buscando adequar os parâmetros de desempenho com a demanda da população.

5.1.11 ANÁLISE DE SENSIBILIDADE COM O MODELO DE SIMULAÇÃO

Uma das premissas adotadas na construção do modelo de simulação foi a importância da consideração de atividades até então desconsideradas em trabalhos semelhantes para estimação do tempo de resposta do sistema; como a manutenção das unidades de resgate, o processo de reposição de equipamentos e os tempos relativos às atividades da central de regulação. Para validar tal hipótese, o modelo de simulação com a configuração do SAMU de Belo Horizonte no ano de 2011 foi rodado sem as três atividades citadas anteriormente (central, manutenção e reposição). Utilizou-se o mesmo número de replicações, tempo de aquecimento e nível de confiança para construção do intervalo relativo ao tempo de resposta do sistema: $I.C. = [16,28 ; 18,25]$ minutos.

Para comparação com o cenário 2011, construiu-se novamente um intervalo de confiança $[\theta_1, \theta_2]$ (95%) para as diferenças das médias em cada replicação ($y_i - y_j$)

com $i \in$ cenário com manutenção, reposição e central e $j \in$ cenário sem estas atividades ($\forall(i,j) \in \{1,2, \dots, 10\}$). Os valores encontrados foram $\theta_1 = 1,76$ e $\theta_2 = 5,15$: $I.C. = [1,76 ; 5,15]$ minutos. Como $\theta_1 > 0$, pode-se afirmar com 95% de confiança que o tempo de resposta do sistema sem as três atividades citadas é menor que o tempo de resposta do sistema com elas incluídas. Considerando os pontos médios dos dois intervalos, tem-se uma diferença percentual de 16,7%. O tempo de resposta médio do sistema de 20,72 minutos teria sido estimado em 17,25 minutos sem a consideração destas atividades.

Outra análise de sensibilidade foi conduzida com o modelo de simulação de maneira a verificar o comportamento do sistema com o aumento na demanda pelos serviços, aqui representado pelo aumento da taxa de chegada das chamadas na central. Sabe-se que a demanda está diretamente relacionada com uma série de fatores: tamanho da população, conhecimento e acesso da população aos serviços de atendimento móveis de urgência, abrangência deste tipo de serviço, políticas de redução de acidentes (por exemplo, campanhas de educação no trânsito), dentre outros. A Figura 24 apresenta, na abscissa o aumento percentual nas taxas de chegada de chamadas na central de regulação em cada faixa de horário em cada dia, na ordenada à esquerda o tempo de resposta médio do sistema em minutos e à direita a variação percentual com relação ao tempo de resposta estimado para o ano de 2011.

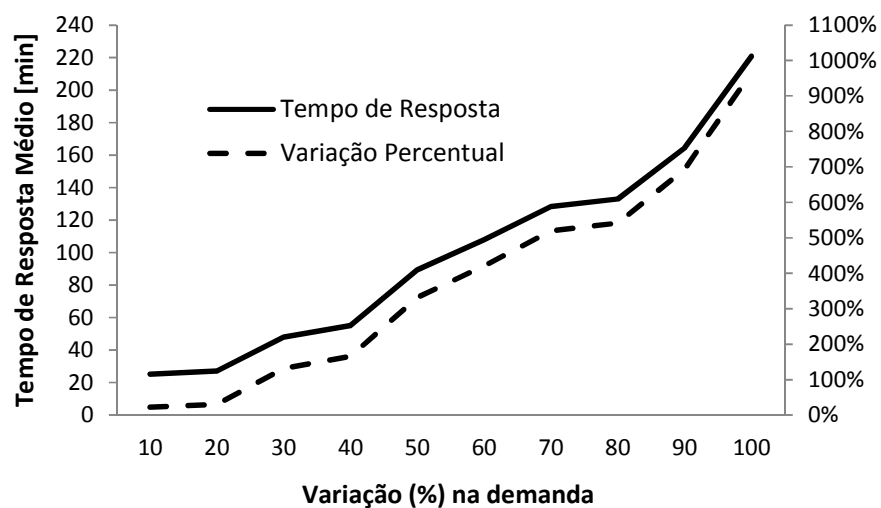


Figura 24: Tempo de resposta médio x aumento da demanda
Fonte: Elaborado pelo autor (2014)

O tempo de resposta médio sofre um 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 21 minutos, sendo que para 40% de aumento este valor mais do que dobra quando comparado ao cenário atual. Acima de 50% de aumento, o tempo de resposta médio ultrapassa a casa de uma hora, sendo que quando se dobra a demanda o valor encontrado fica próximo das quatro horas.

A variável de resposta analisada está diretamente relacionada com a capacidade de atendimento dentro da central de regulação bem como a relação entre quantidade de ocorrências e número de ambulâncias disponíveis. Mantendo este último constante e aumentando o primeiro, é evidente que se deve esperar um aumento no tempo de resposta médio do sistema como um todo. Estabelecidos os parâmetros de desempenho a serem atingidos, torna-se possível propor e simular cenários diferentes de operação de maneira a adaptar o funcionamento do sistema às novas exigências de demanda.

Quando se pensa em gerenciamento de recursos de um sistema de atendimento médico emergencial não se pode esquecer que melhor qualidade em serviço, em geral, implica em maiores custos e investimentos. Independente da forma de financiamento da empresa que gerencia o sistema de atendimento (público, privado ou cooperativo), é desejável que os recursos (não apenas financeiros) sejam alocados da melhor forma possível (MENDONÇA e MORABITO, 2000).

O cenário analisado deixa clara a importância de uma análise de *trade-off* entre possíveis investimentos no sistema e o comportamento deste frente a um aumento na demanda por serviços, sendo analisado por parâmetros de desempenho conhecidos.

Algumas medidas de desempenho podem ser conflitantes em termos dos diferentes interesses das partes envolvidas na operação dos sistemas de atendimento emergencial. O tempo médio de resposta numa região, por exemplo, é uma medida de desempenho externa do sistema, que interessa principalmente ao usuário. Por outro lado, o balanço das cargas de trabalho dos servidores é uma medida de desempenho interna do sistema, que interessa particularmente aos operadores do sistema (IANNONI E MORABITO, 2008).

Uma segunda análise foi feita variando a velocidade média das unidades de resgate. Na Figura 25, o eixo x representa a variação percentual nos parâmetros da distribuição triangular representativa da velocidade (mínimo, moda e máximo), o eixo y à esquerda o tempo de resposta médio do sistema em minutos e à direita a variação percentual com relação ao tempo de resposta estimado para o ano de 2011.

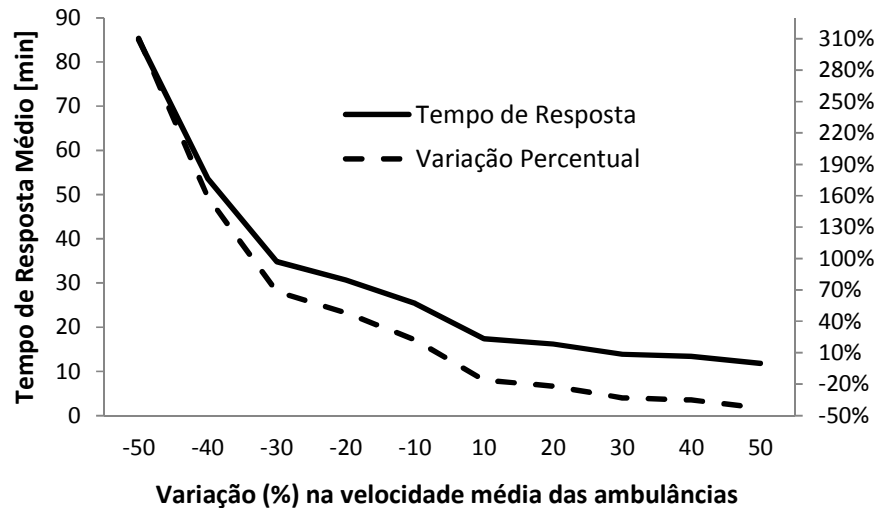


Figura 25: Tempo de resposta médio x velocidade das ambulâncias

Fonte: Elaborado pelo autor (2014)

Para uma diminuição pela metade dos parâmetros representativos da velocidade, o tempo médio de resposta do sistema se aproxima de uma hora e meia, apresentando uma inclinação negativa relativamente constante até o percentual de -30% . Um novo patamar de variação constante se mantém até a variação de 10% na velocidade, quando o tempo de resposta se aproxima de 17,5 minutos, reduzindo 16% quando comparado com o estimado para o ano de 2011. Daí em diante a variação se reduz até o percentual de aumento de 50% com um tempo de resposta estimado em 11,5 minutos.

Fica claro com esta análise que qualquer ganho no sentido de facilitar o deslocamento das unidades ao longo das vias urbanas é de grande importância para o desempenho do sistema, aqui medido em termos do tempo de resposta médio. Algumas medidas estão aquém da gestão do sistema, mas outras podem ser implantadas no sentido de agilizar o resgate, como por exemplo o uso de veículos rápidos para um primeiro atendimento às vítimas até a chegada de uma ambulância apropriada.

A perda de chamadas e o atraso no tempo de resposta estão diretamente relacionados ao conflito entre as variáveis aleatórias da demanda por serviço e as restrições de capacidade do sistema. Dado que, devido a restrições de orçamento, os serviços de atendimento emergencial não podem ser planejados de forma a trabalhar com um número muito grande de servidores, há claramente um importante *trade-off* a ser considerado entre a qualidade de atendimento e os custos de investimento e operação nestes sistemas. Ao analisar sistemas de atendimento emergencial, os fatores probabilísticos relacionados à distribuição temporal e espacial dos servidores e chamadas devem ser considerados, dado que a operação destes sistemas é caracterizada por incertezas com relação à localização e tempo necessário para atender a um determinado chamado (IANNONI e MORABITO, 2006).

De maneira a verificar o potencial da configuração do sistema, uma análise foi conduzida buscando verificar qual o menor tempo de resposta do sistema utilizando a mesma configuração do ano de 2011 (mesmo número de ambulâncias e de bases). Utilizou-se o otimizador *Optquest* integrado com o simulador *Arena* para testar diferentes cenários de alocação das unidades nas bases existentes, com o objetivo de minimizar o tempo de resposta médio. Definiu-se como restrição uma quantidade máxima de 4 unidades em cada base. Utilizando das mesmas 17 unidades básicas e 6 unidades avançadas, a melhor configuração gerada apresentou um tempo de resposta médio do sistema de 18,49 minutos. Este valor apresenta uma diferença percentual com relação ao tempo de resposta médio de 2011 de 10,76%.

Esta análise indica que para ganhos mais expressivos na variável tempo de resposta, uma possibilidade interessante seria o reposicionamento das bases atuais, evitando aumentar exageradamente o número de ambulâncias em cada base. Um estudo de localização considerando as bases atuais e possíveis pontos para implantação de novas bases futuras pode ser interessante para validar a hipótese levantada, sendo ainda possível a determinação de configurações que reduzam o tempo de resposta médio com o mesmo efetivo de unidades móveis, mas com mudanças em suas localizações ao longo do território considerado (SILVA, 2010).

Buscando entender melhor a influência do total de unidades no desempenho geral do sistema, a próxima análise de sensibilidade consistiu em aumentar de 1 em 1 a

quantidade de ambulâncias em operação e computar o tempo de resposta correspondente. A análise foi feita separadamente para as unidades básicas (Figura 26), avançadas (Figura 27) e total de unidades (Figura 28). A configuração inicial foi a mesma do ano de 2011 (17 unidades básicas, 6 avançadas e 23 no total).

A partir de análise do gráfico da Figura 26, é possível ver que o tempo de resposta médio das unidades básicas apresenta um decrescimento constante até o total de 22 unidades. Um novo padrão de decrescimento acontece até o total de 26 unidades, quando então o tempo de resposta não apresenta tendência clara de decrescimento, estabilizando em um valor próximo de 15 minutos.

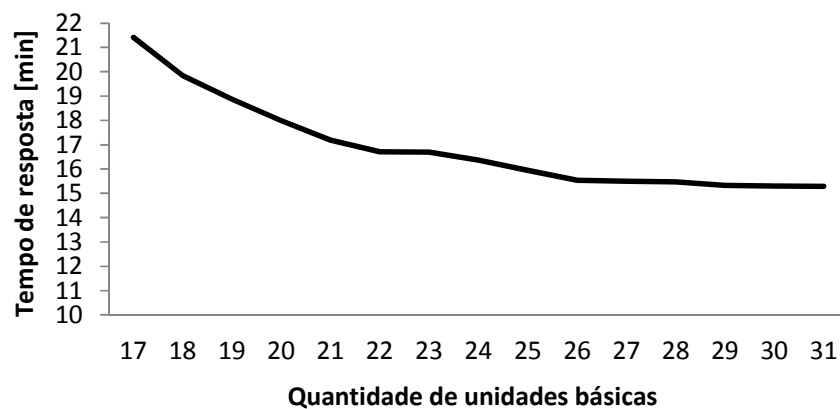


Figura 26: Análise de sensibilidade com o total de unidades básicas
Fonte: Elaborado pelo autor (2014)

Comportamento semelhante pode ser observado no gráfico da Figura 27 relativo às unidades avançadas. O decrescimento é maior na parte inicial do gráfico e vai ficando cada vez menos acentuado com o aumento no número de unidades. O menor tempo de resposta alcançado está próximo dos 12 minutos.

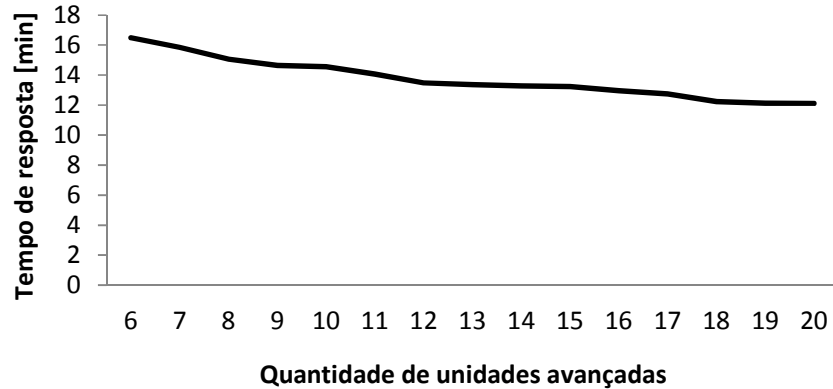


Figura 27: Análise de sensibilidade com o total de unidades avançadas
Fonte: Elaborado pelo autor (2014)

Por fim, o gráfico da Figura 28 apresenta a mesma análise mas considerando o tempo de resposta médio do sistema com o total de unidades (básicas + avançadas). O decréscimo é acentuado até um total de 31 unidades, quando então a inclinação do gráfico começa a diminuir. O tempo de resposta médio se estabiliza com uma quantidade aproximada de 41 unidades.

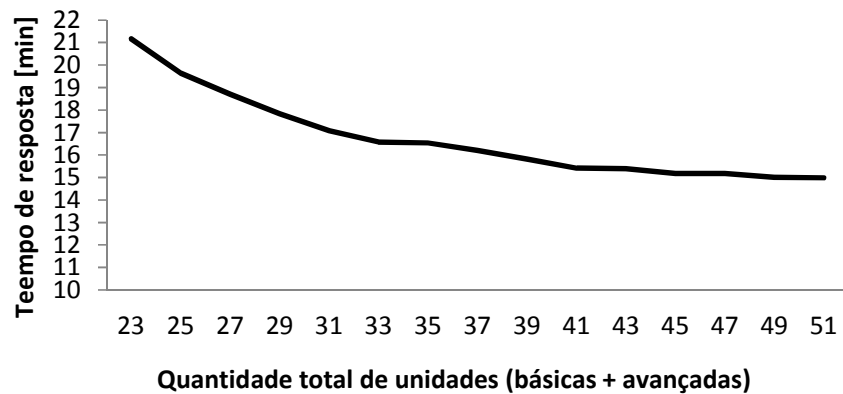


Figura 28: Análise de sensibilidade com o total de unidades
Fonte: Elaborado pelo autor (2014)

Através desta análise é possível verificar que o sistema se satura a partir de uma determinada quantidade de unidades em operação, deixando evidente que existem outros fatores importantes na estimativa do tempo de resposta que não só a quantidade de ambulâncias. Interessante observar também que as atividades relativas à central de regulação, que independem do total de unidades, representam aproximadamente 14% do tempo de resposta, considerando o ano de 2011. Uma maior eficiência dos atendentes,

médicos reguladores e despachantes, bem como uma diminuição nos tempos de deslocamento podem representar ganhos expressivos para o desempenho do sistema.

5.2 MÉTODOS HEURÍSTICOS PROPOSTOS

Para o processo de otimização para simulação, foram consideradas duas heurísticas. A primeira baseada na taxa de utilização de cada base (parâmetro fornecido pelo modelo de simulação) e utilizando dos conceitos da heurística busca tabu. A segunda baseada no tempo de resposta de cada configuração e utilizando da meta-heurística *simulated annealing*.

A heurística busca tabu foi proposta por Glover (1986), com o objetivo de explorar o espaço de soluções de um problema de otimização movendo-se de uma solução para outra que seja seu melhor vizinho. Utiliza-se de uma estrutura de memória para armazenar as soluções geradas em uma lista tabu de maneira a escapar de ótimos locais.

O método de *simulated annealing* foi proposto por Kirkpatrick *et al.* (1983). Em cada iteração, uma nova solução é gerada a partir da solução corrente por uma modificação aleatória nesta. Trocam-se as duas soluções se a nova solução é melhor que a atual. Se for pior, esta ainda pode ser aceita de acordo com uma probabilidade de aceitação que diminui à medida que a heurística avança. Este procedimento serve para escapar de ótimos locais. A probabilidade é definida de acordo com um parâmetro denominado *temperatura*.

As heurísticas foram desenvolvidas de maneira a levar em consideração a estimação dos parâmetros de desempenho do sistema a partir da simulação computacional. Conforme apresentado na seção anterior, estes parâmetros são estimados considerando todas as atividades relativas ao processo de resgate, o que aumenta o tempo computacional necessário para gerar estimativas estatisticamente confiáveis. A noção de explorar a vizinhança de uma solução candidata antes de mover para outra configuração foi adaptada de maneira a acelerar a chegada dos métodos na região de viabilidade do problema.

Seja s a configuração corrente, s° a configuração inicial, s^* a melhor configuração encontrada até o momento, $IterT$ o número da iteração corrente, $Tmax$ o número máximo de iterações sem melhora em s^* , $Temp$ o tempo corrente do procedimento, $TempMax$ o tempo máximo permitido para o procedimento, $MelhorIter$ a iteração mais recente que forneceu s^* , $TU[]$ um vetor com a taxa de utilização de cada base, $LT[]$ uma lista tabu para guardar as últimas λ configurações, $TU(s)$ a taxa de utilização média das unidades na configuração s , TU^* a taxa de utilização a ser atingida, $TR(s)$ o tempo de resposta da configuração s simulada, $TRpadrão$ o tempo de resposta utilizado como restrição do modelo, $f(s)$ a função objetivo da configuração s , $f(s^*)$ a função objetivo da configuração s^* , $iter$ a quantidade de iterações com a mesma quantidade de unidades, β um parâmetro que determina o número máximo de iterações consecutivas que uma base pode ficar sem unidades e Θ o número máximo de iterações consecutivas com a mesma quantidade total de unidades no sistema.

Os parâmetros β , λ , Θ e $Tmax$ serão estimados a partir de testes iniciais feitos com a heurística. O valor inicial para os parâmetros $Iter$ e $MelhorIter$ é zero e a lista tabu ($LT[]$) inicializa vazia. A função objetivo $f(s^*)$ é inicializada com um valor M muito grande, de maneira que a primeira solução viável encontrada possa ser aceita.

O primeiro método desenvolvido, baseado na lógica da meta-heurística busca tabu, opera de maneira resumida utilizando a seguinte lógica: durante Θ iterações, o procedimento mantém fixo o número total de ambulâncias no sistema; aumentando uma unidade na base com maior taxa de utilização e diminuindo uma unidade na base com menor taxa de utilização. Após Θ iterações, o procedimento diminui uma unidade no número total de ambulâncias caso alguma das soluções geradas tenha satisfeito os critérios de tempo de resposta e utilização, ou aumenta uma unidade caso nenhuma das soluções geradas tenha satisfeito os parâmetros de desempenho definidos.

O pseudocódigo do método desenvolvido está apresentado na Figura 29. A configuração inicial (s) é simulada (passo 5) e a taxa de utilização em cada base armazenada de maneira crescente no vetor $TU[]$ (passo 6). Caso o tempo de resposta ($TR(s)$) e a taxa de utilização ($TU(s)$) sejam menores que os valores definidos como padrão (passo 7), verifica-se se a função objetivo ($f(s)$) é menor que a melhor solução encontrada até o momento ($f(s^*)$) – passo 8).

```

1   $s \leftarrow s^\circ, LT[] \leftarrow \emptyset, iter \leftarrow 0, MelhorIter \leftarrow 0$ 
2  Inicialize função de aspiração A
3  while ( $IterT - MelhorIter \leq Tmax$ ) and ( $Temp \leq TempMax$ ) do
4     $IterT \leftarrow IterT + 1$ 
5    Simule s
6    sort TU[]
7    if ( $TR(s) \leq TR_{padr\tilde{a}o}$ ) and ( $TU(s) \leq TU^*$ ) then
8      if ( $f(s) < f(s^*)$ ) then
9         $s^* \leftarrow s, f(s^*) \leftarrow f(s)$ 
10        $MelhorIter \leftarrow IterT$ 
11     end if
12   end if
13   if ( $iter < \Theta$ ) then
14     Verifica Bases 1
15     Movimento de troca 1
16      $iter \leftarrow iter + 1$ 
17   else
18     Verifica Bases 2
19     Movimento de troca 2
20      $iter \leftarrow 0$ 
21   end if
22   Atualiza LT[], atualiza A
23 end while
24 return  $s^*$ 

```

Figura 29: Pseudocódigo do método baseado na Busca Tabu
 Fonte: Elaborado pelo autor (2014)

Em caso afirmativo, atualiza-se a solução (passo 9) e a melhor iteração (*MelhorIter* – passo 10). Neste caso, a vizinhança (Figura X) é definida da seguinte forma: caso o número máximo de iterações (Θ) com o mesmo número de unidades não for atingido (passo 13), a heurística procura aumentar uma unidade na base que possui maior taxa de utilização e diminuir uma unidade na base com menor taxa de utilização, caso esta não esteja com zero unidades e o movimento não for tabu (passo 15). O método *Verifica Bases 1* (passo 14) foi definido para evitar que uma base com zero unidades jamais seja escolhida para receber uma unidade. Caso uma determinada base esteja a β iterações com zero unidades, a mesma terá preferência para receber uma unidade, considerando que o movimento não seja tabu. A Figura 30 exemplifica a vizinhança descrita. A primeira linha representa o total de unidades em cada uma das sete bases e a segunda a taxa de utilização em casa, estimada pelo módulo de simulação. A base cinco apresenta

a maior taxa de utilização (80%) e receberá uma unidade, passando de 1 para 2. A base dois apresenta a menor taxa de utilização (55%) e terá 1 unidade diminuída no total, passando de 3 para 2.

θ^i	2	3	1	0	1	2	1
TU:	60	55	72	0	80	65	78
θ^{i+1}	2	2	1	0	2	2	1

Figura 30: Vizinhança 1 da Busca Tabu
Fonte: Elaborado pelo autor (2014)

Atualiza-se a variável *iter* (passo 16). Se a quantidade máxima de iterações (Θ) com o mesmo número de unidades for atingido (passo 17), a vizinhança é definida da seguinte forma: o método verifica se alguma das soluções geradas atingiu as restrições definidas para o tempo de resposta e a taxa de utilização. Em caso afirmativo, diminui-se uma unidade na base com menor taxa de utilização (caso a mesma possua alguma unidade); em caso negativo, aumenta-se uma unidade na base com maior taxa de utilização (respeitando a restrição da quantidade máxima de unidades em cada base). Esta lógica está indicada nos passos 18 a 20. A Figura 31 apresenta a geração de um novo cenário para o caso em que nenhuma solução viável for encontrada. A base cinco que apresenta a maior taxa de utilização irá receber uma nova unidade, passando de 1 para 2. O número total de ambulâncias no sistema é acrescido de 1.

θ^i	2	3	1	0	1	2	1
TU:	60	55	72	0	80	65	78
θ^{i+1}	2	3	1	0	2	2	1

Figura 31: Vizinhança 2 da Busca Tabu
Fonte: Elaborado pelo autor (2014)

A lista tabu $LT[]$ e a função de aspiração são atualizadas (passo 22) e o método retorna para o passo 4, caso o critério de parada (passo 3) não tenha sido atingido. A lista tabu utilizada pode eliminar ciclos, pois seu objetivo é garantir o não retorno de uma solução já visitada anteriormente por λ iterações. Entretanto, isto pode fazer com que movimentos para soluções que ainda não foram visitadas não sejam alcançados de acordo com as restrições da lista tabu. Por isto, uma função de aspiração é utilizada, retirando sob certas circunstâncias o status tabu de um movimento. De maneira resumida, para cada possível valor v da função objetivo existe um nível de aspiração $A(v)$, em que uma solução $s' \in V$ pode ser gerada se $f(s') < A(f(s))$, mesmo que o movimento m esteja na lista tabu. Para o método proposto, considerou-se $A(f(s) < f(s^*))$, onde s^* é a melhor solução encontrada até o momento, ou seja, aceita-se o movimento tabu se o mesmo levar a uma solução melhor ou igual à melhor solução corrente.

Para a segunda heurística (Figura 32), utilizou-se a lógica do *simulated annealing*.

```

1   $s \leftarrow s^0, T \leftarrow T^0$ 
2  while  $(T > 0.1)$  and  $(Temp \leq TempMax)$  do
3    Gere um vizinho aleatório  $s' \in N(s)$ 
4    Simule  $s'$ 
5     $\Delta = f(s') - f(s)$ 
6    if  $(TR(s') \leq TRpadr\tilde{a}o)$  and  $(TU(s) \leq TU^*)$  then
7      if  $(\Delta < 0)$  then
8         $s \leftarrow s'$ 
9        if  $(f(s') < f(s^*))$  then  $s^* \leftarrow s'$ 
10     else
11       sample  $x \in [0,1]$ 
12       if  $(x < e^{-(\Delta/T)})$  then  $s \leftarrow s'$ 
13     end if
14   else
15     Gere um vizinho aleatório  $s' \in N(s)$ 
16   end if
17    $T \leftarrow T - \alpha x T$ 
18   Atualiza Temp
19 end while
20 return  $s^*$ 

```

Figura 32: Pseudocódigo do método baseado no *Simulated Annealing*

Fonte: Elaborado pelo autor (2014)

Seja T a temperatura corrente, T° a temperatura inicial, $N(s)$ a vizinhança de s , s' um vizinho da solução s , s^0 a configuração inicial, α o parâmetro de resfriamento (utilizou-se o método de resfriamento geométrico - Kirkpatrick *et al.* (1983)), $Temp$ o tempo corrente do procedimento e $TempMax$ o tempo máximo permitido para o procedimento. Os parâmetros T° e α serão estimados a partir de testes iniciais feitos com a heurística.

Em cada iteração, um vizinho s' da solução corrente s é gerado (passo 3) e analisado via simulação (passo 4). O método analisa se o tempo de resposta da configuração s' bem como a taxa de utilização estão dentro dos parâmetros de desempenho definidos (passo 6). Em caso afirmativo, verifica-se se a solução vizinha s' é melhor que a solução corrente s (a primeira solução viável será aceita); em caso afirmativo, atualiza-se o vetor de solução (passo 8). Além disso, se a função objetivo da configuração s' ($f(s')$) for menor que a função objetivo da melhor configuração encontrada até o momento ($f(s^*)$), trocam-se as duas (passo 9). Caso a solução vizinha seja pior que a solução corrente, aceita-se a configuração s' com uma determinada probabilidade baseada na temperatura corrente (passos 11 e 12). Este procedimento busca evitar que o método fique preso em um ótimo local, convergindo mais rapidamente para soluções de boa qualidade.

A vizinhança é definida da seguinte forma: caso a configuração simulada não tenha atingido o tempo de resposta e taxa de utilização padrão (passo 14), gera-se um novo vizinho aumentando uma unidade em uma base escolhida aleatoriamente, desde que a mesma não tenha atingido o número máximo permitido de unidades. A Figura 33 mostra a configuração atual θ^i com sete bases e o total de unidades em cada. O método escolhe aleatoriamente uma das bases (no exemplo, a base dois) e diminui uma unidade, passando de 3 para 2.

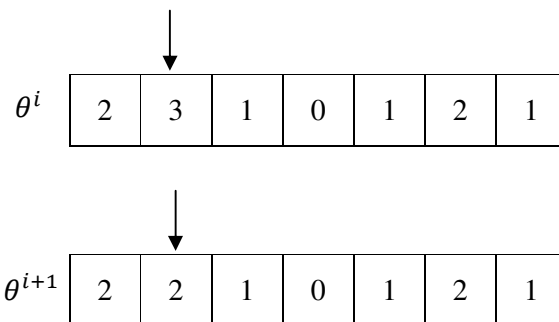


Figura 33: Vizinhança do método *Simulated Annealing*

Fonte: Elaborado pelo autor (2014)

A temperatura T e o tempo $Temp$ são atualizados (passos 17 e 18) e retorna-se para o passo 3. O procedimento é finalizado quando a temperatura corrente atinge um valor limite ou se o tempo máximo para o procedimento for atingido (passo 2). A vizinhança $N(s)$ foi definida de maneira aleatória, aumentando uma unidade caso $TR(s) > TR_{padrão}$ ou diminuindo uma unidade em caso contrário. A escolha da base foi feita utilizando uma distribuição uniforme discreta com probabilidade $(1/n_b = 20) \times 100$.

5.3 METAMODELO DESENVOLVIDO

De maneira a construir um metamodelo de regressão representativo de um processo de simulação, é necessária a escolha de um método para geração dos dados de ajuste. Para tanto, utilizou-se o Modelo Hipercubo Latino (MHL) para amostrar o número de unidades avançadas (Q_A) e básicas (Q_B) em cada base. Diversos autores já desenvolveram estudos comprovando a eficácia do método quando aplicado em modelos de simulação discreta (Jin *et al.*, 2000; Meckesheimer *et al.*, 2002).

O MHL para amostragem divide o espaço amostral de cada uma das k variáveis em n intervalos equiprováveis, e um ponto é escolhido em cada um destes intervalos; isto gera uma matriz de k colunas, formadas pelas amostras de cada variável, que são depois permutadas aleatoriamente, gerando a amostra final de n vetores de k dimensões. Para os experimentos, a partir de conversas com o responsável pela gestão do sistema, considerou-se os intervalos $0 \leq Q_{USB} \leq 4$ e $0 \leq Q_{USA} \leq 4$. Foram propostos 250 experimentos de acordo com o MHL para construção do metamodelo e analisado o tempo de resposta médio para as unidades básicas e avançadas, separadamente. Cada um dos 250 cenários foi analisado no módulo de simulação desenvolvido e anotada a estimativa fornecida do tempo de resposta médio do sistema. Esta informação, juntamente com a configuração, foi introduzida no Minitab 17.1.0 para construção do metamodelo.

Definido o conjunto de pontos da amostra, torna-se necessária a escolha do modelo para representação dos dados gerados. Kleijnen (2009a) afirma que metamodelos baseados em regressões polinomiais podem ser utilizados para explicar o comportamento de modelos de simulação, bem como para predição do valor esperado dos parâmetros de desempenho do sistema dada uma combinação das variáveis de entrada. Jin *et al.* (2001)

afirmam que é possível identificar a significância dos diferentes fatores do experimento diretamente dos coeficientes do modelo de regressão. Para problemas de grandes dimensões, aconselha-se o uso de modelos polinomiais de regressão de primeira ou segunda ordem para reduzir as variáveis de interesse para aquelas mais importantes.

A escolha pelo modelo de regressão linear múltipla, no caso do SAMU-BH, além dos motivos expostos, baseou-se também no número elevado de variáveis (20 bases com dois tipos de ambulância = 40 variáveis), tornando complexo o uso de modelos não lineares. A premissa de relação linear será avaliada através de um teste de hipótese baseado na estatística F da distribuição de Fisher.

A regressão linear múltipla é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis estudadas (variável dependente ou resposta) e o restante das variáveis que descrevem o sistema (variáveis independentes ou explicativas). Sua formulação proposta por dos Santos (2002) é dada por:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (21)$$

onde $\beta_0, \beta_1, \dots, \beta_p$ são os parâmetros estimados pelo modelo, x_1, \dots, x_p são constantes conhecidas (quantidade de unidades em cada base – $p = 20$), ϵ representa os erros das estimativas e Y a variável dependente (tempo de resposta médio das unidades básicas ou avançadas). Os metamodelos foram gerados com auxílio do *software* estatístico Minitab 17.1.0. A análise dos erros do modelo se baseou no gráfico de distribuição normal dos resíduos, auxiliando na verificação da hipótese de normalidade. Para ambos os metamodelos, verificou-se os pontos dispostos em torno de uma linha reta, validando o pressuposto de que os erros seguem uma distribuição normal de média μ e variância σ^2 . A hipótese de linearidade dos dados foi verificada através do gráfico de resíduos versus valores ajustados. Para os metamodelos propostos, observou-se os pontos do gráfico distribuídos de forma dispersa e aleatória em torno da reta que corresponde ao resíduo zero, formando uma mancha de largura uniforme. Dessa forma, pode-se esperar que os erros sejam independentes com média nula e variância constante.

Para verificar se as variáveis independentes contribuem significativamente com informação para explicar linearmente a variação da variável resposta, utilizou-se um

teste de hipóteses baseado na estatística F com nível de significância $\alpha = 5\%$. As hipóteses a testar são: $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ e $H_1: \beta_j \neq 0$ para algum $j, j = 1, \dots, p$; ou de uma maneira alternativa, H_0 : não existe relação linear entre a variável dependente e o conjunto de variáveis independentes utilizadas. O critério de rejeição da hipótese nula se baseia na relação dada por

$$f_0 > f_t = f_\alpha[k, n - k] \quad (22)$$

onde α representa o grau de significância, f_0 o valor observado a partir da regressão, f_t o valor Tabelado para a distribuição F, k os graus de liberdade do modelo de regressão e $n - k$ os graus de liberdade do erro residual. Para os dados utilizados, através da Tabela ANOVA (*analysis of variance*), observou-se um valor de $f_0^{USB} = 46,94 > f_t = 1,614$ e $f_0^{USA} = 24,83 > f_t = 1,614$, concluindo pela rejeição da hipótese nula, ou seja, pode-se afirmar com 95% de confiança que o tempo de resposta do sistema está linearmente relacionado (pela equação de regressão linear) com a quantidade de unidades em cada uma das 20 bases consideradas.

O fato do modelo de regressão ser estatisticamente significativo não implica necessariamente que a relação encontrada seja o modelo mais adequado para estimar e/ou prever o tempo de resposta do sistema a partir de valores das variáveis independentes. Utilizou-se, para tanto, o coeficiente de determinação ajustado, que é uma medida da proporção da variação da variável resposta Y que é explicada pela equação de regressão quando estão envolvidas as variáveis independentes $x_1, x_2, \dots, x_{k=20}$. O metamodelo para as unidades básicas apresentou um coeficiente de determinação ajustado $R_{ajust.}^2 = 0,81$. Sua equação de regressão (23) é dada por:

$$Y_{USB} = 21.0 - 0.0985x_1 - 0.0309x_2 - 0.0968x_3 - 0.202x_4 - 0.167x_5 - 0.132x_6 - 0.126x_7 - 0.133x_8 - 0.117x_9 - 0.0968x_{10} - 0.242x_{11} - 0.251x_{12} - 0.0554x_{13} - 0.102x_{14} - 0.0863x_{15} - 0.0744x_{16} - 0.0578x_{17} - 0.127x_{18} - 0.0488x_{19} - 0.134x_{20} \quad (23)$$

O metamodelo para as unidades avançadas apresentou um coeficiente de determinação ajustado $R^2_{ajust.} = 0,80$. Sua equação de regressão (24) é dada por:

$$Y_{USA} = 20.3 - 0.0521x_1 - 0.0062x_2 - 0.0770x_3 - 0.142x_4 - 0.165x_5 - 0.189x_6 - 0.0636x_7 - 0.0549x_8 - 0.141x_9 - 0.0746x_{10} - 0.219x_{11} - 0.232x_{12} - 0.0380x_{13} - 0.0667x_{14} - 0.0588x_{15} - 0.0605x_{16} - 0.111x_{17} - 0.113x_{18} - 0.0724x_{19} - 0.146x_{20} \quad (24)$$

As Figuras 34 e 35 apresentam, para as unidades básicas e avançadas respectivamente, um gráfico comparando, para cada um dos 250 cenários, a estimativa gerada pelo modelo de simulação e aquela gerada pelo metamodelo.

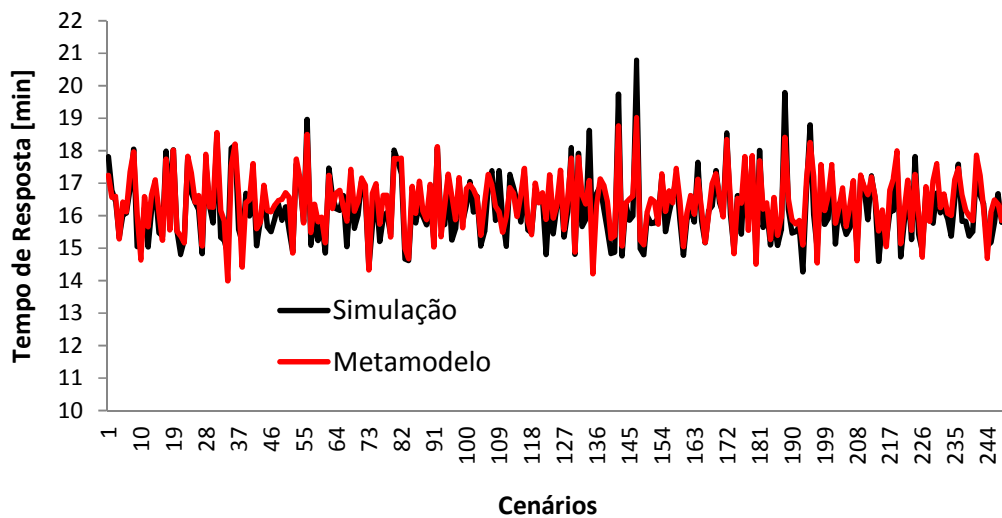


Figura 34: Metamodelo para unidades básicas
Fonte: Elaborado pelo autor (2014)

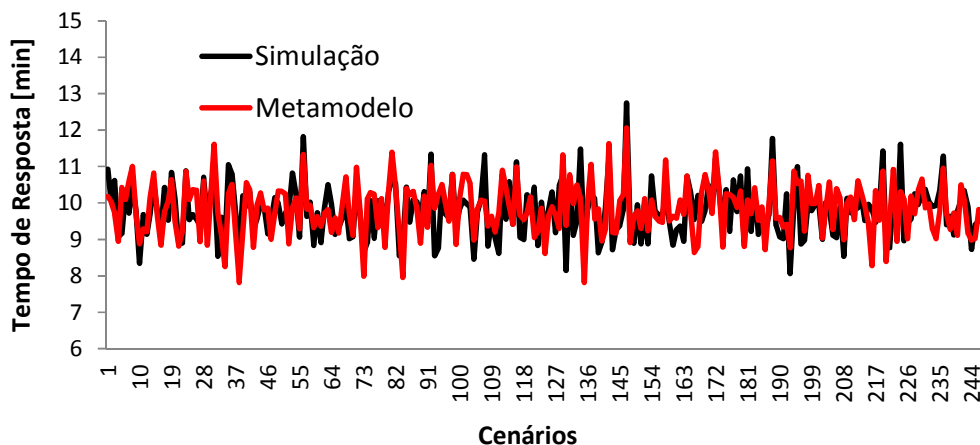


Figura 35: Metamodelo para unidades avançadas
Fonte: Elaborado pelo autor (2014)

Pode-se observar, a partir dos gráficos, que o modelo de regressão se ajusta bem aos dados propostos, com algumas exceções para valores extremos de pico. Os valores de pico nas Figuras acima estão associados com uma menor quantidade de ambulâncias em operação. Desta maneira, a dispersão das unidades ao longo do território considerado tem um peso maior do que nos outros cenários, não se adequando de maneira tão satisfatória à hipótese de linearidade.

5.4 RESULTADOS COMPUTACIONAIS

Para análise e validação de métodos heurísticos, as melhores instâncias de teste são provavelmente aquelas derivadas de aplicações reais (RARDIN e UZSOY, 2001). Segundo os autores, é possível considerar automaticamente uma série de padrões e relações entre as variáveis que influem no comportamento do sistema. Utilizou-se nesta etapa o SAMU de Belo Horizonte conforme detalhado na seção 5.1.10.

Para o processo de otimização via simulação, alguns custos de operação e ativação de bases foram estimados considerando-se dados coletados no Sistema Único de Saúde (disponível em www.datasus.gov.br). Os valores estão de acordo com o trabalho de Nogueira Junior (2011).

i. Custo de ativação da base de operação b (β_b)

Estas despesas são relativas à instalação física da base. O custo médio de ativação de cada base de operação do SAMU de Belo Horizonte corresponde a R\$ 350.000,00.

ii. Custo de aquisição de ambulância do tipo u (η_u)

O valor médio considerado para a aquisição de uma ambulância básica é de R\$ 150.000,00 e de uma ambulância avançada de R\$ 225.000,00.

A metodologia de otimização via simulação proposta foi utilizada considerando o SAMU de Belo Horizonte no ano de 2011. Uma premissa adotada para o processo de otimização é a independência no atendimento entre as unidades básicas e avançadas, ou seja, chamadas que necessitam de unidade básica só são atendidas por este tipo de ambulância e chamadas que necessitam de unidade avançada só são atendidas por ambulâncias avançadas. Esta premissa considera o fato de menos de 9% das chamadas necessitarem unidades avançadas, além destas possuírem taxa de utilização bem menor que as básicas, ou seja, existe uma probabilidade alta de estarem disponíveis na chegada de uma chamada grave e realmente estarem mais próximas (em relação ao tempo) do incidente do que qualquer unidade básica.

Para validar tal hipótese, o SAMU-BH foi simulado com as duas abordagens: atendimentos independentes (alternativa 1: TR^1) e atendimento avançado podendo ser

feito por unidades básicas caso estas estejam mais próximas (alternativa 2: TR^2). Os resultados mostraram que $|TR_{USB}^1 - TR_{USB}^2| < 1,4\%$ e $|TR_{USA}^1 - TR_{USA}^2| < 0,08\%$, com uma diferença de menos de 15 segundos para as básicas e menos de 1 segundo para as avançadas.

Para os testes com o procedimento de otimização para simulação apresentados, os valores para os parâmetros de controle e restrições de desempenho foram definidos de acordo com gerência do Serviço de Atendimento. Para maiores detalhes sugere-se consultar o trabalho de Silva (2010). Os valores utilizados foram: $TR_{USB}^* = 18$ minutos, $TR_{USA}^* = 15$ minutos, $TU_{USB}^* = 50\%$, $TU_{USA}^* = 30\%$, $T_{bUSB}^- = Q_{USB}^- = 0 \forall b = 1, \dots, n$, $T_{bUSA}^- = Q_{USA}^- = 0, \forall b = 1, \dots, n$, $T_{bUSB}^+ = T_{bUSA}^+ = 4, \forall b = 1, \dots, n$ e um tempo máximo de máquina de 24 horas.

Para os parâmetros da heurística baseada no método de *simulated annealing*, foi utilizada a simulação para determinação dos valores mais adequados. A temperatura inicial foi determinada pela média dos custos das soluções vizinhas associadas com a configuração inicial. Os parâmetros utilizados foram definidos como $T_{USB}^0 = 8.200.000$, $T_{USA}^0 = 3.500.000$ e $\alpha = 0,05$.

Para os parâmetros da heurística baseada no método de busca tabu, utilizou-se novamente testes de simulação para identificação dos valores mais adequados. Foram definidos os parâmetros: $\lambda = 7$ iterações (para definição de movimento tabu), $\theta = 15$ (número de iterações com a mesma quantidade de ambulâncias), $\beta = 10$ (número máximo consecutivo de iterações que uma base pode ficar com zero unidades) e $Tmax = 50$ iterações.

As heurísticas foram implementadas no ambiente de programação Dev-C++ 4.9.9.2 e o modelo de simulação implementado no *software* Arena 9.0 da *Rockwell Automation Technologies*. Os testes foram feitos em um computador com processador Intel(R) Pentium(R) M com 1.60 GHz de velocidade e 1.24 GB de memória RAM.

Os dois métodos foram comparados entre si e com o otimizador *OptQuest* da OptTek Systems, Inc. integrado com o *software* de simulação Arena da *Rockwell Automation Technologies*, todos eles considerando a mesma configuração inicial do sistema. O

método baseado no *Simulated Annealing*, por apresentar um componente aleatório, foi otimizado um total de dez vezes, variando a semente geradora dos números aleatórios entre uma otimização e outra. Na Tabela 11 é apresentada a melhor solução encontrada dentre as dez.

A heurística que apresentou os melhores resultados foi a baseada na busca tabu (Tabela 11). A mesma forneceu a melhor solução com o menor número de bases e no menor número de iterações. Para o método *Simulated Annealing*, calculou-se o desvio padrão das melhores soluções geradas em cada uma das dez otimizações. O valor encontrado em termos percentuais, com relação ao custo da melhor configuração, foi de 3,91%.

Tabela 11: Resultados comparativos das heurísticas

Informações	Simulated Annealing	Busca Tabu	OptQuest
Melhor iteração USB	145	111	149
Melhor iteração USA	98	79	106
Nº de unidades USB	21	21	21
Nº de unidades USA	7	7	7
Total de unidades	28	28	28
Nº de bases abertas	13	12	19
% soluções viáveis	42%	30%	23%
Utilização média USB	48%	49%	49%
Utilização média USA	27%	26%	28%
Função objetivo (R\$)	9.275.000	8.925.000	11.375.000
Tempo de Resposta USB [min]	17,82	17,94	17,91
Tempo de Resposta USA [min]	14,45	14,78	14,73

Fonte: Elaborado pelo autor (2014)

Interessante observar que a solução ótima gerada pelos três métodos apresentou a mesma quantidade de unidades básicas (21) e avançadas (7), totalizando 28 ambulâncias em operação no sistema para atingir os critérios de desempenho propostos. Entretanto, os dois métodos propostos baseados nas meta-heurísticas *Simulated Annealing* (SA) e Busca Tabu (BT) conseguiram gerar uma configuração com menos bases do que aquela gerada pelo otimizador *OptQuest*. O método baseado na BT instalou 7 menos bases quando comparado com o *OptQuest*.

Foi possível observar também que existe uma tendência maior de concentração de mais unidades em menos bases nos métodos propostos quando comparado com o *OptQuest*. Na configuração do SAMU-BH para o ano de 2011, haviam apenas duas bases com

mais de uma unidade básica instalada. O método SA instalou 3 bases com 3 unidades em cada e o método BT instalou 6 bases com 3 unidades cada. A função objetivo da melhor solução encontrada pelo método de busca tabu teve uma diferença percentual de 3,92% para o método *simulated annealing* e 27,45% para o OptQuest.

As taxas de utilização das unidades básicas e avançadas apresentaram valores bastante próximos, uma vez que os três métodos apresentaram a mesma quantidade de unidades. As diferenças são devido à própria aleatoriedade das atividades consideradas no modelo de simulação, o que faz com que as estimativas geradas para os parâmetros de desempenho também possuam uma incerteza associada. Com relação às soluções geradas, 30% foram viáveis para o método busca tabu, 42% para o *simulated annealing* e 23% para o OptQuest. Para um sistema com um único tipo de unidade, 20 bases de operação com no máximo 4 ambulâncias em cada, o número de possibilidades está na casa de 10^{14} . Apesar de chegar mais rapidamente na região de viabilidade do problema, ao utilizar uma lógica aleatória de geração de cenários, o método *Simulated Annealing* não conseguiu gerar uma solução melhor do que aquela apresentada pelo método Busca Tabu. Os resultados indicam que, por considerar a taxa de utilização de cada base, o método baseado na busca tabu tende a convergir mais rapidamente para soluções de boa qualidade quando comparado com os demais métodos.

Importante mencionar que, devido ao caráter estocástico dos parâmetros de entrada do módulo de simulação, existe uma aleatoriedade associada às estimativas do tempo de resposta médio em cada método. Estatisticamente, é impossível diferenciá-las em termos de seus valores, de maneira a identificar a melhor. Utilizou-se de várias replicações juntamente com a inclusão de um tempo de aquecimento para construção de uma estimativa do tempo de resposta baseada em um valor médio. Entretanto, o foco da análise deve estar na configuração e nos custos associados com cada, conforme apresentado na Tabela 11. A melhor solução encontrada pelo método BT gerou uma configuração com custo mais de R\$ 3.000.000,00 menor quando comparado com o *OptQuest*.

A Tabela 12 apresenta a distribuição das ambulâncias nas 20 bases consideradas para cada um dos métodos testados. A coluna “Configuração 2011” refere-se à distribuição das unidades no ano de 2011 para o SAMU de Belo Horizonte. As demais colunas

dizem respeito à melhor solução obtida pelo respectivo método de otimização considerando as restrições de tempo de resposta descritas nesta seção (18 minutos para unidades básicas e 15 para unidades avançadas). O desvio padrão e o coeficiente de variação nas últimas duas linhas foram calculados de acordo com a quantidade de ambulâncias alocadas em cada base. Como a quantidade total de unidades é a mesma para os três procedimentos de otimização, o coeficiente de variação serve como um estimador para medir a dispersão na distribuição das unidades. A análise está em consonância com aquela apresentada no trabalho de Zhen *et al.* (2014).

Tabela 12: Distribuição das ambulâncias em cada base

Bases	Regional	Configuração 2011*	Otimizador OptQuest	<i>Simulated Annealing</i>	Busca Tabu
1	Venda Nova	2	1	0	2
2	Venda Nova	1	2	0	0
3	Venda Nova	1	2	1	1
4	Norte	1	1	1	0
5	Pampulha	1	3	0	3
6	Pampulha	1	1	4	2
7	Nordeste	1	1	3	3
8	Noroeste	1	0	0	0
9	Noroeste	1	1	3	3
10	Noroeste	1	2	1	0
11	Leste	2	2	0	1
12	Oeste	1	2	2	3
13	Oeste	1	1	3	3
14	Oeste	1	1	0	0
15	Oeste	1	2	1	2
16	Centro-Sul	1	1	2	3
17	Centro-Sul	2	1	0	0
18	Centro-Sul	1	2	2	0
19	Barreiro	1	1	1	0
20	Barreiro	1	1	4	2
Total de unidades		23	28	28	28
Total de bases		20	19	13	12
Desvio Padrão		0,36	0,66	1,36	1,28
Coef. de Variação		0,31	0,47	0,97	0,91

* para um tempo de resposta médio de 20,72 minutos

Fonte: Elaborado pelo autor (2014)

É possível observar a partir das informações da Tabela 12, que para tempos de resposta mais restritivos, conforme cenário apresentado para otimização, o plano ótimo de

distribuição das unidades se torna cada vez mais desigual ou desbalanceado entre as diversas bases. Na configuração 2011, todas as 20 bases possuíam pelo menos uma unidade (básica ou avançada) alocada. Com o otimizador *OptQuest* foram abertas 19 bases, na otimização com o método *Simulated Annealing*, 7 bases ficaram sem unidades e no método Busca Tabu, 8 bases não foram utilizadas.

O desvio padrão da distribuição das unidades nestes dois últimos métodos é mais do que o dobro quando comparado com o *OptQuest*. Considerando os dois tipos de ambulância, o método *Simulated Annealing* abriu duas bases com 4 unidades cada e três bases com 3 unidades cada. Para o método Busca Tabu, os valores foram nenhuma base com 4 unidades e seis bases com 3 unidades cada. A distribuição mais desigual sugere que os procedimentos de otimização empregados são importantes na alocação das ambulâncias para tempos de resposta mais restritivos, ao invés de simplesmente distribuir as unidades entre as bases de maneira balanceada. A razão para isso pode estar relacionada com o fato de que sob situações de maior exigência do sistema, as distribuições desiguais das ocorrências nas diversas regiões bem como condições de tráfego terão uma influência mais direta em um sistema com uma alocação de unidades menos balanceado. A análise e os resultados encontrados estão de acordo com aqueles apresentados no trabalho de Zhen *et al.* (2014).

Para testar a metodologia completa proposta, os métodos heurísticos desenvolvidos foram implementados juntamente com o metamodelo para verificar sua adequação no sentido de acelerar o processo de busca de uma solução de qualidade. Como o metamodelo gerado não apresentou um ajuste com qualidade tão elevada (coeficiente de determinação ajustado de 0,81 para unidades básicas e 0,80 para unidades avançadas), o mesmo não foi capaz de acelerar o processo de otimização para um $TR_{padr\tilde{a}o}^{USB} = 18$ minutos e $TR_{padr\tilde{a}o}^{USA} = 15$ minutos; considerando que o tempo de resposta médio do sistema em 2011 já estava bastante próximo destes valores ($TR_{2011} = 20,72$ minutos). Conforme explicado, como o metamodelo opera com a lógica: se $|TR_{padr\tilde{a}o} - TR_{metamodelo}| \leq \alpha$ então a configuração gerada é analisada via simulação, o parâmetro α deveria ser calibrado em valores muito baixos para funcionar como um filtro, o que acarretaria na simulação de praticamente todos os cenários gerados. De maneira a contornar tal situação e analisar o funcionamento da metodologia proposta em um

contexto real, utilizou-se o cenário SAMU-BH 2011, mas assumindo uma configuração inicial com 5 unidades básicas e 2 unidades avançadas.

Esta configuração simulada apresentou um $TR_{USB} = 30$ minutos e $TR_{USA} = 27$ minutos. Utilizou-se as mesmas heurísticas apresentadas anteriormente, mas agora com $TR_{padr\tilde{a}o}^{USB} = TR_{padr\tilde{a}o}^{USA} = 15$ minutos. Para o caso específico da heurística busca tabu, caso uma configuração seja analisada pelo metamodelo, é necessário gerar uma estimativa da taxa de utilização em cada base para que o método possa gerar um novo cenário a ser analisado. Esta estimativa foi construída a partir dos testes feitos na geração de dados do metamodelo, utilizando da lógica do MHL. A partir dos 250 experimentos gerados, para cada base em cada quantidade de unidades (de zero a quatro), utilizou-se um valor médio para a taxa de utilização. Desta maneira, construiu-se uma matriz $TU_{n \times m}$ com $n = 20$ bases e $m = 5$ configurações diferentes. Estes valores são utilizados toda vez que um cenário é analisado via metamodelo, quando então não é possível gerar estimativas para as taxas de utilização a partir da simulação computacional.

Analisando os 250 experimentos gerados através do MHL para a construção do metamodelo baseado em regressão linear múltipla, observou-se que a maior diferença absoluta percentual entre o tempo de resposta estimado da simulação e o estimado pelo metamodelo foi de 9,5%. Considerando-se este valor, e através de testes iniciais de sensibilidade conduzidos com as heurísticas, definiu-se que os cenários gerados serão analisados pelo modelo de simulação caso $|TR_{padr\tilde{a}o} - TR_{metamodelo}| \leq \alpha = 2$ minutos. A Tabela 13 apresenta um resumo dos resultados encontrados.

Utilizou-se novamente os métodos heurísticos propostos sem o metamodelo, o otimizador *OptQuest* e cada método heurístico juntamente com o metamodelo desenvolvido. O metamodelo foi utilizado para chegar mais rapidamente na região de viabilidade do problema, quando então são utilizados os métodos heurísticos descritos anteriormente para encontrar uma solução de qualidade. Caso o tempo de resposta estimado pelo metamodelo esteja fora do tempo de resposta padrão (15 minutos), uma nova solução é gerada utilizando dos procedimentos descritos anteriormente.

Tabela 13: Resultados da metodologia proposta

Informações	Simulated Annealing	Busca Tabu	OptQuest	Simulated Annealing + Metamodelo	Busca Tabu + Metamodelo
Melhor iteração USB	154	149	88	184	176
Melhor iteração USA	110	99	175	74	99
Nº de unidades USB	39	41	35	31	30
Nº de unidades USA	10	10	8	7	7
Total de unidades	49	51	43	38	37
Nº de bases abertas	18	17	19	15	14
Utilização média USB	30%	31%	39%	47%	48%
Utilização média USA	24%	23%	28%	26%	27%
Função objetivo	R\$ 14.400.000,00	R\$ 14.350.000,00	R\$ 13.700.000,00	R\$ 11.475.000,00	R\$ 10.975.000,00
Tempo de Resposta USB [min]	14,88	14,91	14,94	14,79	14,85
Tempo de Resposta USA [min]	14,71	14,69	14,95	14,47	14,69

Fonte: Elaborado pelo autor (2014)

É possível perceber, inicialmente, que os métodos sozinhos não foram superiores ao *OptQuest* com relação à melhor solução gerada. Como o tempo de resposta padrão estabelecido como restrição estava distante do tempo de resposta observado do sistema (50% para as unidades básicas e 45% para as unidades avançadas), os dois procedimentos demoraram mais para entrar na região de viabilidade do problema, retornando uma solução aproximadamente 4,53% maior com relação à função objetivo quando comparado com o *OptQuest*. Mesmo utilizando menos bases, os métodos BT e SA utilizaram mais unidades dos dois tipos para alcançar o tempo de resposta proposto de 15 minutos, ficando em média com uma diferença de 14% quando comparado com o total de unidades na melhor solução fornecida pelo *OptQuest*. O método SA apresentou um desvio padrão das dez otimizações que foram feitas de 2,12%.

Comparando-se os resultados fornecidos pelo otimizador do Arena com aqueles observados na Tabela 11 (cujo tempo de resposta padrão para unidades avançadas eram os mesmos 15 minutos e a configuração inicial possuía mais unidades), apesar de ter gerado soluções melhores do que os métodos heurísticos sozinhos, o *OptQuest* não conseguiu alcançar as mesmas sete unidades propostas anteriormente, mantendo o mesmo número de bases.

Quando analisamos os métodos heurísticos juntamente com o metamodelo desenvolvido, pode-se perceber que o mesmo atingiu o objetivo inicialmente proposto de acelerar o processo de busca e gerar uma solução de melhor qualidade. Ao eliminar rapidamente soluções candidatas muito distantes do tempo de resposta proposto, o metamodelo conduz o procedimento de otimização mais velozmente para a região de soluções viáveis do sistema. Ambos os métodos (duas últimas colunas da Tabela 13) conseguiram alocar as mesmas 7 unidades avançadas propostas anteriormente, mas o método baseado na busca tabu conseguiu fornecer uma solução com uma unidade básica a menos quando comparado com o método baseado na lógica do *simulated annealing*.

Novamente é importante destacar a impossibilidade de diferenciar estatisticamente os valores gerados para o tempo de resposta do sistema em cada configuração, devido à aleatoriedade dos dados utilizados no modelo de simulação. Quando analisamos o custo das configurações, aquela baseada na BT juntamente com o metamodelo apresentou um

valor quase R\$ 3.000.000,00 menor quando comparado com a melhor configuração gerada pelo *OptQuest*.

A taxa de utilização das unidades de ambos os métodos ficou um pouco acima dos demais, justamente por operarem com um número menor de unidades, tanto básicas quanto avançadas. O método de busca tabu com o metamodelo conseguiu novamente alocar a menor quantidade de bases, uma a menos do que o método de *simulated annealing* com o metamodelo. A melhor solução encontrada entre todos os procedimentos testados utilizou uma quantidade de 30 unidades básicas, 7 avançadas e 14 bases. Quando comparada com os métodos SA, BT, *OptQuest* e SA com metamodelo, a função objetivo foi, respectivamente, 24%, 23%, 19% e 3,5% menor. Importante destacar que todos os métodos, além da melhor solução encontrada, geraram soluções viáveis próximas do ótimo local fornecido que podem ser utilizadas como alternativas de operação para os gestores do sistema.

6. CONCLUSÕES E TRABALHOS FUTUROS

A maioria dos trabalhos relacionados com a análise e dimensionamento de sistemas móveis de emergência utilizam técnicas de otimização ou simulação de maneira não integrada. O presente trabalho utilizou técnicas de otimização para simulação de maneira a desenvolver um plano de alocação ótimo levando em consideração fatores estocásticos do sistema relacionados com as solicitações de serviço, condições de tráfego, velocidade das unidades, dentre outros. Com a metodologia proposta foi possível resolver de maneira eficiente o problema de localização de unidades de resgate com objetivo de minimizar custos de abertura de bases e aquisição de ambulâncias.

As principais contribuições para a área de estudo de pesquisa operacional em sistemas de saúde dizem respeito à integração efetiva de um simulador baseado em eventos discretos com um componente de otimização baseado nas meta-heurísticas *Simulated Annealing* e Busca Tabu. Ambos os métodos foram capazes de superar o pacote de otimização comercial *OptQuest* integrado com o *software* Arena com relação ao custo das melhores configurações geradas. O trabalho também contribuiu para a área ao analisar o uso de metamodelos para acelerar o processo de busca de uma solução de qualidade no contexto de alocação de ambulâncias. Apesar desta técnica ser amplamente discutida na literatura, nunca havia sido utilizada juntamente com a metodologia de otimização para simulação proposta para sistemas móveis de emergência. Os testes conduzidos com os metamodelos também superaram os resultados fornecidos pelo *OptQuest*.

A metodologia foi aplicada no Serviço de Atendimento Móvel da cidade de Belo Horizonte. Com relação aos resultados computacionais, os testes mostraram que os métodos heurísticos propostos sem o metamodelo, superaram os resultados fornecidos pelo *OptQuest* quando o tempo de resposta atual do sistema está próximo do tempo de resposta estipulado como restrição. Quando o tempo de resposta está longe do padrão, os métodos superaram o otimizador *OptQuest* quando operacionalizados juntamente com os metamodelos desenvolvidos. No primeiro caso, a melhor solução encontrada foi pelo método busca tabu com uma diferença percentual da função objetivo de 27,45% para a melhor solução gerada pelo *OptQuest*. No segundo caso, a melhor solução encontrada

foi pelo método busca tabu com o metamodelo, com uma diferença percentual de 19% quando comparada novamente com o *OptQuest*.

Importante mencionar também que os parâmetros de entrada do modelo de simulação influenciam diretamente no tempo de resposta estimado pelo modelo para o sistema de ambulância. Soluções ainda melhores poderão ser encontradas caso o procedimento proposto seja utilizado considerando novas políticas de operação do sistema, como por exemplo reposição nos hospitais ao invés de unicamente na central de regulação, convênio com mais de uma oficina para realização das manutenções, treinamento dos motoristas, acesso a mais centros de atendimento para evitar esperas prolongadas, utilização de outros tipos de veículos que consigam desenvolver uma velocidade média maior para prestar um atendimento inicial em menor tempo, dentre outros.

Como sugestão para trabalhos futuros, o procedimento proposto pode ser testado utilizando novas abordagens heurísticas e metamodelos mais sofisticados de maneira a verificar sua adequação. Técnicas como a otimização robusta também podem ser utilizadas para resolução do problema de dimensionamento de sistemas de emergência. Esta técnica é baseada na programação estocástica composta da abordagem por cenários e da programação por metas. Outra alternativa seria a utilização de uma função multi-objetivo para considerar a minimização de custos do sistema e minimização do tempo de resposta.

Um estudo para identificar as variáveis mais significativas na construção do metamodelo pode gerar ajustes mais precisos para estimação do tempo de resposta do sistema. Ainda com relação ao metamodelo, sua geração pode ser feita de maneira dinâmica e integrada com o procedimento de otimização via simulação. Cada cenário analisado pode ser utilizado para alimentar o metamodelo e ir gerando melhores aproximações com o aumento no número de iterações.

O uso de algum modelo de otimização para geração de uma solução inicial pode auxiliar o processo de busca e gerar soluções de melhor qualidade. Os testes feitos consideraram sempre a configuração atual do sistema como solução inicial. Os métodos propostos não foram analisados frente a alterações nos parâmetros de entrada, como por exemplo, quantidade de ambulâncias, bases, hospitais ou taxas de chegada de solicitações.

O módulo de simulação pode incorporar algumas atividades que não foram implementadas, como por exemplo, o desvio da ambulância para um outro centro de atendimento caso o escolhido esteja sem vaga ou com falta de profissional da saúde necessário para o atendimento.

Outro aspecto que pode ser abordado é a adequação da metodologia para resolver o problema de sequenciamento das ambulâncias, determinando a quantidade ideal a ser alocada em cada base ao longo do dia. Tal procedimento pode reduzir o número total de unidades a serem utilizadas pelo sistema. Por fim, a utilização da metodologia proposta em sistemas diferentes pode auxiliar na validação de suas etapas, bem como identificar possíveis alterações para torná-la mais genérica e abrangente.

REFERÊNCIAS BIBLIOGRÁFICAS

ABOUELJINANE, L.; SAHIN, E.; JEMAI, Z.; MARTY, J. A simulation study to improve the performance of an emergency medical service: Application to the French Val-de-Marne department. *Simulation Modelling Practice and Theory*, v. 47, p. 46 – 59, 2014.

ADENSO-DÍAZ, B.; RODRÍGUEZ, F. A simple search heuristic for the MCLP: application to the location of ambulance bases in a rural region. *International Journal of Management Science*, v. 25, n. 2, p. 181-187, 1997.

ALIZADEH, F.; XIA, Y. Simulation metamodel estimation with penalized b-splines: a second-order cone programming approach. In: A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, (Eds.), *Proceedings of the 2014 Winter Simulation Conference*, pp. 4061-4062.

ALLAOUI, H.; ARTIBA, A. Integrating simulation and optimization to schedule a hybrid flow shop with maintenance constraints. *Computers & Industrial Engineering*, v. 47, p. 431-450, 2004.

ALLEN, A. B.; FALK, B. G.; HIGGS, E. R.; MCCRACKEN, J. G. A study of waiting time in an emergency department. *CMA Journal*, v. 109, p. 373 – 376, 1973.

ALMEHDAWE, E.; JEWKES, B.; HE, Q. M. A Markovian queueing model for ambulance offload delays. *European Journal of Operational Research*, v. 226, p. 602-614, 2013.

ALMOYNA, M. M.; NITSCHKE, C. A. S. Dos serviços de urgência e sua regulação. *Manual de regulação-médica dos serviços de atendimento médico de urgência*. Secretaria de Saúde do Estado de Santa Catarina, 1999.

ALTIOK, T.; MELAMED, B. *Simulation modeling and analysis with arena*. California: Elsevier, 2007. 440 p.

AMPONSAH, S. K.; AMOAKO, G.; DARKWAH, K. F.; AGYEMAN, E. Location of ambulance emergency medical service in the Kumasi metropolis, Ghana. *African Journal of Mathematics and Computer Science Research*, v. 41, p. 18-26, 2011.

ANGELIS, V.; FELICI, G.; IMPELLUSO, P. Integrating simulation and optimization in health care centre management. *European Journal of Operational Research*, v. 150, p. 101-114, 2003.

ARABANI, A. B.; FARAHANI, R. Z. Facility location dynamics: an overview of classifications and applications. *Computers & Industrial Engineering*, v. 62, p. 408-420, 2012.

ARINGHIERI, R. 2010. An Integrated DE and AB simulation model for EMS management. In: *Proceedings of the 2010 IEEE Workshop on Health Care Management*, pp. 1 – 6.

ARINGHIERI, R.; CARELLO, G.; MORALE, D. Ambulance location through optimization and simulation: the case of Milano urban area. In: ANNUAL CONFERENCE OF THE ITALIAN OPERATIONS RESEARCH SOCIETY OPTIMIZATION AND DECISION SCIENCES, 38., 2007, Genova. *Annals...* Genova: Università degli Studi di Milan, 2007. p. 29.

ARREOLA-RISA, A.; GIMÉNEZ-GARCÍA, V. M.; MARTÍNEZ-PARRA, J. L. Optimizing stochastic production-inventory systems: a heuristic based on simulation and regression analysis. *European Journal of Operational Research*, v. 213, p. 107-118, 2011.

AZADIVAR, F. A tutorial on simulation optimization. In: WINTER SIMULATION CONFERENCE, 24., 1992, Arlington. *Proceedings...* Arlington: 1992. p. 198-204.

BANKS, J.; CARSON II, J. S.; NELSON, B. L.; NICOL, D. M. *Discrete-event system simulation*. NJ: Pearson, 2005. 608 p.

BARTON, R. R. Simulation optimization using metamodels. In: Rossetti, M. D., Hill, R. R., Johansson, B., Dunkin, A., Ingalls, R. G. (Eds.), *Proceedings of the 2009 Winter Simulation Conference*, pp. 230-283.

BARTON, R. R.; MECKESHEIMER, M. *Handbook in OR & MS*. Elsevier, 2006, v. 13, 692 p.

BELL, C. E., ALLEN, D. Optimal Planning of an Emergency Ambulance Service. *Socio-Economic Planning Sciences*, v. 3, p. 95-101, 1969.

BERCHI, R.; FONTANA, G.; PAGLIOSA, A.; BONORA, R.; SESANA, G. A five steps methodology for an Ambulance Planning. In: IEEE WORKSHOP ON HEALTH CARE MANAGEMENT, 2010. *Proceedings...* pp. 1 – 5.

BILES, W. E.; KLEIJNEN, J. P. C.; NIEUWENHUYSE, I. V.; VAN BEERS, W. C. M. Kriging metamodeling in constrained simulation optimization: an explorative study. In: M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, (Eds.), *Proceedings of the 2009 Winter Simulation Conference*, pp. 355-362.

BILLHARDT, H.; LUJAK, M.; SÁNCHEZ-BRUNETE, V.; FERNÁNDEZ, A.; OSSOWSKI, S. Dynamic coordination of ambulances for emergency medical assistance services. *Knowledge Based Systems*, v. 70, p. 268-280, 2014.

BLACKWELL, T. H.; KAUFMAN, J. S. Response time effectiveness: comparison of response time and survival in an urban emergency medical services system. *Academic Emergency Medicine*, v. 9, n. 4, p. 288-295, 2002.

BLANNING, R. W. The construction and implementation of metamodels. *Simulation*, p. 177-184, 1975.

BRAILSFORD, S. C. Tutorial: advances and challenges in healthcare simulation modeling. In: WINTER SIMULATION CONFERENCE, 40., 2007, Washington. *Proceedings...* Washington: 2007. p. 1436-1448.

- BROTCORNE, L.; LAPORTE, G.; SEMET, F. Fast heuristics for large scale covering-location problems. *Computers & Operations Research*, v. 29, p. 651-665, 2002.
- BROTCORNE, L.; LAPORTE, G.; SEMET, F. Ambulance location and relocation models. *European Journal of Operational Research* 147, p. 451-463, 2003.
- BUDGE, S.; INGOLFSSON, A.; ZEROM, D. Empirical analysis of ambulance travel times: the case of Calgary emergency medical services. *Management Science*, v. 56, n. 4, p. 716-723, 2010.
- CARSON, Y. Simulation optimization: Methods and applications. In Andraróttir S., Healy K.J., Withers D.H., Nelson B.L. *Proceedings of the 1997 Winter Simulation Conference*. USA, pp. 118-126, 1997.
- CHAVEESUK, R.; SMITH, A. E. Dual Kriging: an exploratory use in economic metamodeling. *The Engineering Economist*, v. 50, p. 247-271, 2005.
- CHI, M. A more accurate second-order polynomial metamodel using a pseudo-random number assignment strategy. *Journal of the Operational Research Society*, v. 64, p. 198-207, 2013.
- CHURCH, R.; REVELLE, C. The maximal covering location problem. *Papers of the Regional Science Association*, v. 32, p. 101-118, 1974.
- CHWIF, L.; MEDINA, A. C. *Modelagem e simulação de eventos discretos: teoria e aplicações*. São Paulo: Editora do Autor, 2010. 254 p.
- CIMELLARO, G. P.; REINHORN, A. M.; BRUNEAU, M. Performance-based metamodels for healthcare facilities. *Earthquake Engineering & Structural Dynamics*, v. 40, n. 11, p. 1197-1217, 2011.
- DASKIN, M. S. A maximum expected coverage location model: formulation, properties and heuristic solution. *Transportation Science*, v. 17, p. 48-70, 1983.
- DELLINO, G.; MELONI, C. Metamodelo variability in simulation-optimization: a bootstrap analysis. Siena: Departamento de Engenharia de Informação – Universidade de Siena, 2011. 28 p. (Relatório técnico, DII n02)
- DEREKENARIS, G.; GAROFALAKIS, J.; MAKRIS, C.; PRENTZAS, J.; SIOUTAS, S.; TSAKALIDIS, A. Integrating GIS, GPS and GSM technologies for the effective management of ambulances. *Computers, Environment and Urban Systems*, v. 25, p. 267-278, 2001.
- DOERNER, K. F.; GUTJAHR, W. J.; HARTL, R. F.; KARALL, M.; REIMMAN, M. Heuristic solution of an extended double-coverage ambulance location problem for Austria. *Central European Journal of Operations Research*, v. 13, n. 4, p. 325, 2005.
- DOS SANTOS, M. I. C. S. R. *Construção de metamodelos de regressão não linear para simulação de acontecimentos discretos*. 2002. 177 f. Tese (Doutorado em Engenharia e Gestão Industrial) – Instituto Superior Técnico da Universidade Técnica de Lisboa, 2002.

DOS SANTOS, P. M. R.; DOS SANTOS, M. I. R. Using subsystem linear regression metamodels in stochastic simulation. *European Journal of Operational Research*, v. 196, n. 3, p. 1031-1040, 2009.

DURIEUX, S.; PIERREVAL, H. Regression metamodeling for the design of automated manufacturing system composed of parallel machines sharing a material handling resource. *International Journal of Production Economics*, v. 89, p. 21-30, 2004

FANG, K.T.; LI, R.; SUDJIANTO, A. *Design and Modeling for Computer Experiments*. Boca Raton, FL: Computer Science and Data Analysis Series. Chapman & Hall/CRC, 2006, p. 281.

FIGUEIREDO, L. H.; CARVALHO, P. C. P. *Introdução à Geometria Computacional*. Rio de Janeiro; Instituto de Matemática Pura e Aplicada PUC-Rio, 1991. 111 p.

FIGUEIREDO, A. P. S.; LORENA, L. A. N.; CARVALHO, S. V. Localização de Ambulâncias pelo Modelo TEAM – Solução Através do Algoritmo Genético Construtivo. *IV Worcap – Workshop dos Cursos de Computação do INPE*, 2004. Disponível na biblioteca digital URLib <lac.inpe.br/worcap/2004/10.06.15.47>.

FISHWICK, P. A. Neural network in simulation: a comparison with traditional modeling approaches. *Proceedings of the 1989 Winter Simulation Conference*, Washington DC, p. 702-710, 1989.

FITZSIMMONS, J. A. 1971. An emergency medical system simulation model. In: *Proceedings of the 1971 Winter Simulation Conference Proceedings*, pp. 18 – 25.

FITZSIMMONS, J. A.; FITZSIMMONS, M. J. *Administração de serviços: operações, estratégia e tecnologia da informação*. 4a ed. Porto Alegre: Artmed Editora S. A., 2004. 564 p.

FONE, D.; HOLLINGHURST, S.; TEMPLE, M.; ROUND, A.; LESTER, N.; WEIGHTMAN, A.; ROBERTS, K.; COYLE, E.; BEVAN, G.; PALMER, S. Systematic review of the use and value of computer simulation modeling in population health and health care delivery. *Journal of Public Health Medicine*, v. 25, n. 4, p. 325-335, Dec. 2003.

FONSECA, D. J.; NAVARESE, D. Artificial neural networks for job shop simulation. *Advanced Engineering Informatics*, v. 16, n. 4, p. 241-246, 2002.

FONSECA, D. J.; NAVARESE, D. Simulation through artificial neural networks. *Engineering Applications of Artificial Intelligence*, v. 16, p. 177-183, 2003.

FREITAS FILHO, P. J. *Introdução à modelagem e simulação de sistemas com aplicações em Arena*. Florianópolis/SC: Visual Book Editora, 2008. 372 p.

FU, M. C. Optimization for simulation: theory vs. practice. *INFORMS Journal on Computing*, v. 14, n. 3, p. 192-215, Jul. 2002.

FUJIWARA, O.; MAKJAMROEN, T.; GUPTA, K. K. Ambulance deployment analysis: a case study of Bangkok. *European Journal of Operational Research*, v. 31, p. 9-18, 1987.

GALDÁMEZ, E. V. C. *Aplicação das técnicas de planejamento e análise de experimentos na melhoria da qualidade de um processo de fabricação de produtos plásticos*. 2002. 133 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2002.

GARCIA, L. C. *Dimensionamento de recursos de atendimento móvel de urgência da região metropolitana II do estado do Rio de Janeiro*. 2006. 170 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

GENDREAU, M.; LAPORTE, G.; SEMET, F. Solving an ambulance location model by tabu search. *Location Science*, v. 5, n. 2, p. 75-88, 1997.

GENDREAU, M.; LAPORTE, G.; SEMET, F. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing*, v. 27, p. 1641-1653, 2001.

GLOVER, F. Future paths for integer programming and links to artificial intelligence. *Computers and Operations Research*, v. 13, n. 5, p. 533-549, 1986.

GOLDBERG, J. Operations research models for the deployment of emergency service vehicles. *EMS Management Journal*, v. 1, p. 20-39, 2004.

GOLDBERG, J.; DIETRICH, R.; CHEN, J. M.; MITWASI, M. A simulation model of evaluating a set of emergency vehicle base locations: development, validation and usage. *Socio-Econ. Plan. Sci.*, v. 24, n. 2, p. 125-141, 1990.

GONÇALVES, M. B. Métodos de pesquisa operacional em serviços emergenciais. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 26., 1994, Florianópolis. *Anais...* Florianópolis: 1994. p. 597-601.

GUNAL, M. M.; PIDD, M. Discrete event simulation for performance modeling in health care: a review of the literature. *Journal of Simulation*, v. 4, n. 1, p. 42-51, 2010.

HAMAD, H. Validation of metamodels in simulation: a new metric. *Engineering with Computers*, n. 27, p. 309-317, 2011.

HARRELL, F. E. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. New York: Springer, 2001, 572 p.

HAUSNER, J. Determining the travel characteristics of emergency service vehicles. The New York City-Rand Institute, R-1687/HUD/NYC, Abril, 1975.

HENDERSON, S. G., MASON, A. J., 1999. Estimating Ambulance Requirements in Auckland, New Zealand. In: Farrington, P. A., Nembhard, H. B., Sturrock, D. T., Evans, G. W. (Eds.), Proceedings of the 1999 Winter Simulation Conference, pp. 1670 – 1674.

HILL, E. D.; HILL, J. L.; JACOBS, L. M. Planning for emergency ambulance service systems. *The Journal of Emergency Medicine*, v. 1, p. 331-338, 1984.

HUANG, B.; PAN, X. GIS coupled with traffic simulation and optimization for incident response. *Computers, Environment and Urban Systems*, v. 31, p. 116-132, 2007.

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, Jan. 2006.

IANNONI, A. P.; MORABITO, R. Otimização da localização das bases de ambulâncias e do dimensionamento de suas regiões de cobertura em rodovias. *Produção*, v. 18, n. 1, p. 47-63, 2008.

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

IANNONNI, A. P.; MORABITO, R.; SAYDAM, C. Optimizing large-scale emergency medical system operations on highways using the hypercube queuing model. *Socio-Economic Planning Sciences*, v. 45, p. 105-117, 2011.

IOOSS, B., BOUSSOUF, L., FEUILLARD, V., AND MARREL, A. Numerical studies of the metamodel fitting and validation processes. *International Journal of Advances in Systems and Measurements*, n. 3, p. 11-21, 2010.

JALAL, H.; DOWD, B.; SAINFORT, F.; KUNTZ, K. M. Linear regression metamodeling as a tool to summarize and present simulation model results. *Medicine Decision Making*, v. 33, n. 7, p. 880-890, 2013.

JI, Y.; KIM, S. An adaptive radial basis function method using weighted improvement. In: R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, (Eds.), *Proceedings of the 2013 Winter Simulation Conference*, pp. 957-968.

JIN, R.; CHEN, W.; SIMPSON, T.W., 2000. Comparative studies of metamodeling techniques under multiple modeling criteria. In: *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA. AIAA, St. Louis, MO, 2000.

JIN, R.; CHEN, W.; SIMPSON, T.W. Comparative studies of metamodeling techniques under multiple modeling criteria. *Structural and Multidisciplinary Optimization*, v. 23, n. 1, p. 1-13, 2001.

KARAMAN, M. *A genetic algorithm for the multi-level maximal covering ambulance location problem*. 2008. 116 f. Thesis (Master of Science in Industrial Engineering) – Graduate School of Natural and Applied Sciences of Middle East Technical University.

KILMER, R. A.; SMITH, A. E.; SHUMAN, L. J. An emergency department simulation and a neural network metamodels. *Journal Society of Health Systems*, v. 5, n. 3, p. 63-79, 1997.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. *Science*, v. 220, n. 4598, p. 671-680, 1983.

KLEIJNEN, J. P. C. An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, v. 164, p. 287-300, 2005.

KLEIJNEN, J. P. C. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, v. 192, p. 707-716, 2009a.

KLEIJNEN, J. P. C. Factor screening in simulation experiments: review of sequential bifurcation. In: *Advancing the Frontiers in Simulation: a Festschrift in Honor of George S. Fishman, Alexopoulos, C., Goldsman, D., and Wilson, J.R. (Eds.)*. International Series in Operations Research & Management Science, Springer, New York, v. 133, 2009b, p. 153-167.

KLEIJNEN, J. P. C.; CHENG, R. C. H.; FEELDERS, A. J. Bootstrapping and validation of metamodels in simulation. In: D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, (Eds.), *Proceedings of the 1998 Winter Simulation Conference*, pp. 701-706.

KLEIJNEN, J. P. C.; VAN BEERS, W. C. M. Application-driven sequential designs for simulation experiments: Kriging metamodeling. *Journal of the Operational Research Society*, v. 55, p. 876-883, 2004.

KLEIJNEN, J. P. C.; VAN BEERS, W. C. M. Robustness of Kriging when interpolating in random simulation with heterogeneous variances: some experiments. *European Journal of Operational Research*, v. 165, p. 826-834, 2005.

KOCK, O., WEIGL, H., 2003. Modeling Ambulance Service of the Austrian Red Cross. In: Chick, S., Sánchez, P. J., Ferrin, D., Morrice, D. J. (Eds.) *Proceedings of the 2003 Winter Simulation Conference*, pp. 1701 – 1706.

KOLESAR, P.; BLUM, E. Square root laws for fire engine response distances. The New York City-Rand Institute, P-5091, October, 1973.

KOLESAR, P.; WALKER, W. E. An algorithm for the dynamic relocation of fire companies. *Operations Research*, v. 22, p. 249-274, 1974.

KUJLIS, J., PAUL, R. J.; STERGILOULAS, L. K., 2007. Can health care benefit from modeling and simulation methods in the same way as business and manufacturing has? In: Henderson, S. G., Biller, B., Hsieh, M. –H, Shortle, J., Tew, J. D., Barton, R. R. (Eds.), *Proceedings of the 2007 Winter Simulation Conference*, pp. 1449 – 1453.

LARSON, R. C. A hypercube queueing model for facility location and redistricting in urban emergency services. *Computers and Operations Research*, v. 1, p. 67-95, 1974.

LAW, A. M. *Simulation modeling and analysis*. 4. ed. New York: McGraw-Hill, 2007. 768 p.

LEE, S. The role of centrality in ambulance dispatching. *Decision Support Systems*, v. 54, n. 1, p. 282-291, 2012.

LEI, Y.; SONG, L.; WANG, W.; JIANG, C. A metamodel-based representation method for reusable simulation model. In: S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton. (Eds.), *Proceedings of the 2007 Winter Simulation Conference*, pp. 851-858.

LI, Y. F.; NG, S. H.; XIE, M.; GOH, T. N. A systematic comparison of metamodeling techniques for simulation optimization in Decision Support Systems. *Applied Soft Computing*, v. 10, p. 1257-1273, 2010.

LIM, C. S.; MAMAT, R. BRAUNL, T. Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems*, v. 12, n. 2, p. 624 – 632, 2011.

LIN, C. H.; KAO, C. Y.; HUANG, C. Y. Managing emergency department overcrowding via ambulance diversion: a discrete event simulation model. *Journal of the Formosan Medical Association*, v. 22, p. 1-8, 2012.

LOPES, S. L. B.; FERNANDES, R. J. Uma breve revisão do atendimento médico pré-hospitalar. *Revista do Hospital das Clínicas e da Faculdade de Medicina de Ribeirão Preto da USP*, v. 32, n. 4, p. 381-387, 1999.

LOWERY, J. C. Introduction to simulation in health care. In: WINTER SIMULATION CONFERENCE, 28., 1996, Coronado. *Proceedings...* Coronado: 1996. p. 78-84.

LUBICZ, M.; MIELCZAREK, B. Simulation modeling of emergency medical services. *European Journal of Operational Research*, v. 29, p. 178-185, 1987.

MALEKI, M.; MAJLESINASAB, N.; SEPEHRI, M. M. Two new models for redeployment of ambulances. *Computers and Industrial Engineering*, v. 78, p. 271-284, 2014.

MATTESON, D. S.; MCLEAN, M. W.; WOODARD, D. B.; HENDERSON, S. G. Forecasting emergency medical service call arrival rates. *The Annals of Applied Statistics*, v. 5, n. 2, p. 1379 – 1406, 2011.

MAYER, J. D. Paramedic response time and survival from cardiac arrest. *Social Science and Medicine*, v. 13, n. 4, p. 267-271, Dec. 1979.

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

MCALFEE, W. E.; NAQVI, I. A. The relocation of ambulance stations: a successful case study. *European Journal of Operational Research*, v. 75, p. 582-588, 1994.

McKAY, M.D.; BECKMAN, R.J.; CONOVER, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, v. 21, p. 239-245, 1979.

MECKESHEIMER, M. *A framework for metamodel-based design: subsystem metamodel assessment and implementation issues*. 2001. 285 f. Tese (Doutorado em Engenharia Industrial) - Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, 2001.

MECKESHEIMER, M.; BOOKER, A. J.; BARTON, R. R.; SIMPSON, T. W. Computationally inexpensive metamodel assessment strategies. *AIAA Journal*, v. 40, p. 2053–2060, 2002.

MENDONÇA, F. C.; MORABITO, R. Aplicação do modelo hipercubo para análise de um sistema médico-emergencial em rodovia. *Gestão & Produção*, v. 7, n. 1, p. 73-91, Abr. 2000.

MERKURYEVA, G. Response surface-based simulation metamodeling methods. *Applied Optimization*, v. 94, p. 205-215, 2005.

MINISTÉRIO DA SAÚDE. *Portaria nº 1.864 de 29 de setembro de 2003*. Disponível em http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2003/prt1864_29_09_2003.html, acessado em 25 de janeiro de 2012.

MOROHOSI, H.; FURUTA, T. Optimization model and simulation for improving ambulance service system. 11th International Symposium on Operations Research and its Applications in Engineering, Technology and Management (2013), p. 23-25.

NAFARRATE, A. D.; FOWLER, J. W.; WU, T. Design of centralized ambulance diversion policies using simulation-optimization. In: S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, (Eds), *Proceedings of the 2011 Winter Simulation Conference*, pp. 1251 – 1262.

NAOUM-SAWAYA, J.; ELHEDHLI, S. A stochastic optimization model for real-time ambulance redeployment. *Computers & Operations Research*, v. 40, p. 1972-1978, 2013.

National Health System (NHS): London Ambulance Service, 2012, www.londonambulance.nhs.uk/about-us/how-are-we-doing/meeting-our-targets.aspx, acessado em 16 de janeiro de 2012.

NEDDERMEIJER, H. G.; OORTMARSSSEN, G. J. V. A framework for response surface methodology for simulation optimization. In: J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., *Proceeding of the 2000 Winter Simulation Conference*, p. 129-136.

NOGUEIRA JÚNIOR, L. C. *Um estudo para redução do tempo de resposta do SAMU de Belo Horizonte através da realocação das bases de operação*. 2011. 87 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia da Universidade Federal de Minas Gerais, 2011.

NOVAES, A. G. *Sistemas logísticos: transporte, armazenagem e distribuição física de produtos*. São Paulo: Editora Edgard Blucher, 1989. 372 p.

O'KEEFFE, C.; NICHOLL, J.; TURNER, J.; GOODACRE, S. Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. *Emergency Medical Journal*, v. 28, p. 703-706, 2011.

ORGANIZACIÓN PANAMERICANA DE LA SALUD (OPAS), 2009. Información y Análisis de Salud: Situación de Salud en las Américas: Indicadores Básicos 2009. Washington, D.C., 2009.

OSBORNE, A.; TAYLOR, L.; REUBER, M.; GRUNEWALD, R. A.; PARKINSON, M.; DICKSON, J. M. Pre-hospital care after a seizure: evidence base and United Kingdom management guidelines. *Seizure*, v. 14, 2014.

PHAM, J. C.; PATEL, R.; MILLIN, M. G.; KIRSCH, T. D.; CHANMUNGAN, A. The effects of ambulance diversion: A comprehensive review. *Academic Emergency Medicine*, v. 13, n. 11, p. 1220 – 1227, 2006.

PIDD, M. *Computer simulation in management science*. Chichester: John Wiley & Sons Ltd, 1998, 279 p.

PINTO, L. R.; SILVA, P. M. S.; YOUNG, T. P. A generic method to develop simulation models for ambulance systems. *Simulation Modelling Practice and Theory*, v. 51, p. 170-183, 2015.

PISTONE, G; VICARIO, G. Comparing and generating Latin Hypercube designs in Kriging models. *ASIA Adv Stat Anal*, v. 94, p. 353-366, 2010.

PONS, P. T.; HAUKOOS, J. S.; BLUDWORTH, W.; CRIBLEY, T.; PONS, K. A.; MARKOVCHICK, V. J. Paramedic response time: does it affect patient survival? *Acad. Emerg. Med.*, v. 12, n. 7, p. 594-600, 2005.

RAJAGOPALAN, H. K.; SAYDAM, C. A minimum expected response model: formulation, heuristic solution and application. *Socio-Economic Planning Sciences*, v. 43, p. 253-262, 2009.

RAJAGOPALAN, H. K.; VERGARA, F. E.; SAYDAM, C.; XIAO, J. Developing effective meta-heuristics for a probabilistic location model via experimental design. *European Journal of Operational Research*, v. 177, p. 83-101, 2007.

RAJAGOPALAN, H. K.; SAYDAM, C.; XIAO, J. A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research*, v. 25, p. 814-826, 2008.

RAMIREZ-NAFARRATE, A.; HAFIZOGLU, A. B.; GEL, E. S.; FOWLER, J. W. Optimal control policies for ambulance diversion. *European Journal of Operational Research*, v. 236, p. 298-312, 2014.

RAMIREZ-NAFARRATE, A.; HAFIZOGLU, A. B. Comparison of ambulance diversion policies via simulation. In: C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher. (Eds.), Proceedings of the 2012 Winter Simulation Conference.

- RARDIN, R. L.; UZSOY, R. Experimental evaluation of heuristic optimization algorithms: a tutorial. *Journal of Heuristics*, v. 7, p. 261-304, 2001.
- RAUNER, M. S.; SCHAFFHAUSER-LINZATTI, M. M.; NIESSNER, H. Resource planning for ambulance services in mass casualty incidents: a DES-based policy model. *Health Care Management Science*, v. 15, n. 3, p. 254-269, 2012.
- REPEDE, J. F.; BERNARDO, J. J. Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky. *European Journal of Operational Research*, v. 75, p. 567-581, 1994.
- REVELLE, C. S.; EISELT, H. A. Location analysis: a synthesis and survey. *European Journal of Operational Research*, v. 165, p. 1-19, 2005.
- REVELLE, C. S.; HOGAN, K. The maximum availability location problem. *Transportation Science*, v. 23, p. 192-200, 1989.
- ROBINSON, S. Simulation model verification and validation: increasing the users' confidence. In: S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson (Eds), Proceedings of the 1997 Winter Simulation Conference, pp. 53 – 59.
- SAJJADIAN, S. *Dynamic model for real-time ambulance relocations based on coverage variation*. 2009. 67 f. Thesis (Master of Science in the School of Computing Science) – Sharif University of Technology, Tehran, Iran, 2009.
- SALAS, J.D.; SHIN, H.S. Uncertainty analysis of reservoir sedimentation. *Journal of Hydraulic Engineering*, v. 125, n. 4, p. 339-350, 1999.
- SÁNCHEZ-MANGAS, R.; GARCÍA-FERRER, A.; JUAN, A.; ARROYO, A. M. The probability of death in road traffic accidents. How important is a quick medical response? *Accident Analysis and Prevention*, v. 42, p. 1048-1056, 2010.
- SARGENT, R. G. Verification and validation of simulation models. In: S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Bartons (Eds.), Proceedings of the 2007 Winter Simulation Conference, pp. 124 – 137.
- SARTORI, D. E.; SMITH, A. E. A metamodel approach to sensitivity analysis of capital project evaluation. *The Engineering Economist*, v. 43, n. 1, 1997.
- SASAKI, S.; COMBER, A. J.; SUZUKI, H.; BRUNSDON, C. Using genetic algorithms for optimize current and future health planning – the example of ambulance locations. *International Journal of Health Geographics*, v. 9, n. 4, 2010.
- SAVAS, E. S. Simulation and Cost-effectiveness Analysis of a New York's Emergency Ambulance Service. *Management Science*, v. 15, n. 12, p. 608-627, 1969.
- SCHMID, V. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research*, v. 219, n. 3, p. 611-621, 2011.

SCHMID, V.; DOERNER, K. F. Ambulance location and relocation problems with time-dependent travel times. *European Journal of Operational Research*, v. 207, p. 1293-1303, 2010.

SIDDHARTHAN, K.; JONES, W. J.; JOHNSON, J. A. A priority queuing model to reduce waiting times in emergency care. *International Journal of Health Care Quality Assurance*, v. 9, n. 5, p. 10 – 16, 1996.

SILVA, P. M. S. *Análise do serviço de atendimento móvel de urgência (SAMU) de Belo Horizonte via simulação e otimização*. 2010. 134 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia da Universidade Federal de Minas Gerais, 2010.

SILVA, P. M. S., PINTO, L. R., 2010. Emergency Medical System Analysis by Simulation and Optimization. In: Johansson, B., Jain, S., Montoya-Torres, J.,vHugan, J., Yücesan, E. (Eds.), *Proceedings of the 2010 Winter Simulation Conference*, pp. 2422 – 2432.

SINGER, M.; DONOSO, P. Assessing an ambulance service with queuing theory. *Computers & Operations Research*, v. 35, p. 2549 – 2560, 2008.

SLADJANA, A.; GORDANA, P.; ANA, S. Emergency response time after out-of-hospital cardiac arrest. *European Journal of Internal Medicine*, v. 22, p. 386-393, 2011.

SORENSEN, P.; CHURCH, R. Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences*, v. 44, p. 8-18, 2010.

STANDRIDGE, C. R. A tutorial on simulation in health care: applications and issues. In: WINTER SIMULATION CONFERENCE, 31., 1999, Phoenix. *Proceedings...* Phoenix: 1999. p. 49-55.

SU, S. LIANG, S. C. Resource reallocation in an emergency medical service system using computer simulation. *American Journal of Emergency Medicine*, v. 20, n. 7, p. 627-634, 2002.

SU, S., SHIH, C. L., Modeling an Emergency Medical Services System using Computer Simulation. *International Journal of Medical Informatics*, v. 72, p. 57 – 72, 2003.

SWOVELAND, C.; UYENO, D.; VERTINSKY, I.; VICKSON, R. A simulation-based methodology for optimization of ambulance service policies. *Socio-Econ. Plan. Sci.*, v. 7, p. 697-703, 1973.

SWOVELAND, C.; UYENO, D.; VERTINSKY, I.; VICKSON, R. Ambulance location: A probabilistic enumeration approach. *Management Science*, v. 20, p. 686-698, 1973.

TAKEDA, R. A. *Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde*. 2000. 210 f. Tese (Doutorado em Transportes) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, 2000.

- TAKEDA, R. A.; WIDMER, J. A.; MORABITO, R. Aplicação do modelo hipercubo de filas para avaliar a descentralização de ambulâncias em um sistema urbano de atendimento médico de urgência. *Pesquisa Operacional*, v. 24, n. 1, p. 39-71, 2004.
- TAKEDA, R. A.; WIDMER, J. A.; MORABITO, R. Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. *Computers & Operations Research*, v. 34, p. 727-741, 2007.
- TESTI, A.; TANFANI, E.; IVALDI, E.; CARELLO, G.; ARINGHERI, R.; FRAGNELLI, V. Operations research for patient - centered health care delivery. Paris: Franco Angeli, 2010. 432 p.
- TIMERMAN, S.; GONZALEZ, M. M. C.; ZARONI, A. C.; RAMIRES, J. A. F. Emergency medical services: Brazil. *Resuscitation*, v. 70, n. 3, p. 356-359, Sep. 2006.
- TOREGAS, C.; SWAIN, R.; REVELLE, C.; BERGMAN, L. The location of emergency service facilities. *Operations Research*, v. 19, n. 6, p. 1363-1373, Oct. 1971.
- VINCENT, S. Input data analysis. In: *Handbook of Simulation*, ed. J. Banks, p. 55-91, New York: John Wiley and Sons, 1998.
- VUKMIR, R. B. Survival from pre-hospital cardiac arrest is critically dependent upon response time. *Resuscitation*, v. 69, p. 229-234, 2006.
- ZAREKIFAR, M.; BILES, W. E.; EVANS, G. W. Kriging metamodeling in multi-objective simulation optimization. In: M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, (Eds.), *Proceedings of the 2009 Winter Simulation Conference*, pp. 2115-2122.
- ZAYTOON, J.; VILLERMAN-LECOLIER, G. A discrete events meta simulation model. *Advanced Factory Automation, Conference Publication n. 398*, 1994.
- ZHANG, H.; LI, H. Simulation-based optimization for dynamic resource allocation. *Automation in Construction*, v. 13, p. 409-420, 2004.
- ZHEN, L.; WANG, K.; HU, H.; DAOFANG, D. A simulation optimization framework for ambulance deployment and relocation problems. *Computers & Industrial Engineering*, v. 72, p. 12-23, 2014.
- WILDE, E. T. Do emergency medical system response times matter for health outcomes? *Health Economics*, v. 22, n. 7, p. 790-806, 2013.
- WINCH, J. K.; MADU, C. N.; KUEI, C. H. Metamodeling and optimizing a reverse logistics system. *International Journal of Information and Management Sciences*, v. 23, p. 41-58, 2012.
- WU, C. H., HWANG, K. P. Using a Discrete-event Simulation to Balance Ambulance Availability and Demand in Static Deployment Systems. *Academic Emergency Medicine*, v. 16, n. 12, p. 1359 – 1366, 2009.

YI, P.; GEORGE, S. K.; PAUL, J. A. LIN, L. Hospital capacity planning for disaster emergency management. *Socio-Economic Planning Sciences*, v. 44, p. 151-160, 2010.

YOUNG, T.; EATOCK, J. M.; NASEER, A.; LILFORD, R. 2009. Three critical challenges for modelling and simulation in healthcare. In: Rossetti, M. D., Hill, R. R., Johansson, B., Dunkin, A., Ingalls, R. G. (Eds.), *Proceedings of the 2009 Winter Simulation Conference*, pp. 1823-1830.