

**TÍTULO:** Uma estrutura computacional de otimização da simulação baseada em Simulated Annealing para avaliar desempenho de Sistemas Médicos de Emergência: O caso do SAMU das cidades de Ouro Preto e Mariana

**TITLE:** A computational structure for simulation optimization based on Simulated Annealing to evaluate the performance of Emergency Medical Systems: The case of SAMU in the cities of Ouro Preto and Mariana.

**Versão do autor aceita publicada online:** 03 out. 2022

**Publicado online:** 13 out. 2022

**Como citar esse artigo - American Psychological Association (APA):** Morais, T. R. M., Gomes Júnior, A. de C., & Rodrigues, L. F. (2022). Uma estrutura computacional de otimização da simulação baseada em Simulated Annealing para avaliar desempenho de Sistemas Médicos de Emergência: O caso do SAMU das cidades de Ouro Preto e Mariana. *Exacta*. DOI: <https://doi.org/10.5585/exactaep.2022.21836>.

**Thaise Regina Matos de Morais**

[thaise.morais@aluno.ufop.edu.br](mailto:thaise.morais@aluno.ufop.edu.br)

<https://orcid.org/0000-0002-9819-547X>

Universidade Federal de Ouro Preto

Mestre em Engenharia de Produção pela Universidade Federal de Ouro. Possui Pós graduação em Engenharia de Produção pela Pontifícia Universidade Católica de Minas Gerais (PUC MINAS - 2018) e também graduação em Engenharia de Produção Universidade Presidente Antônio Carlos (2016). Atualmente atua como coordenadora de processos.

**Aloísio de Castro Gomes Júnior**

[aloisio.junior@ufop.edu.br](mailto:aloisio.junior@ufop.edu.br)

<https://orcid.org/0000-0003-1206-1209>

Universidade Federal de Ouro Preto

Resumo da Biografia Doutor e Mestre em Engenharia de Produção pela UFMG. Atua desde 2006 no ensino de Engenharia de Produção, com ênfase na área de Pesquisa Operacional, tanto de graduação como de pós-graduação. É professor adjunto da UFOP.

**Lasara Fabricia Rodrigues**

[lasara@dep.ufmg.br](mailto:lasara@dep.ufmg.br)

<https://orcid.org/0000-0001-5427-7919>

Universidade Federal de Minas Gerais

Possui graduação em Engenharia de Produção pela Universidade Federal de Ouro Preto (2004), mestrado em Engenharia de Produção pela Universidade Federal de Minas Gerais (2006) e doutorado em Engenharia de Produção pela Universidade Federal de São Carlos (2014). Atualmente, é professora do Departamento de Engenharia de Produção (DEP) da Universidade Federal de Minas Gerais (UFMG) e dos Programas de Pós-Graduação em Engenharia de Produção da Universidade Federal de Minas Gerais (PPGEP-UFMG) e da Universidade Federal de Ouro Preto (PPGEP-UFOP). Tem experiência na área de Engenharia

de Produção e suas principais linhas de pesquisa são voltadas para as áreas de Teoria de Filas, Simulação, Modelos de Localização Probabilística, Análise Envoltória de Dados e aplicações de Pesquisa Operacional em sistemas logísticos, de manufatura e de saúde.

## RESUMO

O tempo de resposta de um Sistema Médico de Emergência (SME) é uma métrica preponderante de eficiência, visto que prestar atendimento rápido as vítimas de urgência, determina a minimização de sequelas permanentes ao mesmo tempo que maximiza a taxa de sobrevivência do paciente. Neste artigo, propomos um modelo de simulação via otimização, desenvolvido em linguagem Python, capaz de avaliar o desempenho de SME's. Aplicamos ao método proposto, os dados reais de um SME brasileiro e verificamos, a partir dos resultados obtidos, que configurações estratégicas resultariam na redução de aproximadamente 10% no tempo de resposta médio. Além disso, foi verificada a importância de se considerar outras variáveis de forma conjunta ao número de habitantes, na determinação do número de ambulâncias necessário para se atender as demandas de emergência no serviço pré-hospitalar.

**Palavras-chave:** Otimização da Simulação. *Simulated Annealing*. Serviços de emergência médica. Tempo de resposta. SAMU.

## ABSTRACT

The response time of an Emergency Medical System (EMS) is a preponderant metric of efficiency, since providing fast assistance to emergency victims determines the minimization of permanent sequelae while maximizing the patient's survival rate. In this article, we propose a simulation model via optimization, developed in Python language, capable of evaluating the performance of SME's. We applied real data from a Brazilian SME to the proposed method and verified, from the results obtained, which strategic configurations would result in a reduction of approximately 10% in the average response time. In addition, the importance of considering other variables together with the number of inhabitants was verified, in determining the number of ambulances necessary to meet the emergency demands in the pre-hospital service.

**Keywords:** Simulation Optimization. *Simulated Annealing*. Emergency Medical Services. Response time. SAMU.

## 1. INTRODUÇÃO

Os serviços pré-hospitalares têm por objetivo responder eficientemente às demandas da população, fornecendo atendimento em situações de emergências médicas, com o intuito de minimizar sequelas e taxas de mortalidade, melhorando as chances de recuperação do paciente. Neste contexto está inserido o Serviço Médico de Emergência (SME) que busca prestar além da assistência médica rápida, no local de ocorrência, o transporte do paciente para uma unidade de saúde adequada.

O Serviço de Atendimento Móvel de Urgência (SAMU) é um serviço médico pré-hospitalar brasileiro, de responsabilidade governamental, que foi instituído como o componente pré-hospitalar da Política Nacional de Atenção às Urgências. Em seu modelo, o sistema busca prestar assistência a pacientes a partir de regulações médicas, que propõe o melhor atendimento, variando entre orientações por telefone ou o envio de uma unidade móvel de suporte básico ou avançado à vida.

O tempo de resposta (TR), que é caracterizado pelo tempo decorrido entre o recebimento da chamada até a chegada da primeira equipe de socorro ao local de ocorrência da emergência, é um aspecto crítico, sobremaneira em casos onde existe o risco de morte. Além disso, a resposta rápida do sistema a um evento de emergência é uma grande expectativa do público que demanda os serviços. Nesse sentido, o poder público e os gestores dos SME's são desafiados a aplicar os limitados recursos de forma a obter um sistema eficiente.

Diversos trabalhos da literatura médica relacionam o tempo de resposta à taxa de sobrevivência do paciente, sobretudo, em caso de vítimas de traumas ou parada cardiorrespiratória (Pell, 2001; Gonzalez, Cummings, Phelan, Mulekar, & Rodning, 2009; Byrne et al., 2019).

Ainda que os resultados de alguns trabalhos realizados contradigam essa relação (Pons & Markovchick, 2002; Blackwell & Kaufman, 2002), o tempo de resposta é dentre os fatores de avaliação de desempenho, o mais utilizado no processo de tomada de decisões.

Este estudo objetivou desenvolver um método de otimização da simulação capaz de analisar o desempenho de SME's. Na estrutura proposta, a metaheurística *Simulated Annealing* foi integrada ao modelo de simulação, visando além da análise de diferentes cenários, apresentar configurações cujo tempo de resposta médio do sistema fosse reduzido.

O modelo proposto foi desenvolvido em linguagem *Python* e apresentou grande flexibilidade e adaptabilidade na análise de diferentes cenários, apresentando também, facilidade de adaptação para análise de outros SME's com características semelhantes. Ademais, o método de solução aproximada integrado ao modelo, permitiu determinar

configurações estratégicas que melhoram o desempenho do sistema, no que tange ao tempo de resposta.

Para validar a estrutura desenvolvida, utilizamos os dados reais de um SME brasileiro, o SAMU das cidades de Ouro Preto e Mariana. Além de verificar a aplicabilidade e eficiência do método desenvolvido, apresentamos configurações estratégicas que reduziram o tempo de resposta deste sistema, em diferentes cenários em aproximadamente 10%.

Na seção 2, apresentamos a revisão da literatura, apresentado relevantes trabalhos que utilizaram de técnicas e métodos da Pesquisa Operacional para analisar e melhorar desempenho de SME's. Na seção 3, descrevemos a metodologia aplicada e a estrutura dos modelos computacionais desenvolvidos. Na seção 4, apresentamos o sistema real estudado e os resultados dos experimentos realizados. Por fim, as discussões dos resultados, as conclusões do estudo e sugestões de pesquisas futuras são apresentadas na seção 5.

## **2. REVISÃO DA LITERATURA**

As técnicas e metodologias da Pesquisa Operacional são amplamente utilizadas pela comunidade acadêmica na análise e solução dos mais variados problemas, como planejamento de produção (Da Silva, Frazzon, & Casarotto Filho, 2016); análise de capacidade produtiva e redução de lead time (Filho, Silva, & Júnior, 2019); gerenciamento de filas e dimensionamento de pessoal no setor varejista (Moraes & Silva, 2021) e no segmento de reciclagem de resíduos sólidos (Junges, Romio, Simonetto, Campos, & Bobsin, 2021); planejamento e gestão de suprimentos (Felix, Peixoto, & Edwiges, 2021).

Um sistema médico de emergência pode ser representado como um sistema de fila única com diversos servidores controlados por uma central. A aleatoriedade intrínseca à operação das atividades do sistema, dificultam a modelagem analítica, tornando atraente analisá-lo a partir da técnica de simulação. Nessa seção, destacamos alguns estudos desenvolvidos e relacionados a SME's, tal qual é o objeto desse estudo.

O aumento da taxa de sucesso, que é caracterizado pelo número de atendimentos realizados dentro de um limite de tempo, em relação ao total de atendimentos realizados, foi objetivado pelos estudos realizados por Aringhieri, Carello e Moralle (2013) e Aboueljinane et al., 2014. Os primeiros utilizaram a simulação de eventos discretos (SED) para analisar os resultados obtidos dos dois modelos analíticos propostos para avaliar o desempenho de um SME italiano. Obtiveram como resultado a maximização da taxa de sucesso do sistema, a partir de uma maior disponibilidade das ambulâncias obtidas a partir de uma nova política

de despacho. Aboueljinane, Sahin, Jemai e Marty (2014) desenvolveram um modelo com objetivo de melhorar um SME francês, avaliando duas métricas de cobertura. Os resultados demonstraram que a descentralização das bases e localização estratégica das bases, aumentariam a taxa de sucesso do sistema em aproximadamente 5%.

A análise da redução do tempo de resposta foi realizada por Nogueira, Pinto e Silva (2014) e Fukushima e Moriya (2021). Os primeiros verificaram que seria possível minimizar o tempo de resposta do SME brasileiro, a partir de um modelo matemático proposto. Diferentes cenários foram então submetidos a um modelo de simulação nos quais os resultados apontaram para uma redução do tempo de resposta do sistema de forma simultânea à redução dos custos totais do sistema, em virtude da diminuição do número total de bases instaladas. Por outro lado, Fukushima e Moriya (2021) buscaram analisar como as condições do tráfego e clima influenciavam o desempenho do sistema, no que se refere ao tempo de resposta e de transferência dos pacientes para uma unidade de saúde. Os resultados obtidos levaram os autores a concluir que o desempenho do sistema não é afetado pelo clima ou fatores temporais, mas que as condições das estradas (tráfego) impactam consideravelmente, concluindo-se pela relevância da utilização de sistemas GPS para previsões e determinações de rotas alternativas.

Bélanger, Lanzarone, Nicoletta, Ruiz e Soriano (2020) propuseram uma estrutura de otimização da simulação capaz de abordar o Problema de Localização e Despacho de ambulância (*Ambulance Location and Dispatching Problem – ALDP*). O objetivo do modelo proposto era avaliar a disponibilidade das ambulâncias e as medidas de desempenho do sistema, a partir de estimativas resultantes da simulação. Propuseram e compararam duas abordagens de aproximação de disponibilidade das ambulâncias, baseada na distribuição de Bernoulli e Teoria das filas. Os resultados dos testes aplicados demonstraram a eficiência do método para se obter menores tempos de respostas quando se integra as decisões de localização e despacho.

Zhen, Wang, Hu e Chang (2014), McComack e Coates (2015), Zaffar, Rajagopalan, Saydam, Mayorga e Sharer (2016) e Golabian, Arkat, Farughi e Tavakkoli-Moghaddam (2021) também utilizaram de uma estrutura de otimização da simulação para analisar o desempenho de SME's sob diferentes estratégias. Zhen, Wang, Hu e Chang (2014) buscaram solucionar o problema de implantação de um SME, incorporando a metaheurística Algoritmo Genético ao modelo de Simulação. Propuseram um modelo matemático de realocação de ambulâncias e submeteram a este método, os dados reais do SME de Xangai,

China, dividindo o período de um dia em cinco faixas horárias. Apresentaram como resultado, a configuração estratégica do sistema estudado, considerando fatores como tráfegos dependentes de condições temporais e espaciais. McComack e Coates (2015) realizaram um estudo no SME de Londres integrando ao modelo de simulação, a metaheurística Algoritmo Genético. Os resultados demonstram que a otimização da localização da frota, sem aumento de recursos, aumentava a taxa de sobrevivência de pacientes com parada cardíaca. Os autores apontaram também que a remoção estratégica de uma base apresenta pouco impacto na probabilidade de sobrevivência. Zaffar, Rajagopalan, Saydam, Mayorga e Sharer (2016) fizeram uso da otimização de simulação para propor um modelo de análise de desempenho de um SME sob diferentes métricas. Utilizaram a metaheurística Busca Tabu Reativa (RTS) para otimizar e comparar três diferentes objetivos: maximizar a cobertura, minimizar o tempo de resposta médio e maximizar a capacidade de sobrevivência. Os resultados obtidos no estudo, demonstraram que o objetivo de máxima sobrevivência é estatisticamente melhor quando comparado aos demais. Golabian, Arkat, Farughi e Tavakkoli-Moghaddam (2021) desenvolveram uma estrutura de otimização da simulação para solucionar o problema de localização estratégica de bases, utilizando a metaheurística *Simulated Annealing*. Os autores propuseram um modelo onde a ambulância despachada, não necessariamente retornaria para sua base de origem ao final do atendimento. Para tanto, duas estratégias de prioridade de retorno foram analisadas, objetivando maximizar a cobertura dentro de um tempo limite. Aplicaram ao modelo proposto, os dados de um estudo de caso realizado anteriormente, do SME do Condado de Hannover, obtendo-se melhores resultados e validando a eficiência do método na solução de problemas de grande porte.

### **3. METODOLOGIA DA OTIMIZAÇÃO DA SIMULAÇÃO**

A técnica de simulação apresenta como característica, a possibilidade de se alterar os parâmetros de um modelo e observar o desempenho do sistema analisado sob diferentes configurações das variáveis de entrada. Toda análise realizada a partir desta técnica, busca encontrar o conjunto de parâmetros que apresenta os melhores resultados.

A utilização de técnicas da otimização em conjunto com a simulação computacional, surge com o objetivo de apresentar uma boa solução para o problema, apesar de não garantir que esta seja a ótima.

Diversos métodos de soluções podem ser aplicados aos problemas de otimização. A

otimização da simulação (OS) também denominada de otimização via simulação (OvS) visa definir a melhor configuração do sistema, que sujeito às restrições e limitações de recursos alcança um determinado objetivo (Coelho, & Pinto, 2018). Em consonância, Carson e Maria (1997) definem otimização da simulação como o processo pela busca de melhores valores para as variáveis de entrada dentre todas as possibilidades, sem avaliar explicitamente cada possibilidade.

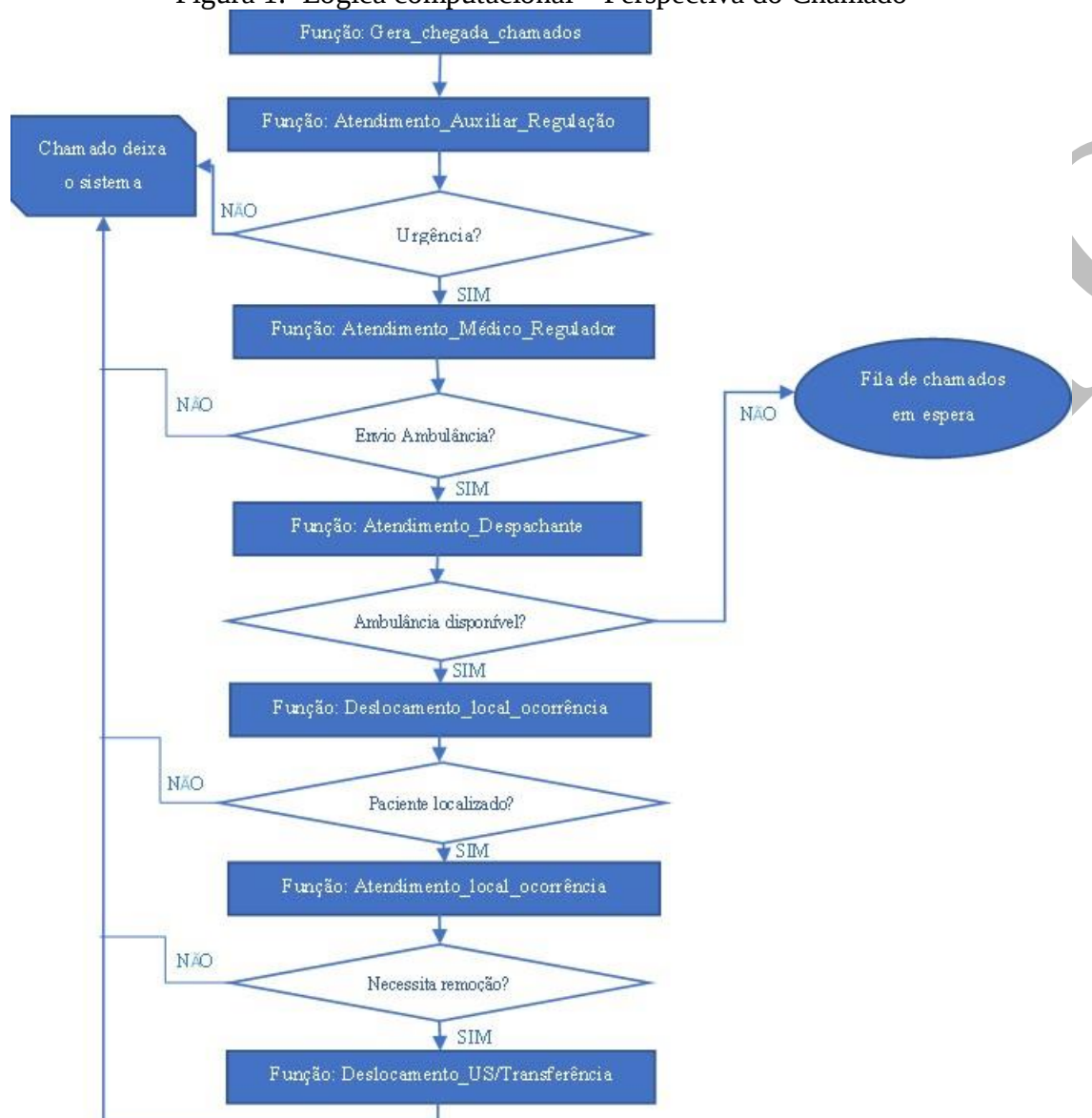
O problema de otimização é formulado a partir de uma função de avaliação que objetiva maximizar ou minimizar uma medida de desempenho, estando ou não sujeita a restrições que delimitam o valor que as variáveis podem assumir. Tal função busca medir a qualidade da solução. As restrições do problema podem ser também relaxadas e integradas à função de avaliação sob a perspectiva de penalidades de forma que as soluções que desrespeitem as restrições do problema não sejam consideradas como a melhor solução (Fisher, 2004). Assim, os parâmetros de entrada do modelo de simulação, correspondem às variáveis de decisão enquanto as saídas, modelam a função objetivo e as restrições do problema estudado.

### 3.1 O modelo computacional de simulação

O modelo computacional de simulação desenvolvido buscou representar fielmente a operação do sistema estudado. Foi implementando em linguagem *Python*, utilizando-se recursos da biblioteca *Simpy*, que é *framework* de simulação de eventos discretos estruturado essencialmente com funções geradoras que permite modelar as interações de entidades ativas e passivas de sistemas reais.

O modelo foi estruturado sob a perspectiva de duas entidades: o chamado, cuja representação esquemática do modelo de simulação é representada pela Figura 1 e ambulância, que tem sua lógica computacional representada pela Figura 2.

Figura 1: Lógica computacional – Perspectiva do Chamado

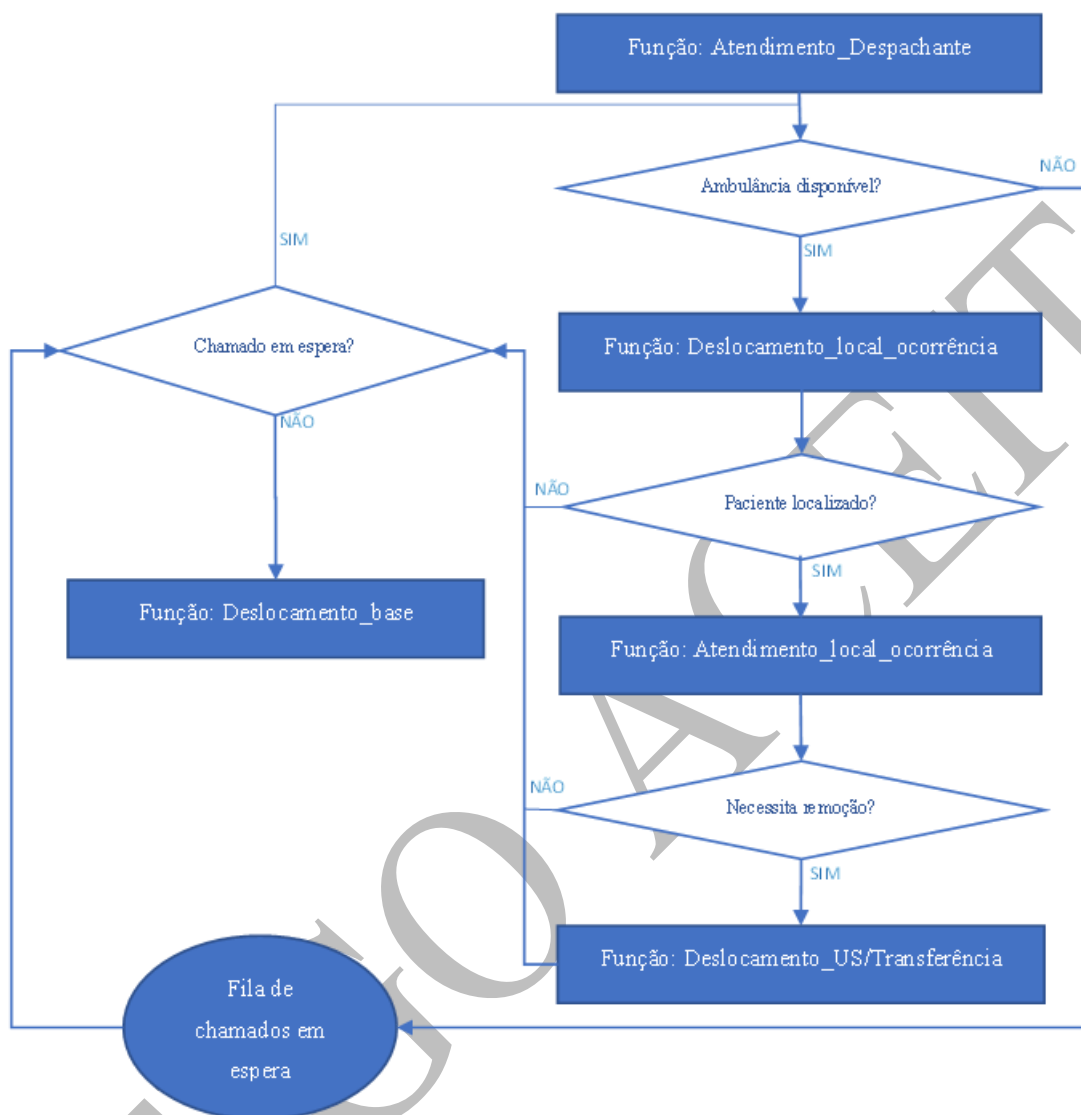


Fonte: Autores.

- Função “Gera\_Chegada\_Chamados”: responsável por gerar a chegada de entidades no sistema, aqui intitulada “Chamados” em acordo com a distribuição teórica representativa da taxa de chegadas. Essa função é também responsável por receber os atributos do chamado, tais como: faixa horária, localização, tipo (A ou B) gerados por funções auxiliares;
- Função “Atendimento\_Auxiliar\_Regulação”: responsável por gerar um número aleatório, segundo uma distribuição de probabilidade, que represente tempo de ocupação do recurso “Auxiliar de regulação” e “chamar” a função que representa o atendimento do médico regulador.

- c. Função “Atendimento\_Medico\_Regulador”: responsável por gerar um número, segundo uma distribuição de probabilidade, representativo do tempo de ocupação do recurso “Médico Regulador” e “chamar” a função que representa a atividade de despacho.
- d. Função “Atendimento\_Despachante”: responsável por gerar um número aleatório representativo do tempo de ocupação do recurso “Despachante”. Essa função é responsável por verificar se a ambulância indicada para realizar o atendimento, em acordo com a política de despacho, está disponível. Caso positivo, o servidor em questão é então requisitado e a função de deslocamento é acionada. Em caso de indisponibilidade, o chamado será alocado em uma “fila”;
- e. Função “Deslocamento\_local\_ocorrência”: Essa função, além de gerar um número aleatório que represente o tempo de deslocamento, é responsável também por calcular o tempo de resposta, que é representado pela soma das atividades de regulação, despacho, espera (se for o caso) e o tempo do deslocamento;
- f. Função “Atendimento\_local\_ocorrência”: função que gera aleatoriamente o número que representará o tempo de deslocamento e que verifica se ocorrerá o atendimento efetivo (em conformidade com a probabilidade do paciente ser localizado). Caso negativo, essa função retira o chamado do sistema, liberando o servidor para um novo atendimento. Em contrário, quantifica o tempo de realização a partir do número aleatório gerado e verifica se haverá a remoção do paciente para uma unidade de saúde (probabilisticamente). Em caso a remoção, a função “Deslocamento\_US\_Transferência” é acionada.
- g. Função “Deslocamento\_US\_Transferência: função que gera o número aleatório que representa o tempo de deslocamento entre o local de ocorrência e a unidade de saúde. Nesta mesma função é gerado o número aleatório que caracteriza o tempo de transferência do paciente para a unidade receptora bem como das atividades administrativas relacionadas. O ciclo do recurso “Ambulância” representado pela Figura 2, é descrito a seguir.

Figura 2: Lógica computacional – Perspectiva da ambulância



Fonte: Autores.

- A função “Atendimento Despachante” ocupa o recurso, caso esteja disponível, acionando a função que representa o deslocamento para o local de ocorrência;
- A função “Deslocamento\_local\_ocorrência atualiza a variável que contabiliza a ocupação do servidor com o valor que foi gerado aleatoriamente para representar o tempo de deslocamento entre a base e o local de origem do chamado;
- A função “Atendimento\_local\_ocorrência atualiza a variável de ocupação do recurso ou o libera, caso não ocorra o efetivo atendimento ou se, após esse, não ocorrer a remoção do paciente para uma unidade de saúde. Em caso de liberação, verifica-se se há chamados aguardando em fila de espera. Caso positivo a ambulância será despachada para esse atendimento. Neste cenário, a função

“Deslocamento\_local\_ocorrência” é acionada e caso contrário, a ambulância desloca-se para a sua base (local de espera) e a variável de ocupação do recurso é novamente atualizada.

- d. A função “Deslocamento\_US\_Transferência” ocupa o recurso e atualiza a sua respectiva variável de ocupação com os tempos referentes às atividades. Finalizados esses tempos, o recurso é liberado. Em seguida, verifica se há chamado em espera. Em cenário positivo, é “chamada” a função “Deslocamento\_local\_ocorrência” e em caso contrário, a ambulância desloca-se para a sua base e a variável de ocupação é atualizada.

As funções auxiliares do modelo, utilizada para se definir os atributos do atendimento foram:

- i. Sorteia\_Faixa\_Chegada: em acordo com a probabilidade de ocorrência, define a faixa horária de chegada do chamado;
- ii. Sorteia\_Átomo: definição da localização (origem) do chamado;
- iii. Sorteia\_hospital: função que define, em acordo com a probabilidade, para qual unidade de saúde o paciente será transferido;

O modelo desenvolvido foi validado a partir da comparação dos resultados obtidos das rodadas de simulação com os dados reais, além dos resultados da análise de sensibilidade.

### 3.2 Modelo computacional de otimização da simulação

Com o objetivo de avaliar cenários que demandavam um grande conjunto de variáveis de entradas, integrou-se ao modelo de simulação, uma estrutura de otimização baseada na metaheurística *Simulated Annealing*.

Esse método de solução aproximada foi proposto por Kirkpatrick, Gelatt e Vecchi, 1983 (1983). Trata-se de um método de otimização que se inspira numa analogia com a termodinâmica, conhecido como recozimento simulado. Este processo, amplamente utilizado no setor metalúrgico, busca elevar a temperatura dos metais à valores superiores a 1000°C e posteriormente resfriá-lo de forma lenta e controlada, no intuito de obter estruturas mais fortes e estáveis, com menor propensão a defeitos.

O *Simulated Annealing* busca melhorar a solução do problema a partir de movimentos condizentes a uma estrutura de vizinhança, em sucessivas iterações, até que um ótimo local seja encontrado. Como estratégia para escapar de ótimos locais, o método aceita-se soluções de piora, segundo uma probabilidade. A temperatura, inicialmente, assume valores altos.

Após determinado número de iterações, o valor da temperatura vai gradativamente diminuindo, segundo uma função ou taxa de resfriamento. Assim, à medida que a temperatura vai tendendo a zero, a probabilidade de se aceitar soluções de piora diminui. Neste cenário, o método se comporta de forma a só realizar o movimento, se a função de avaliação obtiver melhora.

O método é encerrado, retornando a melhor solução encontrada, quando o critério de parada, que pode ser definido em termos de tempo de execução, número máximo de iterações sem alcançar melhora na função de avaliação, ou ainda, um número máximo de iterações é alcançado.

O método de otimização da simulação desenvolvido, tal qual denominamos de SA\_SIMUL é representado pela Figura 3.

Figura 3: Pseudocódigo da estrutura implementada

---

```

algoritmo SA_Simul( $T_0, T_f, Sa_{max}, \alpha$ )
1  Gere uma solução inicial  $s$ 
2   $s^* \leftarrow s$ 
3   $f(s) \leftarrow tr(s)$ 
4   $f(s^*) \leftarrow tr(s)$ 
5   $Iter\_t \leftarrow 0$ 
6   $T \leftarrow T_0$ 
7  enquanto  $T \geq T_f$  faça
8      enquanto  $Iter\_t < Sa_{max}$  faça
9          gere um vizinho  $s'$  aleatoriamente ( $s' \in N(s)$ )
10         execute 5 replicações de simulação para  $s'$  e calcule  $tr$ 
11          $f(s') \leftarrow tr(s')$ 
12         calcule  $\Delta = f(s') - f(s)$ 
13         se ( $\Delta \leq 0$ ) então
14              $s \leftarrow s'$ 
15             se  $f(s') < f(s^*)$  então
16                  $s^* \leftarrow s'$ 
17                  $Iter\_t \leftarrow 0$ 
18         senão
19              $s \leftarrow s'$  com probabilidade  $e^{-\Delta/T}$ 
20              $Iter\_t \leftarrow Iter\_t + 1$ 
21         fim-enquanto
22          $T \leftarrow \alpha \times T$ 
23          $Iter\_t \leftarrow 0$ 
24     fim-enquanto
25     retorne  $s^*$ 
fim-algoritmo

```

---

Fonte: Autores.

O modelo recebe quatro parâmetros de entrada, sendo eles a temperatura inicial ( $T_0$ ), a temperatura final ( $T_f$ ), o número máximo de iterações em cada temperatura ( $SA_{max}$ ) e a taxa de resfriamento da temperatura ( $\alpha$ ).

Inicialmente, uma solução inicial, gerada aleatoriamente, é simulada e o valor da função de avaliação (tempo de resposta médio) é determinado. Essa solução será, no início do método, simultaneamente, a melhor solução encontrada ( $s^*$ ) e a solução corrente ( $s$ ).

Para cada temperatura, enquanto o número de iterações é inferior ao valor de  $SA_{max}$ , é gerada uma solução vizinha, com base na estrutura de vizinhança selecionada (linha 9).

A solução vizinha gerada determina as variáveis de entrada do modelo de simulação, que por sua vez, retorna com o valor da função de avaliação a partir de cinco replicações. Esse valor será utilizado para o cálculo de Delta ( $\Delta$ ):

- i. Caso este valor seja negativo, a solução vizinha avaliada torna-se a solução corrente. Se o valor retornado ( $f(s')$ ) for menor que o valor da melhor solução encontrada até o momento ( $f(s^*)$ ), a solução avaliada nesta iteração torna-se a melhor solução encontrada ( $s^*$ ) e a variável de controle do número de iterações  $c$  recebe o valor 0;
- ii. Caso o valor de delta seja positivo, avalia-se a probabilidade da solução vizinha ( $s'$ ) ser aceita como solução corrente ( $s$ ). Neste cenário, a variável de controle do número de iterações é incrementada em uma unidade.

Quando o valor de  $Iter\_t$  torna-se maior que o valor de  $SA_{max}$ , a variável de controle da temperatura ( $T$ ) é atualizada, recebendo um novo valor, correspondente ao resfriamento e o contador de iterações é zerado.

Quando esta atinge o valor do critério de parada, o método é encerrado, retornando a melhor solução encontrada ( $s^*$ ).

### 3.2.1 Representação da solução do problema

A solução retornada pelo modelo é representada por dois vetores. Em *Python*, este tipo de objeto é chamado de lista e os valores que compõem a lista, de elementos ou itens. A primeira lista representa a localização das bases enquanto a segunda apresenta a respectiva alocação de ambulâncias. A Figura 4 ilustra uma solução da estrutura aqui proposta e deve ser assim interpretada:

- a. Localização: O sistema deve instalar três bases, nas localizações cujos códigos de identificação são 7, 16 e 20;

- b. Na base instalada na localização 7, deve ser alocada duas ambulâncias, uma do tipo USA e outra do tipo USB;
- c. Na segunda base, instalada na localização 16, uma ambulância do tipo USB deve ser alocada;
- d. Na terceira base, instalada na localização 20, mais uma ambulância do tipo USB deve ser alocada.

Figura 4: Exemplo de solução retornada pelo método SA\_SIMUL  
[7, 16, 20], [['USA01', 'USB02'], ['USB01'], ['USB03']]

Fonte: Autores.

### 3.2.2 Determinação da Solução Inicial

O modelo matemático da P-Mediana foi utilizado para se determinar o conjunto de localizações candidatas, que foram utilizadas pela estrutura de otimização da simulação. Essa formulação apresenta-se como a mais popular dentre aquelas associadas aos problemas de localização. Objetiva minimizar a soma ponderada das distâncias das demandas às instalações disponíveis, sendo bastante apropriado a sua utilização, neste estudo. A formulação proposta por Hakimi (1964) é apresentada a seguir:

$$\sum_{i \in N} \sum_{j \in N} w_i d_{ij} x_{ij} \quad (1)$$

sujeito a

$$\sum_{j \in N} x_{ij} = 1, \quad \forall i \in N \quad (2)$$

$$\sum_{j \in N} x_{jj} = p \quad (3)$$

$$x_{ij} \leq x_{jj}, \quad \forall i \in N, j \in N \quad (4)$$

$$x_{ij} \in \{0,1\}, \quad \forall i \in N, j \in N \quad (5)$$

Sendo:

$N$  o conjunto de pontos de demanda;

$i \in N$  determinado cliente ou vértice;

$p$  o número de instalações de serviço ou medianas a serem localizadas;

$w_i$  o peso ou importância do cliente  $i$ ;

$[d_{ij}]_{n \times m}$  a matriz simétrica de distâncias de cada cliente  $i$  à instalação  $j$  com  $d_{ij} = 0, \forall i, j \in N$

$[X_{ij}]_{n \times m}$  a matriz de alocação de cada cliente  $i$ , onde  $x_{ij} = 1$ , se o cliente  $i$  é alocado à instalação  $j$ ,  $x_{ij} = 0$ , caso contrário;

$x_{jj} = 1$  indica que  $j$  é uma mediana e  $x_{jj} = 0$  em caso contrário.

A função objetivo (1) indica a minimização das distâncias ponderadas entre clientes e facilidades. O conjunto de restrições (2) garante que cada cliente  $i$  seja alocado a somente uma instalação  $j$ . As restrições (3) garantem que somente  $p$  facilidades sejam instaladas. As restrições (4) garantem que o cliente somente é atendido por uma facilidade instalada. As restrições (5) impõem variáveis de decisão binárias.

O modelo foi implementado em linguagem AMPL e solucionado através do *solver* CPLEX 12.1.1, em um computador Intel Core I5 1.8 GHz e 16 Gb de memória RAM.

A solução inicial foi definida a partir de um método de escolha aleatório. Sendo assim, de acordo com o número de bases passíveis de localização, a estrutura de geração de solução escolhia aleatoriamente uma, dentre as localidades candidatas disponíveis e após, alocava em cada uma, as ambulâncias disponíveis no cenário avaliado.

### 3.2.3 Estruturas de vizinhança

As estruturas de vizinhanças utilizadas no modelo, consistem em movimentos que objetivam alterar a localização das bases e foram descritas como:

- i. Troca: a estrutura realiza a troca de  $n$  localizações dentre as candidatas, onde  $n$  é determinado de forma aleatória e varia entre 1 e  $m-1$ , sendo  $m$ , o número de bases da solução corrente. As localizações retiradas, bem como as inseridas à solução corrente, são selecionadas aleatoriamente;
- ii. Aumenta número de bases: a estrutura insere  $n$  localizações à solução corrente, onde  $n$  é determinado de forma aleatória e varia entre 1 e  $p-m$ , sendo  $p$  o número máximo de bases a localizar e  $m$ , o número de bases da solução corrente. As novas localizações inseridas à solução corrente, são selecionadas aleatoriamente;
- iii. Reduz número de bases: a estrutura retira  $n$  localizações da solução corrente, onde  $n$  é determinado de forma aleatória e varia entre 1 e  $m-p$ , sendo  $m$  o número de bases da solução corrente e  $p$  o número mínimo de bases a localizar. As localizações retiradas da solução corrente, são selecionadas aleatoriamente.

Após gerada a solução vizinha, as ambulâncias são sorteadas aleatoriamente e inseridas em determinada posição de uma lista de vetores, cujo tamanho equivale ao tamanho da solução

corrente, também determinada de forma aleatória. Associa-se a alocação das ambulâncias à base, a partir dos índices da lista, ou seja, da posição que ocupam na lista.

### 3.3.4 Parâmetros de controle do modelo

Os parâmetros do modelo de otimização da simulação, foram estimados a partir de experimentos, onde diferentes conjuntos dos parâmetros de entrada foram testados no modelo desenvolvido e analisados sob a perspectiva de qualidade da solução encontrada e também, da eficiência computacional. O Quadro 1 apresenta os parâmetros que apresentaram melhores resultados e foram, portanto, utilizados nas execuções do modelo.

Quadro 1: Parâmetros do modelo de otimização da simulação

| Parâmetro                                       | Valor |
|-------------------------------------------------|-------|
| Temperatura Inicial ( $T_i$ )                   | 150   |
| Temperatura Final ( $T_f$ )                     | 0,01  |
| Taxa de resfriamento ( $\alpha$ )               | 0,95  |
| Número máximo de iterações sem melhora (SA_MAX) | 50    |
| Número de replicações do modelo de simulação    | 5     |

Fonte: Resultados da Pesquisa.

## 4 RESULTADO DOS EXPERIMENTOS

Os serviços móveis de urgência brasileiros são inspirados no modelo operacional dos serviços pré-hospitalares franceses, por adotar um processo de regulação para embasar a decisão de enviar ou não as unidades móveis. Entretanto, adquiriu características próprias em decorrência da adaptação às condições locais.

O atendimento se inicia com o recebimento de uma chamada telefônica pelo profissional responsável que busca validar a urgência, coletando todas as informações pertinentes a serem repassadas ao médico regulador, ou encerrando, em caso de não se tratar de urgências. O médico regulador, por sua vez, busca determinar o atendimento necessário embasado em protocolos técnicos, optando por enviar ou não, uma unidade móvel ao local de ocorrência. Em caso de envio, o profissional responsável pelo despacho localiza, em acordo com a política adotada pelo sistema, a ambulância disponível para realizar o atendimento. Em caso de indisponibilidade total, o chamado aguardará por atendimento em fila.

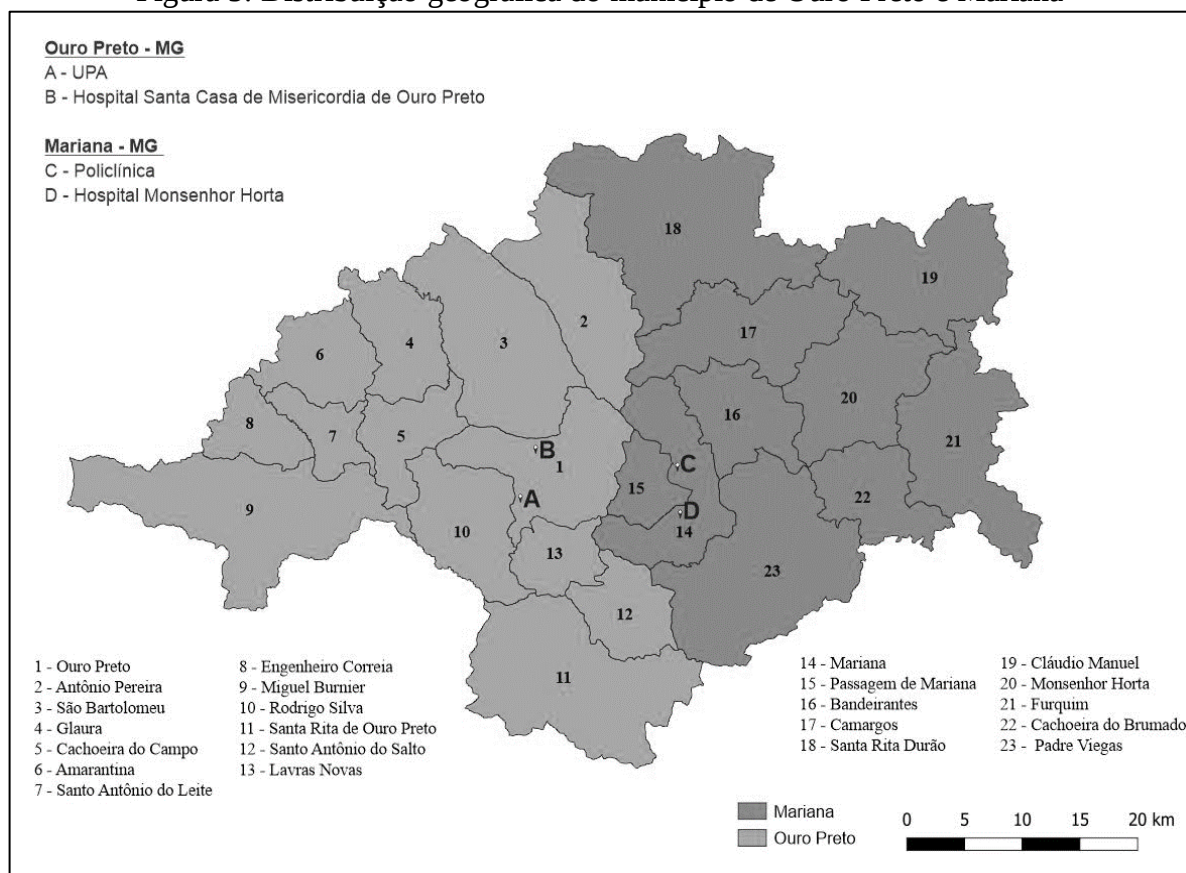
A ambulância despachada inicia então, seu deslocamento para o local de ocorrência. Caso, ao chegar no local, o paciente não seja localizado, verifica-se a existência de um chamado em espera, deslocando para o novo local de ocorrência, em caso positivo, ou, retornando para a base, se não houver nenhum atendimento pendente.

Por outro lado, se localizado o paciente, inicia-se o atendimento e a partir da gravidade real do paciente, verificada in loco, decide-se por removê-lo para uma unidade de saúde, tal qual é definida pelo médico regulador. Em casos onde o paciente recebe alta após o atendimento in loco, a ambulância torna-se disponível a realizar um novo atendimento (em caso de filas) ou para retornar a sua base. Em contrapartida, havendo necessidade de remover o paciente à uma unidade de saúde, a equipe desloca-se para a unidade designada, encerrando o atendimento, após transferir o paciente aos cuidados da equipe médica receptora.

#### *4.1 O Sistema real estudado*

O SME objeto de estudo desta pesquisa possui sedes situadas em Ouro Preto e Mariana, duas das cidades mais antigas do estado de Minas Gerais, Brasil. Os municípios são formados por sua sede e distritos. A área territorial dos municípios totaliza 2.440,073 Km<sup>2</sup> com uma população estimada de 136.654 habitantes (IBGE, 2022). A concentração populacional dos distritos ouropretanos totaliza aproximadamente 42% da população total, enquanto o total de habitantes dos dez distritos marianenses equivalem a aproximadamente 29% (IBGE, 2022). A Figura 5 ilustra a distribuição geográfica das áreas cobertas pelo sistema.

Figura 5: Distribuição geográfica do município de Ouro Preto e Mariana



Fonte: Autores.

O serviço de atendimento móvel a emergências local conta atualmente com quatro ambulâncias, sendo uma de suporte avançado (USA) e três de suporte básico, onde a primeira é despachada para atender os casos de maior gravidade, enquanto as demais são designadas àqueles atendimentos cujo estado de saúde do paciente, ainda que de menor gravidade, requer atenção.

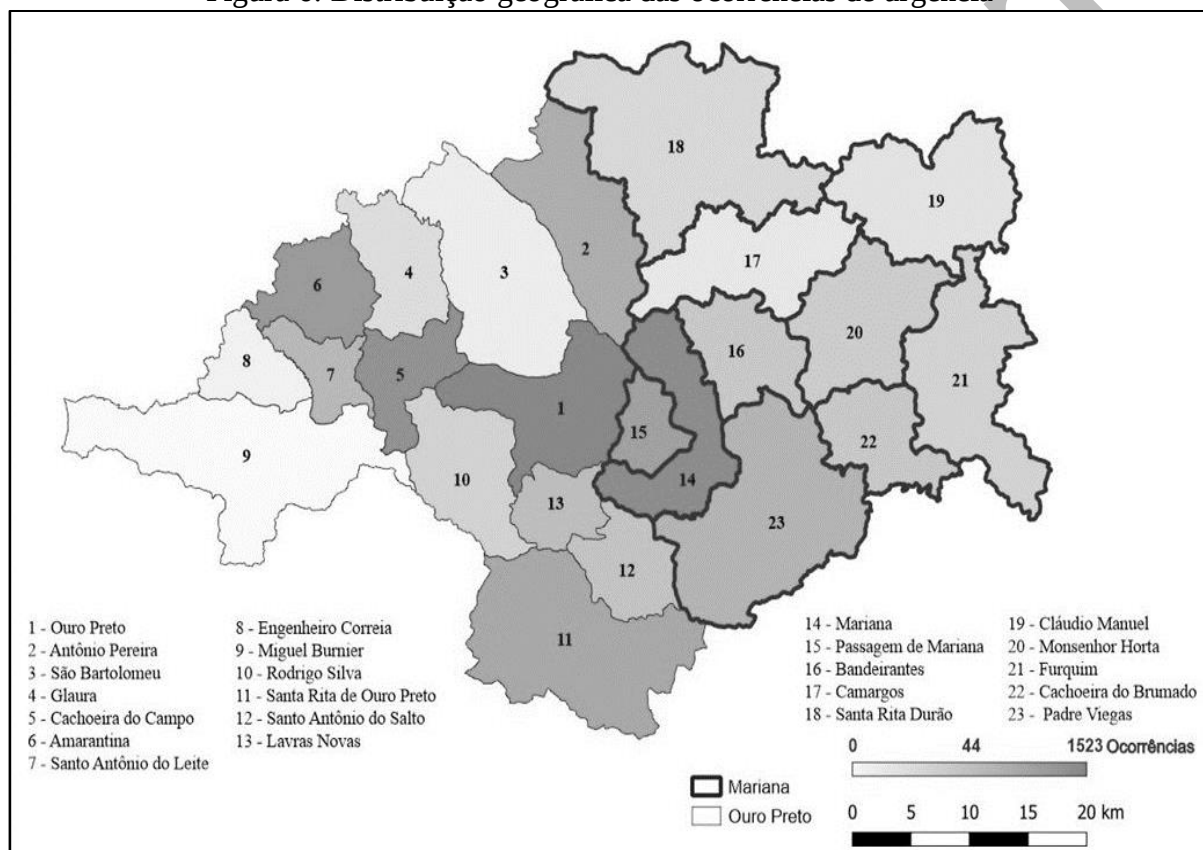
Com o objetivo de representar com maior precisão o sistema estudado e para fins de simplificação, os bairros das sedes dos municípios e os distritos foram agrupados formando microrregiões, considerando aspectos como localização, complexidade geográfica e frequência de chamados. Essa estruturação resultou na formação de 34 células, identificadas neste estudo, por números arábicos sequenciais.

Os veículos do sistema são alocados em bases descentralizadas, sendo que a USA e USB1 ficam alocadas na base instalada na sede de Ouro Preto (célula 4), a USB2 no distrito de Cachoeira do Campo (célula 14), enquanto a USB3 fica alocada na sede de Mariana (célula 26)

O sistema adota a política regionalizada para determinar o despacho das unidades de resgate. Assim, cada base/equipe é designada a atender uma área específica. Em caso de indisponibilidades, a unidade mais próxima disponível é então despachada.

A demanda proveniente de cada localidade coberta pelo sistema é representada pela Figura 6, onde a intensidade da cor é diretamente proporcional ao número de chamados originados na região.

Figura 6: Distribuição geográfica das ocorrências de urgência



Fonte: Autores.

#### 4.2 Cenários analisados

Os experimentos realizados consistiram em avaliar o desempenho do sistema estudado, utilizando o modelo de simulação desenvolvido. Para cada cenário avaliado, os resultados eram obtidos a partir de 5 replicações. Em cada replicação, um ano de operação do sistema, equivalente a 525.600 minutos foram simulados.

Como resultado da execução do modelo matemático, foram definidas as 10 localizações candidatas para a instalação das bases, sendo elas: 4, 8, 11, 13, 14, 19, 22, 24, 26, 27 a serem testadas pelo modelo de otimização da simulação.

Em todos os cenários analisados, as operações do sistema eram simuladas considerando as localizações de bases e alocações de veículos reais. Posteriormente, o modelo de otimização da simulação buscava encontrar uma configuração alternativa, nas mesmas condições (demanda, política de despacho, número de ambulâncias) que resultasse em um tempo médio de resposta reduzido. Ressalta-se que em sua configuração real, o sistema conta com uma base localizada na célula 4 com as ambulâncias USA e USB1 alocadas, uma base instalada na célula 14, com a unidade USB2 alocada e a terceira base localizada na célula 26 onde está alocada a ambulância USB3. Nas subseções a seguir, cada cenário analisado será descrito, assim como os resultados obtidos.

#### 4.2.1 *Localização e alocação estratégica de bases e ambulâncias*

A Tabela 1 apresenta os resultados obtidos da análise de 8 experimentos realizados, tais quais serão descritos posteriormente. Para cada cenário avaliado, nomeado na primeira coluna, apresenta-se duas linhas, com os resultados obtidos para a configuração real do sistema e da configuração alternativa, sendo a segunda, a melhor solução encontrada pelo modelo de otimização da simulação. Ressalta a exceção do cenário 8, onde constam três linhas de resultados necessárias à análise do cenário em questão.

Na segunda coluna, consta o tempo de resposta médio para cada configuração em minutos. Na terceira, o número de bases instaladas, em cada configuração. Na quarta coluna, constam o número médio de chamados em espera, seguido do tempo médio de espera, descritos na quinta coluna. Por fim, na sexta coluna, são apresentados a taxa de ocupação média dos servidores do sistema.

Tabela 1: Indicadores de desempenho do SME estudado

|                  |             | Tempo médio de resposta (min) | Número de bases instaladas | Número médio e clientes em fila | Tempo médio de espera (min) | Taxa média de ocupação (%) |
|------------------|-------------|-------------------------------|----------------------------|---------------------------------|-----------------------------|----------------------------|
| <b>Cenário 1</b> | Real        | 20,12                         | 3                          | 0,62                            | 1,09                        | 17,00                      |
|                  | Alternativa | 18,13                         | 4                          | 0,99                            | 1,00                        | 17,50                      |
| <b>Cenário 2</b> | Real        | 20,48                         | 3                          | 0,62                            | 1,08                        | 17,00                      |
|                  | Alternativa | 18,28                         | 4                          | 1,31                            | 9,77                        | 16,75                      |
| <b>Cenário 3</b> | Real        | 20,63                         | 3                          | 0,61                            | 1,28                        | 18,00                      |
|                  | Alternativa | 18,73                         | 3                          | 1,31                            | 9,93                        | 20,75                      |
| <b>Cenário 4</b> | Real        | 20,78                         | 3                          | 0,61                            | 1,32                        | 19,00                      |
|                  | Alternativa | 18,70                         | 3                          | 1,12                            | 9,57                        | 18,75                      |
| <b>Cenário 5</b> | Real        | 21,07                         | 3                          | 0,64                            | 1,32                        | 20,25                      |
|                  | Alternativa | 18,73                         | 3                          | 1,75                            | 10,80                       | 23,50                      |
| <b>Cenário 6</b> | Real        | 19,88                         | 3                          | 0,60                            | 1,07                        | 16,00                      |
|                  | Alternativa | 18,08                         | 3                          | 1,22                            | 9,81                        | 17,50                      |
| <b>Cenário 7</b> | Real        | 32,75                         | 2                          | 0,77                            | 1,97                        | 21,00                      |
|                  | Alternativa | 27,88                         | 3                          | 2,88                            | 27,12                       | 26,00                      |
| <b>Cenário 8</b> | Ouro Preto  | 775                           | 1                          | 389                             | 400,20                      | 95,00                      |
|                  | Mariana     | 535                           | 1                          | 283                             | 258,00                      | 98,00                      |
|                  | Alternativa | 251                           | 1                          | 463                             | 243,00                      | 98,00                      |

Fonte: Resultados da Pesquisa

#### *Cenário 1 - Configuração real do sistema*

Neste cenário, buscou-se avaliar as métricas de desempenho do sistema com a configuração real do sistema, com 3 bases descentralizadas e política de despacho regionalizada.

Verificou-se pelos resultados, que o sistema apresentava um tempo médio de resposta de aproximadamente 20 minutos com baixo congestionamento, com apenas 0,62 chamados

aguardando por aproximadamente 1 minuto em fila. A ocupação dos servidores também se apresentou baixa, com taxa de ocupação média em 17%.

O modelo de OvS apresentou como melhor solução encontrada a configuração [4,8,14,27],[USB1],[[USA],[USB2],[USB3]]. A solução alternativa, que conta com 4 bases instaladas, reduziu o tempo de resposta médio do sistema em aproximadamente 10%, mantendo com valores baixos, os indicadores de fila e ocupação. Nessa configuração, a sede de Ouro Preto teria duas bases instaladas, sendo a nova base localizada na célula 8. A base da sede de Mariana, deveria ser instalada na célula 27, permanecendo as duas bases já existentes nas células 4 e 14.

#### *Cenário 2 – Aumento da demanda em 5%*

Objetivou-se com essa análise, compreender o comportamento do sistema em um cenário de aumento da demanda em 5%. Os resultados demonstraram que o sistema absorveria esse aumento de demanda com alterações mínimas em seus indicadores de desempenho, quando comparado com a demanda atual. O tempo médio de resposta seria aumentando em apenas 1,75% enquanto a taxa de ocupação dos servidores e indicadores de fila, não seriam afetados.

O modelo de otimização da simulação retornou a solução [4,11,14,27],[[USB1],[USA],[USB2],[USB3], onde o tempo médio de resposta sofreria redução superior a 10%, na mesma condição de aumento de demanda. Nesta configuração, com 4 bases instaladas, a célula 11 deveria receber uma base instalada, tal qual alocaria a unidade avançada. Em Mariana, a base deveria ser instalada na célula 27, mantendo as demais 2 bases já instaladas. Apesar do aumento em termos percentuais dos indicadores de fila, por se tratar de um sistema de grande disponibilidade, esses ainda permaneceriam com valores baixos, na configuração alternativa.

#### *Cenário 3 - Aumento da demanda em 10%*

Com objetivo similar ao cenário anteriormente descrito, nesse, o sistema sofreria um incremento de 10% em relação à demanda atual. Os resultados também demonstraram pequenas variações nos indicadores de desempenho avaliados, onde o tempo de resposta aumentou somente 2,5% e a taxa de ocupação dos servidores, incrementada em 1 ponto percentual. A solução alternativa retornada pelo modelo proposto, [11,14,27],[[USB1,USA],[USB2],[USB3]] resultaria em um tempo de resposta 9,20% menor que a configuração real, na mesma condição de aumento de demanda. Os indicadores de fila e de

ocupação, sofreriam acréscimos consideráveis em termos percentuais, mas ainda apresentariam valores baixos e aceitáveis de desempenho. A configuração alternativa contaria com 3 bases instaladas. Em Ouro Preto, a instalação deveria ocorrer na célula 11 e em Mariana, na célula 27. A terceira base permaneceria no mesmo local atual de instalação.

#### *Cenário 4 - Aumento da demanda em 15%*

Neste cenário, o aumento da demanda em 15 pontos percentuais foi analisado. Enquanto os indicadores de fila ficaram praticamente inalterados, o tempo de resposta sofreu acréscimo de aproximadamente 3,20% e a taxa de ocupação, 2%.

A solução alternativa apresentada pelo nosso modelo, para esse cenário, foi a mesma resultante da análise do cenário apresentado na subseção 5.1.3. Com esta, haveria redução do tempo de resposta em 10% e da taxa de ocupação em 0,75%. Em contrapartida, os indicadores de fila teriam comportamento similar às análises anteriores, permanecendo com valores baixos, apesar do aumento expressivo em termos percentuais.

#### *Cenário 5 – Aumento da demanda em 20%*

Simulando o aumento em 20 pontos percentuais na demanda atendida pelo sistema, verificou-se que o tempo de resposta seria aumentado em 4,5% enquanto a taxa de ocupação seria incrementada em 2%. Os indicadores de fila seriam fracamente afetados por essa condição. A melhor solução encontrada pelo modelo proposto foi [8,14,27], [[USB1,USA], [USB2], [USB3]]. Nesta, uma base deveria ser instalada na célula 8, na sede de Ouro Preto, e em Mariana, a base deveria estar localizada na célula 27. A terceira base permaneceria instalada na célula 14. Com essa configuração, haveria uma redução de 11,10% no tempo de resposta médio, porém com um aumento de 3,25% na taxa de ocupação. Em relação aos indicadores de fila, observou-se o mesmo comportamento descrito nos cenários anteriores.

#### *Cenário 6 – Alteração da Política de Despacho*

Conforme anteriormente citado, o sistema objeto de estudo aplica a política de despacho regionalizada e/ou “Disponível mais próxima”, em caso de indisponibilidade daquela ambulância que preferencialmente atenderia determinado chamado. Entretanto, as ambulâncias do município de Ouro Preto são empenhadas para realizar atendimentos no município de Mariana, mas o contrário não se aplica.

Sendo assim, buscou-se verificar, com a análise desse cenário, o comportamento do sistema, caso houvesse uma alteração na política, onde a única e principal mudança seria o despacho de qualquer ambulância do sistema para realizar atendimentos, independentemente da

localização de origem, ou seja, a USB3 também realizaria atendimentos no município de Ouro Preto.

Os resultados demonstraram uma pequena redução no tempo de resposta médio de aproximadamente 1,20%. Essa alteração também resultaria na redução de 1% na taxa de ocupação média dos servidores, enquanto os indicadores de fila teriam irrelevantes alterações.

Neste contexto, a solução alternativa retornada pelo modelo proposto foi [11,14,27], [[USB1,USA], [USB2],[USB3]], tal qual também foi apresentada na análise dos cenários 3 e 4. Com a configuração alternativa, o tempo de resposta seria reduzido em 9,00%, porém a taxa de ocupação ficaria 1,5% maior. Comportamento similar às demais análises, também foram obtidos nesta, para os indicadores de fila.

#### *Cenário 7 – Retirada de uma ambulância do sistema*

Diante do baixo congestionamento e também das baixas taxas de ocupação, optou-se por verificar o comportamento do sistema quando uma ambulância era retirada e a demanda total, absorvidas pelas demais unidades. Na avaliação do sistema, a unidade USB2, alocada na base instalada na célula 14, foi a escolhida para ser eliminada da configuração do sistema, visto ser a unidade responsável pela cobertura da região de menor demanda. Sendo assim, simulou-se o sistema, com apenas duas bases instaladas, uma em cada município. Os resultados demonstraram que a retirada dessa unidade, acarretaria em um aumento de aproximadamente 63% no tempo de resposta médio. O número de chamados e o tempo de espera aguardando em fila, sofreriam aumentos inexpressivos, assim como a taxa média de ocupação dos servidores.

Em relação à configuração alternativa, o nosso modelo retornou a solução [8, 14,27], [[USB1], [USA], [USB3]]. Com essa configuração, os três servidores reproduziriam um tempo de resposta médio seria menor em 15 pontos percentuais aproximadamente. Os valores dos indicadores de fila e ocupação sofreriam aumentos expressivos, entretanto, tais valores ainda podem ser considerados aceitáveis.

#### *Cenário 8 – Retirada de três ambulâncias do sistema*

A Portaria GM/MS nº 1864 que regulamenta o Serviço Atendimento Móvel de Urgência, tal qual é o sistema objeto de estudo, define no Art. 3º, §3º, que o número de ambulâncias do sistema deve ser definido em acordo com número de habitantes atendidos.

Neste sentido, como o número total de habitantes do sistema em estudo é inferior a 150.000, buscou-se verificar o comportamento do sistema, caso este fosse configurado somente com uma ambulância do tipo USB. Como neste, as populações de dois municípios são atendidas, analisou-se a alocação dessa ambulância nas bases localizadas na sede de cada município. Os resultados dos indicadores de tempo de resposta apresentaram valores proibitivos, levando-se em consideração o objetivo do sistema. Quando a ambulância foi alocada na base do município de Ouro Preto, o tempo médio de resposta do sistema foi de aproximadamente 30 dias enquanto, quando alocada em Mariana, este foi de aproximadamente 22 dias. Verifica-se também que um grande número de chamados permaneceriam em fila, aguardando por atendimento, com consequentes altos tempos de espera, tais quais, também são considerados impraticáveis neste tipo de sistema. A taxa de utilização da ambulância indica uma ocupação do servidor extremamente alta, o que também é impraticável em sistemas de emergência médica.

O modelo de otimização da simulação indicou que a base, em um cenário do sistema somente contar com uma ambulância, deveria ser instalada na célula 11. Assim, o tempo de resposta médio do sistema seria menor, 68% quando comparado com a atual localização da base do município de Ouro Preto e 53% menor, se comparado com a atual localização da base em Mariana. Por outro lado, os indicadores de fila para a configuração retornada pelo modelo, apresentaram piora, com o número médio de chamados aguardando em fila 15% maior quando comparado com atual base do município de Ouro Preto e 39% maior, se comparado com atual localização em Mariana. O tempo máximo de espera também apresentou piora de aproximadamente 37% quando comparado com ambas as localizações atuais.

Apesar da expressiva melhora obtida no tempo de resposta médio do sistema, os valores ainda são impraticáveis nesse tipo de serviço. De mesma forma, a taxa de ocupação da equipe também é incompatível com o objetivo de desempenho desse tipo de sistema.

#### *4.2.2 Tempo de resposta preconizado por leis internacionais*

No Brasil, não há leis que definam o tempo de resposta ideal para o tempo de resposta aos chamados de emergência. Neste cenário buscou-se avaliar quantas ambulâncias seriam necessárias para se alcançar o tempo médio de resposta preconizados pelas leis internacionais de atendimento emergencial. A Organização Mundial da Saúde indica 8 minutos como sendo o tempo de resposta ideal para os atendimentos emergenciais (Nogueira, Pinto, & Silva, 2014). Por outro lado, leis regedoras de SME aplicadas em países da América do Norte, indicam que 90% dos chamados urbanos devem ser atendidos em até

10 minutos enquanto para os chamados oriundos de zona rural, esse tempo deve ser de até 30 minutos (Nair & Miller-Hooks, 2009).

Neste sentido, buscou-se neste cenário identificar a configuração necessária para que o sistema objeto de estudo alcançasse um tempo de resposta médio próximo a 8 minutos.

Ressalta-se que nessa análise, não houve restrições de número de bases, ou seja, todas as 34 células foram consideradas locais candidatos.

O modelo retornou uma solução com 30 ambulâncias, sendo 4 do tipo USA e 26 do tipo USB alocadas em 23 bases descentralizadas. Com essa configuração, o sistema teria um tempo de resposta médio de 8,28 minutos e seria capaz de atender 77,49% dos chamados oriundos de áreas urbanas com um tempo inferior a 10 minutos e 64,75% dos chamados de zonas rurais em um tempo menor que 30 minutos.

## 5 CONCLUSÕES

A aleatoriedade inerente às operações dos sistemas médicos de emergência, torna a gestão de seus processos bastante desafiadora. Neste sentido, desenvolver ferramentas que sejam eficientes e facilmente adaptáveis a análises de novos cenários é bastante relevante.

Assim como nos trabalhos de Nogueira, Pinto e Silva (2014) e Fukushima e Moriya (2021), era objetivo desta pesquisa identificar meios de se reduzir o tempo de resposta de SME's.

Mas, de forma contrária, não utilizamos softwares comerciais para realizar análises estatísticas, simulação e otimização via simulação. Para tanto, desenvolvemos uma estrutura computacional, em linguagem *Python*, que representava as operações do sistema estudado. Nosso modelo se mostrou eficiente ao representar às operações do SAMU de Ouro Preto e Mariana, cujos resultados da simulação de um período correspondente a um ano, eram obtidos em apenas 2 segundos.

A estrutura desenvolvida proporcionou ainda, grande facilidade e adaptabilidade para se analisar cenários alternativos. De mesma forma, verifica-se também a adaptação do modelo para analisar sistemas com características semelhantes.

No modelo de otimização da simulação, a meta-heurística *Simulated Annealing* foi integrada à estrutura de simulação de eventos discretos e observou-se a eficiência do método de solução aproximada. Resultados satisfatórios foram obtidos com custos computacionais e tempos de execução adequados, visto que o tempo médio para se obter as soluções foi de 7

horas e 32 minutos, tempo considerado hábil para se apresentar uma solução para um problema de médio prazo.

O modelo foi capaz de retornar uma configuração estratégica que reduzira o tempo de resposta médio do sistema em aproximadamente 10%. A mesma redução foi obtida com as soluções alternativas apresentadas pelo modelo, para os cenários de aumento de demanda. Ao se verificar a existência de uma configuração alternativa para os cenários de redução do número de ambulâncias, verificou-se a possibilidade de se reduzir o tempo médio de resposta em 15%, a partir de uma nova configuração, caso fosse retirado uma ambulância do sistema. Já para o cenário onde o sistema operaria somente com uma ambulância, adequando-se ao regulamentado pela portaria nacional, que define que uma ambulância deve ser alocada para atender a cada 150 mil habitantes, o modelo desenvolvido apresentou uma solução onde o tempo de resposta médio seria reduzido em pelo menos 53%, ainda que sendo este, extremamente elevado.

Os resultados obtidos indicam também, a necessidade de se considerar outros aspectos variáveis, tais como, extensão territorial, complexidades geográficas e infraestruturas locais, em conjunto com o número de habitantes, ao se determinar o número necessário de ambulâncias para se atender as demandas de urgência no serviço pré-hospitalar, cuja configuração atenda os indicadores de desempenho almejados.

Salientamos que o SME estudado não dispõe de um sistema informatizado e que todos os dados utilizados na pesquisa foram informados, manualmente, pelos operadores do sistema. Considera-se esta uma limitação da pesquisa, visto que em situações de elevada pressão, devido às urgências atendidas, as informações podem ter sido inseridas de forma não precisa.

Como sugestão de trabalho futuros e de melhoria para o modelo desenvolvido, sugere-se (i) integrar meta-heurísticas alternativas, como Algoritmos Genéticos ou GRASP (*Greedy Randomized Adaptive Search Procedure*) com o objetivo de comparar suas respectivas eficiências e (ii) a incorporação de uma função de avaliação multiobjetivo, que considere além do tempo de resposta, outras métricas, buscando encontrar soluções que melhorem o desempenho do sistema a partir de múltiplos indicadores.

Os autores gostariam de agradecer à PROPPI/UFOP, CAPES e FAPEMIG (APQ-04373-17) pelo auxílio financeiro à esta pesquisa.

## REFERÊNCIAS

- Aboueljinnane, L., Sahin, E., Jemai, Z., & Marty, J. (2014). 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*, 47, 46–59.  
<https://doi.org/10.1016/j.simpat.2014.05.007>
- Aringhieri, R., Carello, G., & Morale, D. (2013). Supporting decision making to improve the performance of an Italian Emergency Medical Service. *Annals of Operations Research*, 236(1), 131–148.  
<https://doi.org/10.1007/s10479-013-1487-0>
- Bélanger, V., Lanzarone, E., Nicoletta, V., Ruiz, A., & Soriano, P. (2020). A recursive simulation-optimization framework for the ambulance location and dispatching problem. *European Journal of Operational Research*, 286(2), 713–725.  
<https://doi.org/10.1016/j.ejor.2020.03.041>
- Blackwell, T. H., & Kaufman, J. S. (2002). Response Time Effectiveness: Comparison of Response Time and Survival in an Urban Emergency Medical Services System. *Academic Emergency Medicine*, 9(4), 288–295.  
<https://doi.org/10.1197/aemj.9.4.288>
- Byrne, J. P., Mann, N. C., Dai, M., Mason, S. A., Karanicolas, P., Rizoli, S., & Nathens, A. B. (2019). Association Between Emergency Medical Service Response Time and Motor Vehicle Crash Mortality in the United States. *JAMA Surgery*, 154(4), 286–293.  
<https://doi.org/10.1001/jamasurg.2018.5097>
- Carson, Y., & Maria, A. (1997). Simulation optimization. Proceedings of the 29th Conference on Winter Simulation - WSC '97.  
<https://doi.org/10.1145/268437.268460>
- Coelho, G. F., & Pinto, L. R. (2018). Kriging-based simulation optimization: An emergency medical system application. *Journal of the Operational Research Society*, 69(12), 2006–2020.  
<https://doi.org/10.1080/01605682.2017.1418149>

Da Silva, G. F., Frazzon, E. M., & Casarotto Filho, N. (2016). Proposta de método de otimização-simulação para apoiar a tomada de decisão em internalização de processos produtivos. *Exacta*, 13(3), 155–166.

<https://doi.org/10.5585/exactaep.v13n3.5967>

Felix, J. C., Peixoto, C. S., & Edwige, C. A. (2021). O planejamento de suprimentos utilizando-se a teoria de filas e simulação computacional em uma oficina de recuperação de componentes de vagões. *Exacta*.

<https://doi.org/10.5585/exactaep.2021.19577>

Filho, V. H. dos S., Silva, L. de O., & Júnior, J. da S. F. (2019). Simulação a eventos discretos em uma empresa Brasileira prestadora de serviços imobiliários no Sudoeste Mineiro. *Exacta*, 17(2), 29–45.

<https://doi.org/10.5585/exactaep.v17n2.8034>

Fisher, M. L. (2004). The Lagrangian Relaxation Method for Solving Integer Programming Problems. *Management Science*, 50(12\_supplement), 1861–1871.

<https://doi.org/10.1287/mnsc.1040.0263>

Fukushima, F., & Moriya, T. (2021). Objective evaluation study on the shortest time interval from fire department departure to hospital arrival in emergency medical services using a global positioning system — potential for time savings during ambulance running. *IATSS Research*, 45(2), 182–189.

<https://doi.org/10.1016/j.iatssr.2020.08.001>

Golabian, H., Arkat, J., Farughi, H., & Tavakkoli-Moghaddam, R. (2021). A simulation-optimization algorithm for return strategies in emergency medical systems. *SIMULATION*, 97(9), 565–588.

<https://doi.org/10.1177/00375497211006175>

Gonzalez, R. P., Cummings, G. R., Phelan, H. A., Mulekar, M. S., & Rodning, C. B. (2009). Does increased emergency medical services prehospital time affect patient mortality in rural motor vehicle crashes? A statewide analysis. *The American Journal of Surgery*, 197(1), 30–34.

<https://doi.org/10.1016/j.amjsurg.2007.11.018>

Hakimi, S. L. (1964). Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*, 12(3), 450–459.

<https://doi.org/10.1287/opre.12.3.450>

Instituto Brasileiro de Geografia e Estatística (2022).

<https://cidades.ibge.gov.br/brasil/panorama>.

Junges, V. D. C., Romio, A. M., Simonetto, E. D. O., Campos, S. A. P. de, & Bobsin, D. (2021). Simulação baseada em dinâmica de sistemas: avaliação de cenários em uma associação de reciclagem. *Exacta*.

<https://doi.org/10.5585/exactaep.2021.17058>

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680.

<https://doi.org/10.1126/science.220.4598.671>

McCormack, R., & Coates, G. (2015). A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *European Journal of Operational Research*, 247(1), 294–309.

<https://doi.org/10.1016/j.ejor.2015.05.040>

Moraes, D. G. de, & Silva, A. M. da. (2021). Simulação discreta aplicada à gestão de filas no varejo. *Exacta*, 19(3), 659–677.

<https://doi.org/10.5585/exactaep.2021.14842>

Nair, R., & Miller-Hooks, E. (2009). Evaluation of Relocation Strategies for Emergency Medical Service Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, 2137(1), 63–73.

<https://doi.org/10.3141/2137-08>

Nogueira, L. C., Pinto, L. R., & Silva, P. M. S. (2014). Reducing Emergency Medical Service response time via the reallocation of ambulance bases. *Health Care Management Science*, 19(1), 31–42.

<https://doi.org/10.1007/s10729-014-9280-4>

Pell, J. P. (2001). Effect of reducing ambulance response times on deaths from out of hospital cardiac arrest: cohort study. *BMJ*, 322(7299), 1385–1388.

<https://doi.org/10.1136/bmj.322.7299.1385>

Pons, P. T., & Markovchick, V. J. (2002). Eight minutes or less: does the ambulance response time guideline impact trauma patient outcome? *The Journal of Emergency Medicine*, 23(1), 43–48.

[https://doi.org/10.1016/s0736-4679\(02\)00460-2](https://doi.org/10.1016/s0736-4679(02)00460-2)

Zaffar, M. A., Rajagopalan, H. K., Saydam, C., Mayorga, M., & Sharer, E. (2016). Coverage, survivability or response time: A comparative study of performance statistics used in ambulance location models via simulation–optimization. *Operations Research for Health Care*, 11, 1–12.

<https://doi.org/10.1016/j.orhc.2016.08.001>

Zhen, L., Wang, K., Hu, H., & Chang, D. (2014). A simulation optimization framework for ambulance deployment and relocation problems. *Computers & Industrial Engineering*, 72, 12–23.

<https://doi.org/10.1016/j.cie.2014.03.008>