

UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

Bárbara Regina Pinto e Oliveira

Alocação de leitos de internação com o
uso da abordagem de Otimização via
Simulação

Belo Horizonte

2019

Bárbara Regina Pinto e Oliveira

Alocação de leitos de internação com o uso da abordagem de Otimização via Simulação

Versão final da tese apresentada ao programa de pós-graduação em Engenharia de Produção da Universidade Federal de Minas Gerais, como requisito para a obtenção do título de Doutor em Engenharia de Produção na linha de pesquisa Otimização de Sistemas Logísticos de Grande Porte.

Orientador: Luiz Ricardo Pinto

Coorientador: João Antônio de Vasconcelos

Belo Horizonte

2019

O48a

Oliveira, Bárbara Regina Pinto e.

Alocação de leitos de internação com o uso da abordagem de Otimização via Simulação [recurso eletrônico] / Bárbara Regina Pinto e Oliveira. – 2019.

1 recurso eletrônico (xiv, 83 f. : il., color.) : pdf.

Orientador: Luiz Ricardo Pinto.

Coorientador: João Antônio de Vasconcelos.

Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 75-83.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia de produção - Teses. 2. Simulação (Computadores) - Teses. 3. Otimização - Teses. 4. Algoritmos genéticos - Teses. I. Pinto, Luiz Ricardo. II. Vasconcelos, João Antônio de. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 658.5(043)



FOLHA DE APROVAÇÃO

Alocação de leitos de internação com o uso da abordagem de Otimização via Simulação

BÁRBARA REGINA PINTO E OLIVEIRA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO, como requisito para obtenção do grau de Doutor em ENGENHARIA DE PRODUÇÃO, área de concentração PESQUISA OPERACIONAL E INTERVENÇÃO EM SISTEMAS SOCIOTÉCNICOS, linha de pesquisa Otimização e Simulação de Sistemas Logíst. e de Grande Porte.

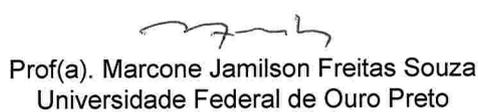
Aprovada em 09 de outubro de 2019, pela banca constituída pelos membros:


Prof(a). Luiz Ricardo Pinto - Orientador
UFMG


Prof(a). João Antônio de Vasconcelos
UFMG


Prof(a). Frederico Gadelha Guimaraes
UFMG


Prof(a). Samuel Vieira Conceição
Universidade Federal de Minas Gerais


Prof(a). Marcene Jamilson Freitas Souza
Universidade Federal de Ouro Preto


Prof(a). Silvio Hamacher
PUC - Rio

Belo Horizonte, 9 de outubro de 2019.

Agradecimentos

Agradeço meu orientador, Prof. Dr. Luiz Ricardo Pinto, e coorientador, Prof. Dr. João Antônio de Vasconcelos, pelo conhecimento compartilhado e motivação no decorrer desta pesquisa.

Aos professores do Departamento de Engenharia de Produção da Universidade Federal de Minas Gerais (UFMG), pela minha formação acadêmica e profissional.

Aos meus pais e minha irmã, pelo apoio e incentivo ao longo de toda minha trajetória acadêmica.

Aos meus amigos e esposo, pela paciência, carinho e compreensão no dia-a-dia.

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

À equipe do Núcleo de Educação em Saúde Coletiva (NESCON), especialmente ao Francisco Carlos Cardoso de Campos por ter me apresentado ao problema estudado nesta pesquisa, e pela colaboração no fornecimento de informações importantíssimas.

Por fim, agradeço também à CAPES pelo apoio financeiro ao desenvolvimento desta pesquisa.

Nunca tenha certeza de nada. A sabedoria começa com a dúvida.

Freud

Resumo

Nesta pesquisa é proposta uma abordagem de Otimização via Simulação para planejamento da capacidade e alocação de leitos de internação. A otimização é conduzida pelo Algoritmo Genético NSGA-II, adequado à formulação e resolução de problemas com múltiplos objetivos. As soluções geradas pelo NSGA-II a cada iteração são avaliadas via Simulação Paralela, pois a avaliação analítica das soluções implicaria em diversas suposições irrealistas e simplistas do problema estudado. A abordagem proposta foi aplicada em uma das macro-regiões de saúde de Minas Gerais, onde cerca de 77% dos pacientes são assistidos pelo sistema público de saúde. Conforme resultados alcançados, 1.179 leitos públicos seriam necessários para minimizar a recusa dos pacientes residentes nesta macro-região, destes, 188 leitos deveriam ser contratados do sistema privado. Embora essa solução aumente os custos diários dos leitos em 13%, ela também é capaz de reduzir as taxas de recusa em cerca de 1/30 dos valores registrados atualmente. Diante disto, a metodologia configura uma alternativa aos métodos empíricos e suposições simplistas aplicados atualmente para planejamento do sistema de internação brasileiro.

Palavras-chave: Otimização via Simulação; Algoritmos Genéticos; Simulação Paralela; Alocação de leitos de internação.

Abstract

This research proposes a Simulation Optimisation approach for planning hospital beds capacity and allocation. The optimisation is conducted by the NSGA-II, a Genetic Algorithm suitable for framing and solving problems with multiple objectives. Each solution generated by it is evaluated through Parallel Simulation, because its analytical evaluation would imply on making several unrealistic and simplistic assumptions. An application was performed in one of the health regions of Minas Gerais, Brazil, where the public health system assists nearly 77% of the patients. The results showed that 1,179 public beds would be appropriate to minimise patients refusal rates, and that 188 more beds should be outsourced from the private system to achieve better service levels. Although this solution increases the beds daily costs by 13%, it reduces the refusal rates about 1/30th of its current values. Therefore, it is an alternative to empirical methods currently applied for planning the Brazilian hospitalisation system.

Keywords: Simulation-Optimization; Genetic Algorithms; Distributed Simulation; Hospital beds allocation.

Lista de Figuras

2.1	Lógica dos AGs	13
2.2	Problema com curva Pareto convexa	16
2.3	Problema com curva Pareto não convexa	17
2.4	Cálculo do hipervolume de um conjunto Pareto	22
2.5	Interação entre os pacotes de otimização e simulação	24
2.6	Método Global para computação paralela dos Algoritmos Evolucionários	27
2.7	Estratégia de Granulação para computação paralela dos Algoritmos Evolucionários	27
4.1	Fluxograma de implementação do Algoritmo NSGA-II	39
4.2	Detalhamento da avaliação dos indivíduos no NSGA-II	39
4.3	Representação dos indivíduos no NSGA-II	40
4.4	Simulação do sistema de internação SUS	43
4.5	Simulação do sistema de internação SUS	46
4.6	Estratégias de avaliação das soluções	47
4.7	Cálculo da Distância de Multidão	49
4.8	Cruzamento dos indivíduos no NSGA-II	51
4.9	Mutação dos indivíduos no NSGA-II	52
4.10	Processo de Elitismo NSGA-II	53
5.1	Macro-regiões de saúde em Minas Gerais	55
5.2	Triângulo sul mineiro	56
5.3	Análise de <i>nger</i> - Primeira fronteira de indivíduos não dominados	61
5.4	Análise de <i>nger</i> - Hipervolume	62
5.5	Análise de p_m e p_c - Hipervolume	64
5.6	Primeira fronteira de indivíduos não dominados	66

5.7 Alocação dos pacientes obstétricos	70
--	----

Lista de Tabelas

2.1	Algoritmos Genéticos Multiobjetivos	18
2.2	Métricas para avaliação dos MOEAs	21
3.1	Compatibilidade entre os leitos de internação e especialidades	31
3.2	Relações de precedência entre os leitos	31
3.3	Custo médio diário de internação nos leitos	33
3.4	Atendimento de pacientes cirúrgicos de União de Minas na rede SUS	36
5.1	Internações nas macro-regiões de saúde de Minas Gerais	56
5.2	População referência e proporção de internações em leitos gerais	57
5.3	Demanda esperada para cada especialidade	58
5.4	Probabilidade de encaminhamento para as UTIs	58
5.5	Distribuição do tempo de internação dos pacientes	59
5.6	Parâmetros do algoritmo NSGA-II	64
5.7	Tempo de cada replicação do algoritmo	65
5.8	Resultados das soluções em destaque	67
5.9	Número de leitos em Araxá	68
5.10	Ocupação média dos leitos em Araxá	69
5.11	Distância máxima percorrida pelos pacientes	69

Lista de Siglas

CNES	Cadastro Nacional de Estabelecimentos de Saúde
DATASUS	Departamento de Informática do Sistema Único de Saúde
DM	Distância de Multidão
ER	<i>Error Ratio</i>
AG	Algoritmo Genético
AGA	Algoritmo Genético de Agrupamento
SIG	Sistema de Informação Geográfica
GD	<i>Generational Distance</i>
HV	<i>Hypervolume</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IGD	<i>Inverted Generational Distance</i>
MOEA	<i>Multiobjective Evolutionary Algorithm</i>
MOP	<i>Multiobjective Optimization Problem</i>
NESCON	<i>Núcleo de Educação em Saúde Coletiva</i>
NSGA	<i>Nondominated Sorting Genetic Algorithm</i>
NPGA	<i>Niched Pareto Genetic Algorithm</i>
PAES	<i>Pareto Archived Evolution Strategy</i>
PIB	Produto Interno Bruto
PO	Pesquisa Operacional
SOP	<i>Scalar Objective Optimization Problem</i>
S	<i>Spacing</i>
SPEA	Strength Pareto Evolutionary Algorithm
SUS	Sistema Único de Saúde
UTI	Unidade de Tratamento Intensivo
VEGA	<i>Vector Evaluated Genetic Algorithm</i>
WBGA	<i>Weight Based Genetic Algorithm</i> Algoritmo Genético Baseado em Pesos

Sumário

1	Introdução	1
1.1	Objetivos	4
1.2	Contribuições	4
1.3	Organização do Texto	5
2	Referencial Bibliográfico	7
2.1	Problemas de localização e alocação na área da saúde	7
2.1.1	Histórico e Aplicações	7
2.1.2	Problema de Programação de Admissão de Pacientes	8
2.1.3	Principais Abordagens	9
2.1.4	Metodologias Aplicáveis	11
2.2	Algoritmos Genéticos	12
2.2.1	Aplicações e Conceitos	12
2.2.2	Problemas com Múltiplos Objetivos	15
2.2.3	Principais Abordagens e Algoritmos	15
2.2.4	Desempenho dos Algoritmos	19
2.3	Otimização via Simulação	22
2.3.1	Conceitos Básicos	24
2.3.2	Limitações da Otimização via Simulação	25
3	Formulação Matemática	29
3.1	Conjuntos	29
3.2	Variáveis de Decisão	30
3.3	Parâmetros	30
3.4	Funções Objetivo	32

3.5	Modelagem Matemática	33
4	Metodologia	38
4.1	Representação dos Indivíduos na População	39
4.2	Geração da Primeira População de Indivíduos	40
4.3	Alocação dos Pacientes	41
4.4	Avaliação dos Indivíduos via Simulação	42
4.5	Classificação dos Indivíduos	46
4.6	Operações de Evolução e Elitismo	50
4.6.1	Cruzamento	50
4.6.2	Mutação	51
4.6.3	Elitismo	52
5	Apresentação dos Resultados	54
5.1	Definição da macro-região de saúde	54
5.2	Dados da macro-região de saúde em estudo	57
5.3	Parâmetros do algoritmo NSGA-II	59
5.4	Principais Resultados	64
6	Conclusões	71
	Referências Bibliográficas	75

Capítulo 1

Introdução

Os sistemas de saúde vêm sendo desafiados pelo aumento do custo total de operação e a ocorrência de importantes alterações demográficas e epidemiológicas. Em 2015, cerca de 9% do Produto Interno Bruto (PIB) brasileiro, i.e., R\$ 546 bilhões foram gastos no consumo final de bens e serviços da área da saúde (IBGE, 2015). Isso representa um aumento de mais de 76% em relação ao ano de 2010. A partir destes dados, é evidente a importância de se desenvolver pesquisas que busquem melhorar a qualidade dos serviços assistenciais através de sistemas mais eficientes.

O planejamento dos sistemas de internação está entre as políticas públicas mais debatidas no Brasil, onde cerca de 77% da população é assistida pelo sistema público de saúde (ANS, 2018). Em 2018, o Brasil contava com um total de 432.788 leitos, sendo 300.280 (69%) exclusivos do Sistema Único de Saúde (SUS), distribuídos desigualmente e concentrados basicamente nos grandes centros urbanos (CNES, 2018).

Apesar de estar entre os problemas mais debatidos na área da saúde, até 2015 a estimativa das demandas dos pacientes e o planejamento do sistema público de internação brasileiro eram pautados na Portaria nº 1.101/2002 (Brasil, 2002), cujos princípios norteadores foram baseados na realidade assistencial daquele momento. Além disso, ao invés de considerar parâmetros específicos para cada tipo de leito e paciente, esta portaria utiliza como referência valores médios e globais de internação, e.g. tempo médio entre as chegadas e permanência dos pacientes no sistema.

Desta forma, enquanto alguns leitos registravam altas taxas de ocupação e recusa, outros apresentavam baixos níveis de utilização, contribuindo fortemente para o alto desperdício de recursos públicos.

Em 2015 a Portaria nº 1.101/2002 (Brasil, 2002) foi revogada, entrando então em vigor a

Portaria nº 1.631/2015 (Brasil, 2015), que consiste na revisão dos parâmetros de planejamento propostos no documento anterior. Mesmo representando um importante passo para melhoria da estimativa das necessidades assistenciais da população brasileira, na prática as diretrizes dispostas na Portaria nº 1.631/2015 (Brasil, 2015) ainda não foram aplicadas.

Ressalta-se que esta não é uma problemática exclusivamente brasileira. Outros países enfrentam problemas similares, e na ausência de diretrizes adequadas, decisões estratégicas de planejamento público são frequentemente fundamentadas em perspectivas puramente políticas ou pragmáticas.

A partir do exposto, torna-se evidente a necessidade de estabelecer uma metodologia eficiente para planejar e otimizar o sistema público de internação brasileiro, que considere de maneira conjunta as especificidades das especialidades demandadas e recursos ofertados, e.g., ocupação desejável para cada tipo de leito e tempo de internação dos pacientes para cada tipo de especialidade.

Esta pesquisa utilizou como referência o problema de otimização da alocação dos leitos SUS em uma das macro-regiões de saúde do estado de Minas Gerais, devido principalmente à existência de um amplo banco de dados disponível pelo Departamento de Informática do SUS (DATASUS), de onde foram coletadas as seguintes informações:

1. Número total de leitos registrados nos municípios.
2. Demanda de internações para cada especialidade.
3. Tempo de internação dos pacientes nos leitos.
4. Custo diário dos leitos.
5. Percentual de beneficiários de plano de saúde para cada especialidade.

Basicamente, três alternativas podem viabilizar a otimização do sistema estudado: (i) aquisição de novos leitos públicos; (ii) contratação de leitos do sistema privado ou (iii) realocação dos leitos SUS existentes.

Diante das restrições financeiras a primeira alternativa não foi considerada, i.e., investigamos se realocando os leitos e fazendo novas contratações do sistema privado seria possível atender a demanda existente com custos e taxas de recusa mínimos.

Apesar dos grandes esforços empregados em pesquisas voltadas ao projeto desses sistemas, na prática os métodos exatos de otimização se mostraram ineficientes na avaliação de problemas

desta natureza, i.e., problemas NP-difíceis de grande porte. Os principais obstáculos destas metodologias estão relacionados ao baixo nível ou mesmo falta de estruturação dos problemas relacionados à área da saúde, à aleatoriedade e dimensão dos sistemas estudados (Ahmadi-Javid et al., 2017).

Quanto aos modelos de simulação, amplamente utilizados na reprodução e análise de sistemas dinâmicos, o grande desafio consiste na redução do tempo de simulação.

Outras abordagens foram, então, propostas para encontrar as soluções mais eficientes para esses problemas e investigar suas performances com recursos computacionais limitados. Dentre as quais destaca-se a Otimização via Simulação (OvS), utilizada nesta pesquisa. Na OvS a interação entre estas técnicas se dá através da otimização de variáveis de desempenho avaliadas por modelos dinâmicos de simulação (Fu et al., 2015).

Segundo Lin et al. (2013), poucas rotinas de otimização incluídas nos pacotes de simulação são adequadas ao tratamento de problemas com múltiplos objetivos. Por outro lado, nos problemas de alocação de recursos públicos é comum a existência de objetivos conflitantes.

Diante disto, nesta pesquisa a otimização é conduzida pelo Algoritmo Genético NSGA-II, proposto em Deb et al. (2002). As soluções são avaliadas via Simulação a Eventos Discretos. A integração destas metodologias se justifica pela inviabilidade de se formular analiticamente uma das funções objetivo do problema estudado. Conforme será detalhado, a simulação funciona como uma "caixa preta" que avalia esta função para determinado conjunto de *inputs* fornecidos pela otimização.

Os resultados da simulação são utilizados pelo NSGA-II para comparar as soluções encontradas e estabelecer novas direções de busca do algoritmo. A interação entre estes métodos continua até que um critério de parada seja alcançado.

A principal premissa verificada nesta pesquisa é que o planejamento do sistema público de internação a partir da metodologia proposta seja capaz de minimizar as taxas de recusa e o custo dos leitos, e também de reduzir os problemas agravados pelo planejamento ineficiente deste sistema, como: (i) elevado risco de infecção dos pacientes alocados em condições precárias, (ii) longas filas de espera para internação dos pacientes, e (iii) elevado desperdício de recurso público investido em leitos mal estimados.

1.1 Objetivos

O principal objetivo desta pesquisa consiste em otimizar a alocação de leitos de internação conectando a Simulação a Eventos Discretos, geralmente aplicável a problemas com baixo nível de estruturação, aos Algoritmos Genéticos, utilizados na formulação e otimização de problemas com múltiplos objetivos.

Para alcançar este objetivo, esta pesquisa apresenta como objetivos específicos:

- Sugerir uma formulação matemática para otimizar a alocação de leitos de internação;
- Implementar o Algoritmo Genético NSGA-II para conduzir o processo de otimização;
- Desenvolver um modelo de simulação para avaliar as funções objetivo do problema, para posterior ordenação e comparação das soluções pela otimização;
- Elaborar uma proposta de integração dos modelos em um processo de Otimização via Simulação.

1.2 Contribuições

A principal contribuição desta pesquisa consiste em propôr um método de Otimização via Simulação para otimizar a alocação de leitos de internação.

A principal premissa averiguada é que a integração do Algoritmo Genético NSGA-II ao processo de Otimização via Simulação seja mais adequada para resolver este tipo de problema do que os métodos empíricos e suposições simplistas aplicados atualmente para planejamento do sistema de internação brasileiro.

A partir das análises feitas também foi averiguado que a metodologia proposta nesta pesquisa atende algumas das demandas latentes da Otimização via Simulação levantadas em Amaran et al. (2014), como:

1. Capacidade de representar problemas NP-difíceis, com múltiplos objetivos e variáveis contínuas e discretas trabalhadas conjuntamente.
2. Tratamento adequado de múltiplos *outputs* na simulação;
3. Consolidação e cooperação de diversas técnicas em algoritmos híbridos.

Além destes aspectos científicos, a pesquisa também contribui propondo uma formulação mais abrangente ao problema estudado, não limitando-se a um município ou recurso específico, mas buscando otimizar o sistema de internação como um todo. Ao contrário do proposto, verificou-se que grande parte das pesquisas levantadas se limita ao estudo de um recurso/sistema específico em determinada localidade/installação, quando na prática, nas redes de saúde, observa-se a interação de várias facilidades hospitalares com diferentes recursos e capacidades de atendimento.

No intuito de abordar esta diversidade de demandas e serviços das redes de saúde, os parâmetros e variáveis do problema estudado foram discriminados por especialidade, município de residência e atendimento dos pacientes, e tipo de leito. Isto garante a abordagem de subsistemas mais homogêneos, evitando os problemas resultantes da análise indevida de valores médios e globais de internação já retratados: (i) elevadas taxas de recusa devido à subestimação de leitos menos especializados, e.g. clínicos, e (ii) desperdício dos recursos públicos devido à superestimação de leitos mais especializados, e.g. UTIs.

Além disso, ao contrário de diversas pesquisas, a função custo diário dos leitos não foi representada por uma constante, mas sim por uma função que depende do número e tipo de leito ofertado, de forma à considerar na otimização a lógica básica da economia de escala.

Finalmente, conforme sugerido em Jones (2011), considerou-se também a relação entre o número de leitos alocados em cada local e a recusa máxima admissível para determinar as taxas de ocupação desejáveis para cada conjunto de leitos.

1.3 Organização do Texto

O restante deste texto está organizado conforme detalhado na sequência. O segundo capítulo apresenta uma revisão teórica acerca dos principais problemas de alocação de recursos na área da saúde, da utilização de Algoritmos Genéticos na formulação e resolução de problemas com múltiplos objetivos e a combinação desses algoritmos em abordagens de Otimização via Simulação. O terceiro capítulo propõe uma formulação matemática ao problema de alocação de leitos de internação. O quarto capítulo propõe a aplicação do Algoritmo Genético NSGA-II para resolução do problema destacado, e a integração da Simulação a Eventos Discretos para avaliação das soluções obtidas. O quinto capítulo apresenta os resultados alcançados na aplicação da abordagem de Otimização via Simulação em uma das macro-regiões de saúde de Minas Gerais. Por fim, o sexto e último capítulo apresenta as conclusões geradas a partir da análise dos resultados alcançados

nesta pesquisa, indicando também possíveis trabalhos futuros que poderão ser desenvolvidos para continuidade e aprofundamento das contribuições práticas e teóricas aqui levantadas.

Capítulo 2

Referencial Bibliográfico

Neste capítulo é apresentado o referencial bibliográfico relacionado à pesquisa. O objetivo desta revisão consiste em fornecer aos leitores um arcabouço teórico necessário para fundamentar as contribuições desta pesquisa. Os seguintes conteúdos foram abrangidos:

1. Problemas de localização e alocação na área da saúde
2. Algoritmos Genéticos
3. Otimização via Simulação

2.1 Problemas de localização e alocação na área da saúde

Nos últimos anos as redes de saúde vêm sendo desafiadas pela diminuição do crescimento econômico mundial, pelo aumento do custo total de operação, e significativas alterações demográficas e epidemiológicas. Pesquisas realizadas pelo IBGE revelaram que R\$ 446 bilhões foram gastos em indústrias e prestadoras de serviços da área de saúde no Brasil em 2015. Este valor representa cerca de 9% do PIB nacional, reforçando a importância de se desenvolver pesquisas no que se refere à correta alocação dos recursos destinados ao setor de saúde do país.

2.1.1 Histórico e Aplicações

Na última década constatou-se um aumento no interesse da aplicação da PO na otimização de sistemas na área da saúde, isto se deve, principalmente, à capacidade de analisá-los através de modelos abstratos. Além disso, os hospitais e indústrias da área de saúde precisaram reduzir

seus custos de operação, fazendo dos métodos da PO importantes ferramentas no processo de tomada de decisão (Diefenbach and Kozan, 2011).

Dentre os problemas abordados pela PO na área da saúde, destacam-se:

- Dimensionamento de sistemas hospitalares (e.g. Zhou et al., 2018; Bidhandi et al., 2019).
- Localização e alocação de facilidades (e.g. Zhang et al., 2016; Zarrinpoor et al., 2017).
- Projeto de operação de sistemas de urgência e emergência (e.g. Nogueira et al., 2016; Aringhieri et al., 2017).

Na área da saúde, os impactos destas decisões vão muito além das métricas econômicas comumente utilizadas pelos gestores. Nestes casos, o aumento dos índices de morbidade e mortalidade também estão entre os critérios avaliados (Ahmadi-Javid et al., 2017). A incompreensão destas peculiaridades, bem como dos objetivos e restrições destes problemas, fizeram com que diversas iniciativas falhassem na tentativa de propor melhorias às redes de saúde (Rich and Piercy, 2013).

2.1.2 Problema de Programação de Admissão de Pacientes

O Problema de Programação de Admissão de Pacientes (*Patient Admission Scheduling Problem* - *PASP*) consiste em designar os pacientes de um sistema de saúde para os leitos compatíveis com suas especialidades, buscando maximizar a eficiência do tratamento, o conforto do paciente e a ocupação dos hospitais, satisfazendo as restrições médicas necessárias e observando, quando possível, as preferências dos pacientes Bastos et al. (2019).

O PASP está entre os problemas mais debatidos na área da saúde, pois internações canceladas ou atrasadas influenciam diretamente na qualidade dos serviços prestados e na dotação de recursos aos hospitais públicos (Bachouch et al., 2012).

Além disso, o planejamento ineficiente desses sistemas pode refletir no aumento do tempo de espera e também do risco de infecção dos pacientes, que muitas vezes são internados em ambientes inapropriados enquanto aguardam a liberação da internação (Pinto et al., 2014).

Neste âmbito, Jones (2011) destaca a importância de se analisar a relação entre as taxas de recusa e ocupação desejável dos leitos. No trabalho dos autores, os sistemas de internação são comparados às filas de espera Erlang-B, em que um pequeno incremento na ocupação dos recursos ofertados resulta no aumento exponencial dos tempos de espera dos usuários.

Entretanto, a definição da ocupação adequada desses recursos é bastante complexa, pois, depende dos seguintes fatores:

1. **Tipologia dos recursos:** leitos mais dedicados ou específicos, e.g., neonatal, devem manter taxas de ocupação mais baixas.
2. **Capacidade dos sistemas:** pequenos hospitais devem manter níveis mais baixos de ocupação para operarem com taxas de recusa semelhantes às observadas em grandes centros de internação.

Atualmente, a análise desse problema está entre as demandas mais debatidas pelas políticas públicas no Brasil, onde cerca de 77% da população é assistida pelo sistema público (ANS, 2018). Os leitos disponíveis para internação, entretanto, não são suficientes para atender esta demanda. Conforme informações levantadas no Cadastro Nacional de Estabelecimentos da Saúde (CNES), em 2018 o Brasil contava com 432.788 leitos, sendo 300.280 (69%) exclusivos do sistema público, distribuídos desigualmente e concentrados basicamente nos grandes centros urbanos.

Além disto, enquanto alguns tipos de leito registram altos índices de recusa e ocupação, outros indicam baixa ocupação, contribuindo fortemente no desperdício dos recursos públicos investidos. Dentre as causas do planejamento ineficiente destes recursos, destaca-se o fato de que os gestores ao planejar os sistemas de internação, frequentemente, se baseiam em parâmetros medianos e globais, acarretando na subestimação de leitos menos especializados, e.g. clínico, e na superestimação de leitos mais especializados, e.g. neonatologia.

Diante do exposto, verifica-se a necessidade de se estabelecer um método mais acurado para otimizar a alocação de leitos nas redes de saúde, que considere parâmetros específicos dos serviços de internação, e.g. tipologia dos recursos, e, das necessidades assistenciais, e.g. faixa etária e tempo de internação dos pacientes.

É importante ressaltar que assim como todos os problemas de otimização, o PASP pode ser hierarquicamente classificado nos seguintes níveis de decisão: (i) estratégico, (e.g. Chen et al., 2010; Lee and Kwak, 2011), (ii) tático, (e.g. Bair et al., 2010; Holm et al., 2013) e (iii) operacional, (e.g. Ceschia and Schaerf, 2011; Range et al., 2014).

2.1.3 Principais Abordagens

Em Ahmadi-Javid et al. (2017) os problemas de localização e alocação são classificados em dois tipos.

A primeira categoria consiste no problema da p-mediana, cujo objetivo é minimizar a distância ponderada percorrida entre os pontos de oferta e demanda do sistema estudado. Neste

caso, cada usuário é alocado à facilidade mais próxima. Apesar de estar entre as abordagens mais utilizadas na localização e alocação de facilidades, as soluções por ela propostas nem sempre são equitativas, forçando alguns usuários à deslocarem mais na rede que outros (Rahman and Smith, 2000).

A segunda categoria é formada por problemas de cobertura, em que se considera as restrições de distância máxima permitida no deslocamento dos usuários. Apesar da popularidade deste tipo de abordagem, encontrar soluções ótimas nos problema de cobertura não é uma tarefa fácil, principalmente em países em desenvolvimento, onde pode não ser possível garantir os recursos necessários ao atendimento das restrições de distância.

Nesta tese, propomos uma abordagem diferente para alocar os leitos de internação. Neste caso: (i) a alocação dos pacientes respeita uma distância máxima permitida ao deslocamento, (ii) os leitos só podem ser retirados ou realocados para outra unidade médica do mesmo local e (iii) novos leitos não podem ser adquiridos pelo sistema público, entretanto, mais leitos podem ser contratados do sistema privado.

Desta forma, o número máximo de leitos que cada local poderá receber é dado pela soma dos leitos privados e públicos já existentes nestas localidades. O problema é, portanto, capacitado, e por isto nem sempre será possível encontrar uma solução que garanta que 100% dos pacientes alocados sejam atendidos, i.e. o paciente será recusado caso o leito em que foi alocado não esteja disponível no momento da internação.

Diferente das abordagens comumente utilizadas, os pacientes ou usuários do sistema estudado podem não ser alocados aos leitos das unidades mais próximas, isto se as restrições de distância forem respeitadas e a recusa e custos dos leitos forem minimizados.

Diversas pesquisas de alocação de recursos públicos levantadas neste referencial consideram apenas um único objetivo nos problemas analisados (e.g. Bruno et al., 2016; Guerriero et al., 2016; Cao et al., 2018), seja este de maximização da qualidade dos serviços ofertados ou de minimização das distâncias totais percorridas pelos usuários.

Entretanto, na prática é comum a existência de problemas com múltiplos e conflitantes objetivos (*Multiobjective Optimisation Problems – MOPs*), que necessitam de abordagens mais específicas na procura por soluções não dominadas (Stummer et al., 2004).

Segundo Gunantara (2018), na tentativa de considerar múltiplos objetivos nas formulações matemáticas, algumas aplicações de métodos da PO tratam apenas uma função objetivo do problema, enquanto as demais são incluídas nas restrições, sujeitas a limites inferiores ou superiores

que definem a região de viabilidade dos problemas. Outra opção consiste na agregação dos múltiplos objetivos na função objetivo escalar. Posteriormente, serão mostradas nesta pesquisa as principais implicações na utilização destas abordagens.

Recentemente, os critérios de equidade espacial, cobertura, acessibilidade e custos foram considerados em Zhang et al. (2016) para planejar a localização e alocação de facilidades da área de saúde em Hong-Kong. Evidentemente, os objetivos analisados são conflitantes, pois, para se prover acessibilidade é preciso instalar as facilidades próximas às regiões mais povoadas, culminando em soluções de maior custo e menor equidade espacial. Por sua vez, a busca por custos operacionais menores tende a concentrar a capacidade do sistema em uma ou poucas facilidades nos grandes centros urbanos, de forma a reduzir os gastos com instalação e manutenção de pequenas facilidades, e fomentar a economia de escala, piorando por sua vez os critérios de acessibilidade e equidade espacial.

A utilização de formulações que abranjam os conflitos mencionados anteriormente, garantiria maior equilíbrio dos critérios de otimização.

2.1.4 Metodologias Aplicáveis

Diferentes formulações matemáticas são utilizadas para modelagem de problemas de localização e alocação de facilidades ou instalações. A programação inteira mista destaca-se entre os modelos já desenvolvidos (Afshari and Peng, 2014), e.g. Bastos et al. (2019).

Apesar dos grandes esforços empregados em pesquisas voltadas ao projeto desses sistemas, na prática os métodos exatos se mostraram ineficientes na avaliação dos problemas de grande porte desta natureza (Ahmadi-Javid et al., 2017). Assim como em outros problemas de otimização combinatória, os seguintes fatores dificultam ou inviabilizam a aplicação de métodos exatos para otimizar a alocação de recursos públicos (Jones et al., 2002):

- Grande quantidade de variáveis binárias e inteiras;
- Não linearidade dos problemas;
- Formulações analíticas complexas ou inviáveis;

Além disso, a estocasticidade e aleatoriedade intrínseca nas operações dos sistemas de interação dificultam a definição do número ótimo de leitos, fazendo com que junções de técnicas de simulação, teoria das filas, e modelagem matemática se tornem cada vez mais comuns.

Em Pinto et al. (2014), os autores propõem um modelo analítico baseado em Teoria das Filas para otimizar a alocação de leitos de internação. A viabilidade e o comportamento das configurações propostas pela otimização são avaliados via simulação.

Ao analisar problemas de grande porte, entretanto, diversas pesquisas recorrem a heurísticas que possam gerar soluções eficientes com esforço computacional razoável (Adewumi and Ali, 2010).

Os Algoritmos Genéticos, e.g., Shi et al. (2017), se destacam entre algumas das heurísticas aplicadas na resolução de problemas de otimização na área da saúde, seguido dos métodos de *Particle Swarm Optimization (PSO)* e *Simulated Annealing*, e.g. Chen and Lin (2017) e Eskandari-Khanghahi et al. (2018).

2.2 Algoritmos Genéticos

Os Algoritmos Genéticos (AGs) englobam algoritmos de otimização estocástica baseados no princípio da Evolução Natural de Charles Darwin, em que os indivíduos mais adaptados às condições ambientais presentes têm chances maiores de sobreviverem e reproduzirem (Bäck, 1963).

Estes algoritmos têm recebido grande atenção da comunidade científica na procura por métodos mais eficientes na resolução de problemas de otimização de grande porte e múltiplos objetivos.

2.2.1 Aplicações e Conceitos

Os Algoritmos Genéticos são utilizados na resolução de problemas no campo da saúde (e.g. Shi et al., 2017), manufatura (e.g. Li et al., 2016), finanças/economia (e.g. Kozeny, 2015) e logística (e.g. Soleimani and Kannan, 2015).

Segundo Coello (2006), os AGs apresentam algumas vantagens relação aos métodos clássicos de otimização:

1. Os AGs não dependem de informações do gradiente das funções objetivo no processo de otimização, permitindo sua aplicação em problemas menos estruturados da PO.
2. Os AGs avaliam diversas soluções em conjunto, garantindo sua aplicação em problemas com múltiplos objetivos.
3. Os AGs utilizam operadores estocásticos na condução da otimização, favorecendo a continuidade de soluções desejáveis ao associá-las à maiores chances de evolução.

Segundo Deb (2011), esses princípios garantem maior aplicabilidade dos AGs em problemas NP-Difíceis com múltiplos objetivos e restrições/objetivos não lineares ou estocásticos.

Na terminologia específica dos AGs, um indivíduo nada mais é que uma solução candidata ao problema analisado. A Figura 2.1 apresenta uma síntese da lógica dos AGs.

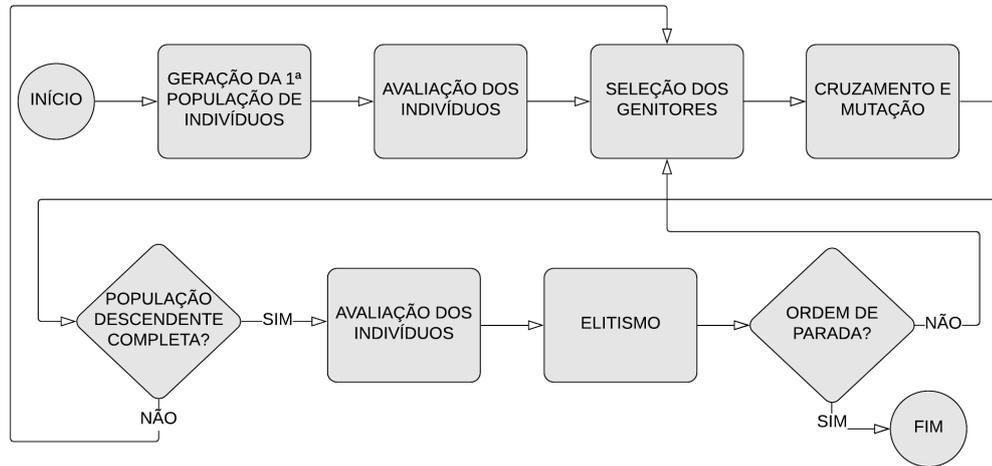


Figura 2.1: Lógica dos AGs

Inicialmente, uma população com $npop$ indivíduos é definida. Em diversas pesquisas (e.g. Kozeny, 2015; Soleimani and Kannan, 2015), essa geração inicial de soluções é aleatória. Entretanto, alguns estudos (e.g Li et al., 2016; Shi et al., 2017) propõem a utilização de um conhecimento específico ou metodologia na proposição de soluções iniciais, buscando melhor desempenho e convergência dos algoritmos.

Definida a geração inicial, os indivíduos são avaliados conforme os objetivos e restrições do problema, e a evolução ocorre por meio dos processos de cruzamento e mutação dessas soluções. O cruzamento é a combinação de dois ou mais indivíduos genitores para definição de um ou mais indivíduos descendentes. A mutação consiste na perturbação de um indivíduo específico. Ao reintroduzir diversidade genética à população, a mutação permite explorar novas buscas locais na vizinhança deste indivíduo. As probabilidades de cruzamento, $p_c \in [0,1]$, e mutação das soluções, $p_m \in [0,1]$, determinam a proporção de indivíduos que participarão desses processos, i.e. $p_c \times npop$ e $p_m \times npop$, e a proporção de indivíduos que serão apenas copiados na próxima geração, i.e. $(1 - p_c) \times npop$ e $(1 - p_m) \times npop$.

Os operadores evolutivos buscam manter a diversidade dos indivíduos de geração em geração, evitando a convergência prematura do algoritmo.

A seleção dos indivíduos que participarão destes processos depende da forma como os algorit-

mos são implementados, mas, é importante que indivíduos descendentes tenham grandes chances de serem mais bem avaliados, desta forma, a evolução da espécie será realmente observada, e a diversidade da população será de fato mantida.

Segundo Zhou et al. (2011), a avaliação dos indivíduos se diferencia nos AGs aplicados aos problemas de otimização de um objetivo (*Scalar Objective Optimization Problems - SOPs*) e múltiplos objetivos (*Multiobjective Optimization Problems - MOPs*). Enquanto nos SOPs é possível classificar e diferenciar as soluções por meio dos métodos de ordenação simples, nos MOPs a ordenação e comparação das soluções é parcial, o que será mais bem detalhado na sequência pelo conceito de dominância.

O processo de elitismo é o último realizado, e seleciona os indivíduos mais bem avaliados da população formada por genitores e descendentes para compor a próxima geração, que passará pelos mesmos processos evolutivos. O principal objetivo do elitismo é garantir que soluções mais eficientes não sejam perdidas devido aos fatores randômicos do processo de evolução (Zitzler et al., 2004).

O número máximo de gerações é o indicador mais utilizado como critério de parada dos AGs. Entretanto, outros indicadores estatísticos também podem ser usados como critérios de parada.

Nos AGs a escolha da estrutura para representar os indivíduos na população é extremamente importante. A representação mais simplificada é a matricial. Nela cada linha representa um indivíduo, já as colunas indicam os valores das variáveis e funções objetivo daquele indivíduo.

Além desta estrutura simplificada, outras foram propostas às distintas aplicações desses algoritmos. Dentre as quais, cita-se os Algoritmos Genéticos de Agrupamento (AGAs), em que os indivíduos são representados por grupos de genes. Pitaksringkarn and Taylor (2005) utilizam esta estrutura de agrupamento no problema de localização dos pontos de coleta de um produto agrícola da Tailândia. Neste caso, 73 pontos de coleta foram divididos entre os cinco estados da região nordeste da Tailândia.

Em Adewumi and Ali (2010) outro tipo de estrutura é proposta para representar os indivíduos em um Algoritmo Genético Multinível aplicado ao problema de alocação dos alojamentos estudantis. No estágio inicial do algoritmo as populações são estruturas com tridimensionais, em que os índices i , j e p definem o número de estudantes da categoria i alocados no corredor j na solução proposta pelo indivíduo p . No último estágio as populações são estruturas com dimensões i , k , l , p , que definem o número de estudantes da categoria i alocados no andar k do bloco l na solução proposta pelo indivíduo p .

2.2.2 Problemas com Múltiplos Objetivos

Por utilizarem uma abordagem baseada na análise de diversas soluções em conjunto, os AGs são indicados na resolução de problemas de otimização com múltiplos objetivos (Konak et al., 2006). Os MOPs apresentam a seguinte formulação matemática genérica:

$$\begin{aligned} \min/\max f(x) &= f_1(x), f_2(x), \dots, f_M(x) \\ \text{s.a. } x &\in \Omega \end{aligned} \quad (2.1)$$

Em que Ω é o espaço de decisão, $x \in \Omega$ é o vetor de decisão com N variáveis, $x = (x_1, x_2, \dots, x_N)$, e $f(x)$ é o vetor de M funções objetivo $f(x) = f_1(x), f_2(x), \dots, f_M(x)$.

Nos problemas restritos, o espaço viável Ω é definido conforme Equação (2.2).

$$\Omega = \begin{cases} g_j(x) = g_j(x_1, x_2, \dots, x_N) \leq 0 & j = 1, 2, \dots, J \\ h_k(x) = h_k(x_1, x_2, \dots, x_N) = 0 & k = 1, 2, \dots, K \\ x_i^L \leq x_i \leq x_i^U & i = 1, 2, \dots, N \end{cases} \quad (2.2)$$

Para cada ponto/vetor x do espaço viável Ω ($\Omega \subset \mathbf{R}^N$) existe um ponto/vetor z no espaço dos objetivos θ ($\theta \subset \mathbf{R}^M$) definido por $z = (f_1(x), f_2(x), \dots, f_M(x))$.

2.2.3 Principais Abordagens e Algoritmos

Existem três tipos de métodos para abordar problemas desta natureza (Farahani et al., 2010). O primeiro consiste no Método ξ -restrito, em que um dos objetivos é otimizado enquanto os demais são incluídos no conjunto de restrições, e.g. (Tavana et al., 2018).

A Equação (2.3) exemplifica a transformação de um problema biobjetivo em um problema com apenas um objetivo. Neste caso, a função f_1 é incluída no conjunto de restrições, reduzindo o espaço de soluções viáveis ao conjunto de pontos com f_1 inferior à ξ .

$$\begin{aligned} \text{Min } f_2 \\ \text{s.a. } f_1 \leq \xi \end{aligned} \quad (2.3)$$

Outra opção seria formular o problema agrupando os objetivos por uma única função escalar, e.g. Karmellos et al. (2015). Neste caso, conforme indicado na Equação (2.4), à cada função objetivo é atribuído um peso w_m proporcional ao grau de importância do critério de otimização.

$$\begin{aligned} \text{Min } z &= w_1 \times f_1 + w_2 \times f_2 \\ \text{s.a. } &w_1 + w_2 = 1 \end{aligned} \tag{2.4}$$

Apesar de simples, esta abordagem apresenta algumas limitações, uma vez que (i) requer conhecimento prévio acerca da importância relativa dos objetivos de otimização, (ii) não considera os conflitos existentes entre estes objetivos e (iii) é inadequada na solução de problemas cujo espaço de busca não seja convexo.

Conforme ilustrado nas Figuras 2.2 e 2.3, se a curva Pareto não for convexa inexistirá um valor de α para o qual a solução ótima de um SOP será obtida na parte não convexa da curva.

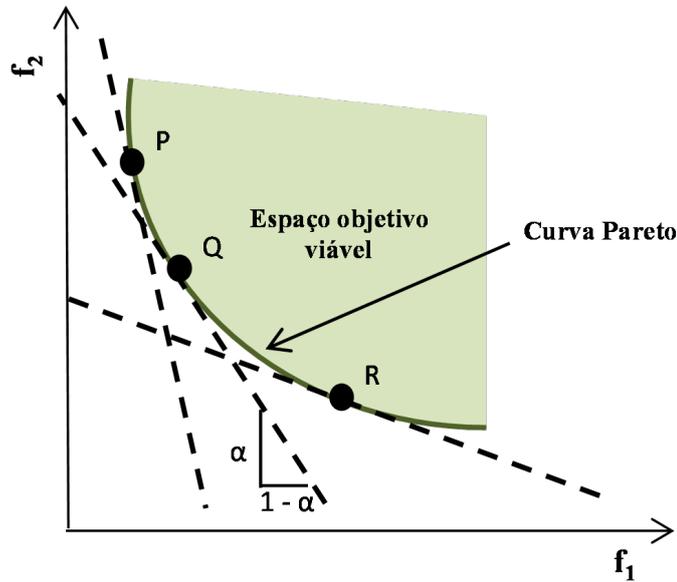


Figura 2.2: Problema com curva Pareto convexa
Fonte: Adaptado de Das and Dennis (1997).

Apesar de obter alguns pontos da curva Pareto (R e M), a solução P da Figura 2.3 nunca será uma solução obtida pelo problema resultante da combinação linear destes objetivos, pois, a tangente à curva neste ponto também intercepta Q com uma inclinação diferente da tangente à curva neste outro ponto. Desta forma, a direção de busca do ótimo seria desviada para Q a fim de se obter uma função objetivo menor para o SOP.

Na terceira abordagem busca-se soluções que balanceiem os conflitos dos múltiplos objetivos (Zhou et al., 2011). Para tanto, é preciso definir e utilizar um método para avaliar e ordenar as soluções geradas, dentre os quais, destaca-se o critério de dominância, definido na sequência.

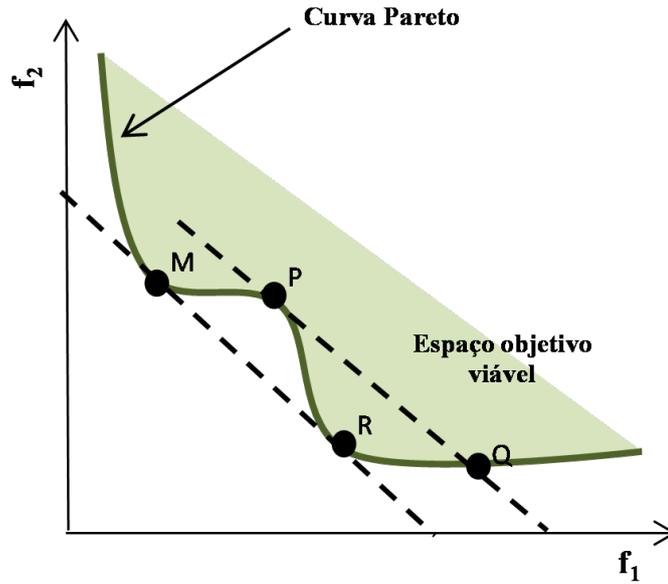


Figura 2.3: Problema com curva Pareto não convexa
 Fonte: Adaptado de Das and Dennis (1997).

Definição 1. Dados dois vetores de solução x^1 e x^2 , com suas respectivas funções objetivo $f(x^1)$ e $f(x^2)$, $f(x^1)$ domina $f(x^2)$, i.e. $f(x^1) \prec f(x^2)$, se em um problema de minimização as seguintes condições forem satisfeitas:

1. $\forall m = 1, 2, \dots, M: f_m(x^1) \leq f_m(x^2)$
2. $\exists m = 1, 2, \dots, M: f_m(x^1) < f_m(x^2)$

Neste caso, $f(x^1)$ domina $f(x^2)$, se $f(x^1)$ não for pior que $f(x^2)$ em todas as funções objetivo, e se $f(x^1)$ é melhor que $f(x^2)$ em pelo menos uma função objetivo.

Segundo Zhou et al. (2011), uma solução viável $x^* \in \Omega$ é Pareto-Ótima se \nexists uma solução $y \in \Omega$ tal que $f(y) \prec f(x^*)$. Ainda de acordo com Zhou et al. (2011), a imagem desse conjunto de soluções no espaço de funções objetivo consiste na fronteira Pareto-Ótima.

Nos MOPs a otimização é conduzida para encontrar ou se aproximar dessa fronteira, retornando ao final da execução do algoritmo soluções mutuamente não dominadas. Na ausência de qualquer outra informação adicional, nenhuma dessas soluções pode ser considerada melhor do que as outras que estão na mesma fronteira (Deb et al., 2002).

A Tabela 2.1 lista alguns dos Algoritmos Evolucionários Multiobjetivos (*Multiobjective Evolutionary Algorithms - MOEAs*) disponíveis na literatura específica, e apresentam suas principais diferenças.

Tabela 2.1: Algoritmos Genéticos Multiobjetivos

Algoritmo	Classificação	Diversidade	Elitismo	Vantagens	Desvantagens
MOGA (Fonseca and Fleming, 1993)	Ordenação baseada na dominância do conjunto Pareto.	Penalização de indivíduos localizados em fronteiras com elevada densidade de soluções.	Não	Variante simples dos ACGs mono-objetivos.	Dificuldade para convergência do algoritmo.
NSGA (Srinivas and Deb, 1994)	Ordenação baseada no critério de dominância.	Penalização de indivíduos localizados em vizinhanças com elevada densidade de soluções.	Não	Velocidade de convergência.	Dificuldade no estabelecimento de parâmetros para formação dos nichos de soluções.
NPGA (reynolds Horn et al., 1994)	Nenhuma <i>fitness</i> atribuída aos indivíduos.	Penalização de indivíduos localizados em vizinhanças com elevada densidade de soluções.	Não	Processo de seleção simples (Torneio Binário).	Dificuldade no estabelecimento de parâmetros para formação dos nichos de soluções.
PAES (Knowles and Corne, 1999)	Escolha entre indivíduos genitores e descendentes baseada no critério de dominância.	Penalização de indivíduos localizados em nichos/células mais densos do espaço de soluções.	Sim	Eficiência computacional, e facilidade para implementar.	Performance do algoritmo sensível ao tamanho dos nichos/células do espaço de soluções.
SPEA-II (Zitzler et al., 2001)	Ordenação baseada no número de soluções da população externa que são dominadas e que dominam o indivíduo.	Diversidade baseada no método do k-ésimo vizinho mais próximo.	Sim	Garante que as soluções extremas sejam preservadas.	Elevado esforço computacional despendido na avaliação e na manutenção da diversidade dos indivíduos.
NSGA-II (Deb et al., 2002)	Ordenação baseada nos critérios de dominância e diversidade dos indivíduos.	Penalização de indivíduos com menor valor da distância de multidão.	Sim	Não necessita estabelecer um parâmetro para formação de clusters/nichos.	A aplicação da distância de multidão está limitada ao espaço das funções objetivo.
MOEA/D (Zhang and Li, 2007)	Ordenação feita para cada subproblema de otimização escalar.	A diversidade dos subproblemas resultantes da decomposição leva à mesma diversidade da população.	Sim	Esforço computacional razoável requerido na resolução dos subproblemas decompostos.	Metodologia de decomposição sensível ao formato da superfície de resposta do problema.
NSGA-III (Deb and Jain, 2014)	Ordenação baseada nos critérios de dominância e diversidade dos indivíduos.	Criação de nichos para seleção dos pontos ainda pouco representados na população.	Sim	Dispersão das soluções na fronteira Pareto auxiliada pela proposição de pontos de referência, adequada à problemas com mais de três objetivos.	Esforço computacional razoável requerido na definição dos nichos de solução.
RYEA (Cheng et al., 2016)	Classificação conforme os critérios de convergência e diversidade do indivíduo em relação ao vetor de referência ao qual está associado.	A diversidade é mensurada pelo ângulo entre as soluções candidatas e os vetores de referência.	Sim	Estratégia adaptativa e regenerativa de ajuste dos vetores de referência.	Vetores de referência não adaptados às características geométricas da fronteira Pareto-Ótima.

Em Zhang and Xing (2017) os autores dividem a evolução dos MOEAs em três gerações. Na primeira geração (1993-1998), grande parte dos métodos de seleção se basearam no critério de dominância, já as estratégias de diversidade se basearam em mecanismos de compartilhamento da *fitness*, e.g. MOGA, NPGA e NSGA. A segunda geração (1995-2002) é marcada por métodos de seleção baseados em estratégias de arquivamento externo e inclusão do processo de elitismo na evolução dos indivíduos, e.g. PAES, SPEA-II, NSGA-II. Desde 2003, diferentes MOEAs foram propostos (Zhang and Xing, 2017), e.g. MOEA/D, NSGA-III e RVEA .

Nesta pesquisa optou-se por utilizar uma adaptação do algoritmo NSGA-II proposto em Deb et al. (2002). O NSGA é um Algoritmo Genético que utiliza o critério de dominância na ordenação dos indivíduos, permitindo, portanto, a abordagem de problemas de otimização com múltiplos objetivos. O NSGA-II incorpora o processo de elitismo e o conceito de distância de multidão para garantir a diversidade dos indivíduos e a convergência do algoritmo. O detalhamento da implementação deste algoritmo está apresentado no Capítulo 4.

2.2.4 Desempenho dos Algoritmos

A definição de métricas apropriadas para mensuração do desempenho dos MOEAs auxilia na validação e comparação dos algoritmos. Entretanto, conforme destacado em Abraham and Jain (2005), nos MOEAs os seguintes fatores dificultam a aplicação destes indicadores:

1. Nas iterações uma população de indivíduos é avaliada ao mesmo tempo.
2. A natureza estocástica dos MOEAs demanda que múltiplas replicações sejam feitas para adequada mensuração do desempenho.
3. Os MOEAs podem ser avaliados por diferentes aspectos, e.g. diversidade dos indivíduos e convergência dos algoritmos.

As métricas se dividem nas seguintes categorias: (i) qualificam a eficiência do algoritmo em termos do esforço computacional requisitado, (ii) avaliam sua eficácia através da medição de acurácia, robustez, escalabilidade e convergência (Coello et al., 2007).

Segundo Zhou et al. (2011) e Deb (2011) os principais indicadores utilizados mensuram a convergência ou aproximação do algoritmo ao conjunto Pareto-Ótimo e a diversidade das populações de indivíduos.

A Tabela 2.2 apresenta algumas destas métricas, i.e., associam um valor para cada fronteira de soluções não dominadas, refletindo sua qualidade segundo um critério específico (Zitzler et al., 2003).

Tabela 2.2: Métricas para avaliação dos MOEAs

Métrica	Referência	Definição	Observações
<i>Spacing (S)</i>	Schott (1995)	Estimativa do espaçamento das soluções não dominadas sendo analisadas.	O cálculo não fornece informações quanto à convergência para o conjunto Pareto-Ótimo; portanto, não é necessário que este seja conhecido.
<i>Hypervolume (HV)</i>	Zitzler and Thiele (1998)	Medida da área do espaço objetivo coberta pelo conjunto de soluções não dominadas analisadas.	O cálculo requer a definição de um ponto de referência no espaço de busca.
<i>Coverage (C)</i>	Zitzler and Thiele (1998)	Comparação da dominância de dois conjuntos de soluções não dominadas encontrados.	O cálculo não fornece informações quanto à convergência para o conjunto Pareto-Ótimo; portanto, não é necessário que este seja conhecido.
<i>Error Ratio (ER)</i>	Van Veldhuizen (1999)	Proporção de um conjunto de soluções não dominadas encontradas que participam do conjunto Pareto-Ótimo.	O cálculo requer que o conjunto Pareto-Ótimo seja conhecido.
<i>Generational Distance (GD)</i>	Van Veldhuizen (1999)	Distância entre o conjunto de soluções não dominadas encontradas e as soluções mais próximas inseridas no conjunto Pareto-Ótimo.	O cálculo requer que o conjunto Pareto-Ótimo seja conhecido.
<i>Inverted Generational Distance (IGD)</i>	Coello and Sierra (2004)	Distância entre as soluções do conjunto Pareto-Ótimo e as soluções não dominadas mais próximas encontradas.	O cálculo requer que o conjunto Pareto-Ótimo seja conhecido.

A principal limitação das métricas listadas consiste na necessidade de se conhecer a fronteira Pareto-Ótima do problema para que sejam calculadas ou estimadas. Entretanto, em problemas reais esta informação geralmente é inacessível. Devido a este motivo, optamos por utilizar o hipervolume para estimar a convergência do algoritmo ao longo de cada geração.

O hipervolume (HV) foi proposto por (Zitzler and Thiele, 1998), e mensura o espaço objetivo coberto pela área formada entre os pontos de um conjunto Pareto e certo ponto deste espaço definido como referência.

Conforme observado na Figura 2.4, quanto maior o valor de HV, melhor é o conjunto Pareto, pois, um alto valor de HV indica que houve boa dispersão ou diversidade das soluções, além de boa convergência do algoritmo.

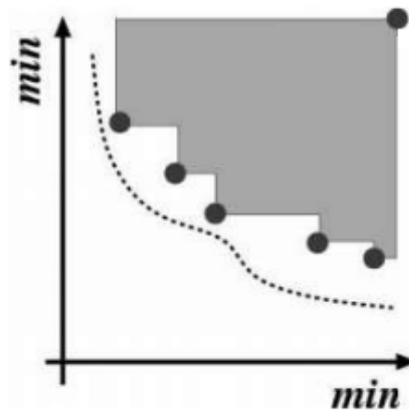


Figura 2.4: Cálculo do hipervolume de um conjunto Pareto
Fonte: Adaptado de Fonseca et al. (2005)

2.3 Otimização via Simulação

Segundo Lee et al. (2008), alguns fatores associados aos problemas de otimização podem dificultar sua modelagem analítica: (i) aleatoriedade dos parâmetros de entrada, (ii) dinâmica não linear, e (iii) existência de múltiplos e conflitantes objetivos. Ainda de acordo com esses autores, apesar de muito utilizada na abordagem destes sistemas, a Simulação a Eventos Discretos é capaz de avaliar apenas alternativas indicadas, o que pode ser considerado insuficiente do ponto de vista da otimização. Nestes casos, a integração entre estas diferentes abordagens pode ser uma solução mais adequada.

A Simulação a Eventos Discretos é apenas uma das técnicas usadas para análise e melhoria das redes de saúde. Outros métodos de simulação também são usados nesta área de estudo, e.g.,

Teoria das Filas e Simulação Baseada em Agentes.

Em Belciug and Gorunescu (2016) os autores propõem um Algoritmo Genético Híbrido com modelo de filas para estudar a alocação de pacientes e otimizar os custos e taxas de recusa. Existem diversas diferenças entre este trabalho e o proposto nesta pesquisa. Primeiramente, em Belciug and Gorunescu (2016) os autores utilizam um modelo de filas para representar o sistema em estudo. Além disso, enquanto eles analisam apenas pacientes de geriatria em um hospital específico, nós analisamos diversas especialidades e localidades. Outro ponto importante consiste na simplificação de parâmetros estocásticos, como o tempo de internação e o intervalo de tempo entre as chegadas dos pacientes. Em Belciug and Gorunescu (2016) os autores utilizam um intervalo de 30% de variância em torno da média, enquanto nesta pesquisa nós atribuímos uma distribuição de probabilidade para cada especialidade.

Em Jun et al. (1999), Jacobson et al. (2006) e Günal and Pidd (2010) diversas aplicações da Simulação a Eventos Discretos na área da saúde são apresentadas.

Em Jun et al. (1999) e Jacobson et al. (2006) os autores estudam o impacto da alocação de recursos hospitalares e programação das internações na eficiência das redes de saúde. Além disso, os autores fazem uma revisão de estudos que buscam otimizar a alocação de recursos humanos e materiais na área da saúde.

Em Günal and Pidd (2010) as aplicações são divididas nas seguintes categorias: emergência, internação, ambulatório, e diversas outras unidades como as UTIs e os centros de análises clínicas. Os autores afirmam que pelo fato de vários estudos terem sido feitos para sistemas ou facilidades específicas, as abordagens aplicadas são dificilmente generalizáveis. Além disso, estas pesquisas costumam fazer suposições muito simplistas da influência e relações que existem entre as diferentes facilidades dos sistemas estudados.

Os múltiplos *outputs* retornados, e a capacidade de representar estas relações, tornam os modelos de Simulação à Eventos Discretos adequados para abordar este tipo de problema (Jun et al., 1999).

Até o ano 2000, as técnicas de otimização e simulação foram mantidas praticamente separadas, mesmo com todas as pesquisas relevantes para combiná-las (Fu, 2002). Nas últimas décadas, porém, diversas rotinas de otimização foram incluídas em pacotes de simulação. De forma geral, o objetivo destas rotinas é buscar a melhor configuração dos sistemas simulados, de forma a maximizar ou minimizar um ou múltiplos critérios de interesse (Lin et al., 2013).

2.3.1 Conceitos Básicos

Segundo Pierreval and Paris (2000) a integração destas técnicas pode ser feita de duas maneiras. A primeira é baseada na otimização da superfície de resposta do problema, definida pela simulação de um conjunto de soluções de interesse. E a segunda consiste em explorar e simular de maneira iterativa as soluções candidatas do problema. Neste caso, é preciso garantir conexão direta entre o algoritmo de otimização e a simulação.

Conforme Figura 2.5, nesta metodologia a otimização repassa um design inicial ao modelo de simulação. Este design é então simulado e a mensuração do desempenho desta configuração é retornada à otimização, que utiliza esta informação para definir novas direções de busca do algoritmo. Esse processo se repete até que um critério de parada seja alcançado (Fu, 2013).

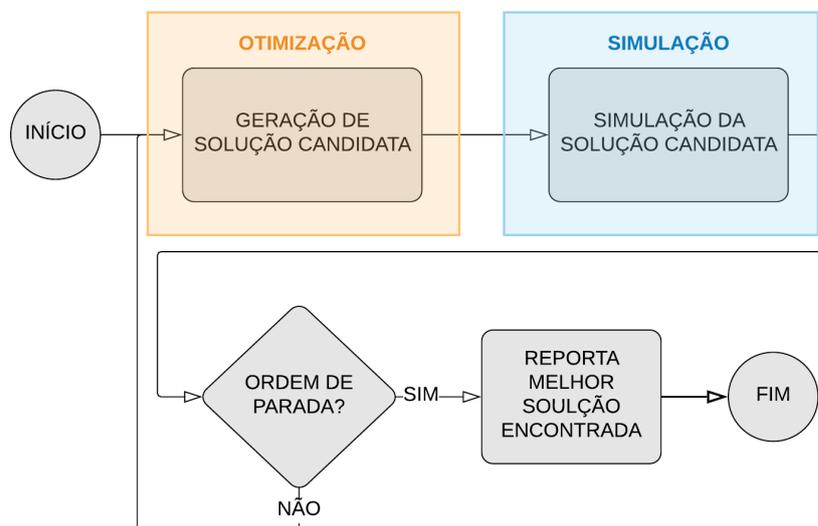


Figura 2.5: Interação entre os pacotes de otimização e simulação
Fonte: Adaptado de Kelton and Law (2000).

A interação entre essas técnicas, também chamada de Otimização via Simulação, consiste, portanto, na otimização de variáveis de desempenho medidas por modelos estocásticos de Simulação a Eventos Discretos (Fu, 2001).

Em Juan et al. (2015) os autores listam diversas aplicações da abordagem de Otimização via Simulação no campo da saúde (e.g. Kasaie and Kelton, 2013), produção (e.g. Gansterer et al., 2014) e logística (e.g. Alizadeh et al., 2011).

Na área da saúde, Kasaie and Kelton (2013) sugerem a interação de técnicas de otimização e simulação na alocação de recursos limitados destinados ao controle de epidemias. Na pesquisa os autores propõem um modelo de Simulação Baseada em Agentes para reproduzir o comportamento

dinâmico do sistema estudado sob determinadas configurações.

Segundo Swisher et al. (2000), em um modelo de simulação com p *inputs* $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_p\}$, e q *outputs* $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_q\}$, onde Φ é função de Ψ , $\Phi = \Phi(\Psi)$, o objetivo da Otimização via Simulação seria definir para quais valores de Ψ a função $\Phi(\Psi)$, resposta da simulação, é otimizada.

Ao contrário da abordagem de programação matemática, na Otimização via Simulação a modelagem analítica dos objetivos e restrições do sistema simulado nem sempre é viável. Nestes casos, a simulação funciona como uma caixa preta que avalia as funções do problema para determinado conjunto de *inputs* (Amaran et al., 2014).

Devido ao elevado esforço requerido nesse processo, Amaran et al. (2014) ressaltam a importância de se considerar as variáveis de entrada mais relevantes ao problema, reduzir ao máximo seus intervalos de variação e fornecer boas soluções iniciais ao algoritmo.

2.3.2 Limitações da Otimização via Simulação

O esforço requerido no processo de avaliação das soluções via simulação poderá exceder o desejável na resolução de um modelo linear de médio porte, configurando uma dicotomia entre o tempo gasto na geração e análise das soluções candidatas (Fu, 2001).

Além disso, segundo Banks et al. (2000), a principal dificuldade da Otimização via Simulação consiste na natureza estocástica dos sistemas modelados, pois, nestes casos, a performance de determinada configuração é estimada, e não definida de maneira exata. Desta forma, é impossível assegurar que uma alternativa é melhor que outra, reduzindo a eficiência dos algoritmos de otimização, que buscam mover-se em direções de melhoria. A utilização de diversas replicações, ou simulação de extensos períodos, está entre as principais estratégias utilizadas na redução da variância do critério de desempenho otimizado. Entretanto, nestes casos, pouquíssimas alternativas poderiam ser exploradas devido ao elevado esforço computacional requerido.

Fica claro, então, a importância de se pesquisar outras abordagens de Otimização via Simulação mais eficientes e eficazes, i.e. que analisem o mínimo de soluções possível e mapeiem grande parte da fronteira Pareto-Ótima (Lin et al., 2013).

Se ainda assim a simulação das soluções propostas pela otimização exigir grande esforço computacional, outras estratégias podem ser usadas para contornar esta limitação, e.g. definição de superfícies de resposta ou metamodelos.

Os metamodelos consistem em formulações simplificadas que representam a lógica de um

sistema, implícita no código e estrutura dos modelos originais de simulação. Segundo Barton (1992), os metamodelos são aplicados nas diversas etapas dos projetos de simulação, e.g., na identificação dos parâmetros mais impactantes na performance de um sistema, ou na condução de cenários para analisar a dinâmica e o comportamento de um sistema em diferentes configurações.

Os metamodelos buscam um melhor entendimento das relações existentes entre *inputs* e *outputs* na otimização das superfícies de resposta dos problemas estudados (Simpson et al., 2001).

Entretanto, segundo Banks et al. (2000), a existência de um quantitativo significativo de configurações ou combinações, e a diversidade do conjunto de variáveis (discretas ou contínuas, quantitativas ou qualitativas) dificulta a definição de um metamodelo confiável.

Portanto, nesta pesquisa buscou-se outra estratégia para redução do tempo gasto na simulação das soluções propostas pela otimização. Neste caso, as replicações e experimentos requisitados foram simulados de maneira simultânea em computadores conectados, técnica conhecida como Simulação ou Computação Paralela.

Segundo Pierreval and Paris (2000), os AGs apresentam características intrínsecas que facilitam essa paralelização em diferentes processadores. Existem duas formas de paralelizar estes algoritmos:

1. **Método Global:** analisa diferentes conjuntos de indivíduos da população em paralelo, deixando os operadores e processos evolutivos para o programa "mestre", Figura 2.6.
2. **Estratégia de Granulação:** a população é dividida em distintos conjuntos, os operadores e processos evolutivos são conduzidos pelos próprios algoritmos de otimização e simulação locais. Neste caso, um operador é definido para migração dos indivíduos entre subpopulações, garantindo a propagação de soluções eficientes e evitando a convergência prematura do algoritmo de otimização, Figura 2.7.

Nesta pesquisa foi utilizado o Método Global de computação paralela. A otimização, conduzida pelo algoritmo NSGA-II, é feita pelo programa mestre, e apenas a etapa de simulação das soluções candidatas é paralelizada.

Em Banks et al. (2000), os autores afirmam que fatores randômicos intrínsecos à estratégia de otimização de heurísticas robustas ¹, como os AGs, reforçam sua tolerância quanto à variação

¹Uma heurística é dita robusta quando sua eficácia não depende da estrutura do problema. Os algoritmos desta natureza podem ser aplicados em problemas com variáveis de diferentes domínios, sendo pouco sensíveis à variação destes *inputs* (Hong and Nelson, 2009).

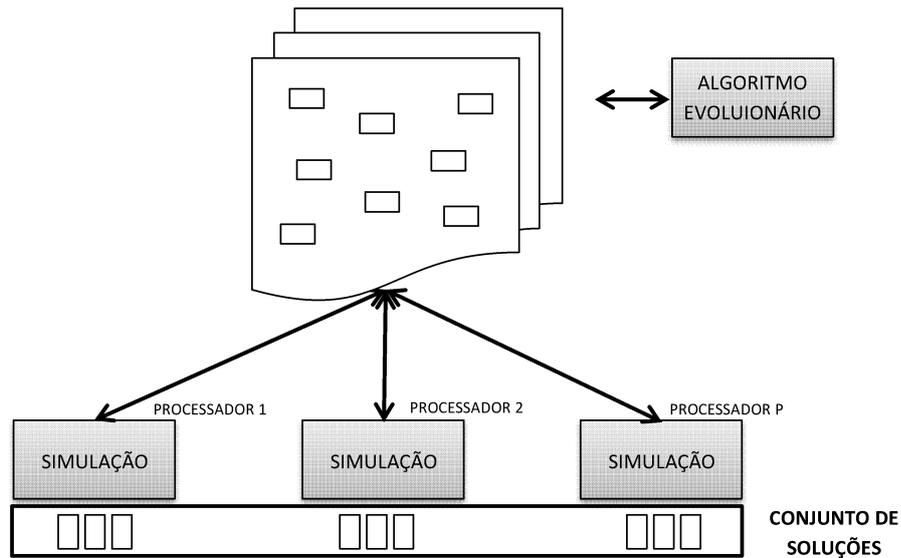


Figura 2.6: Método Global para computação paralela dos Algoritmos Evolucionários
 Fonte: Adaptado de Pierreal and Paris (2000).

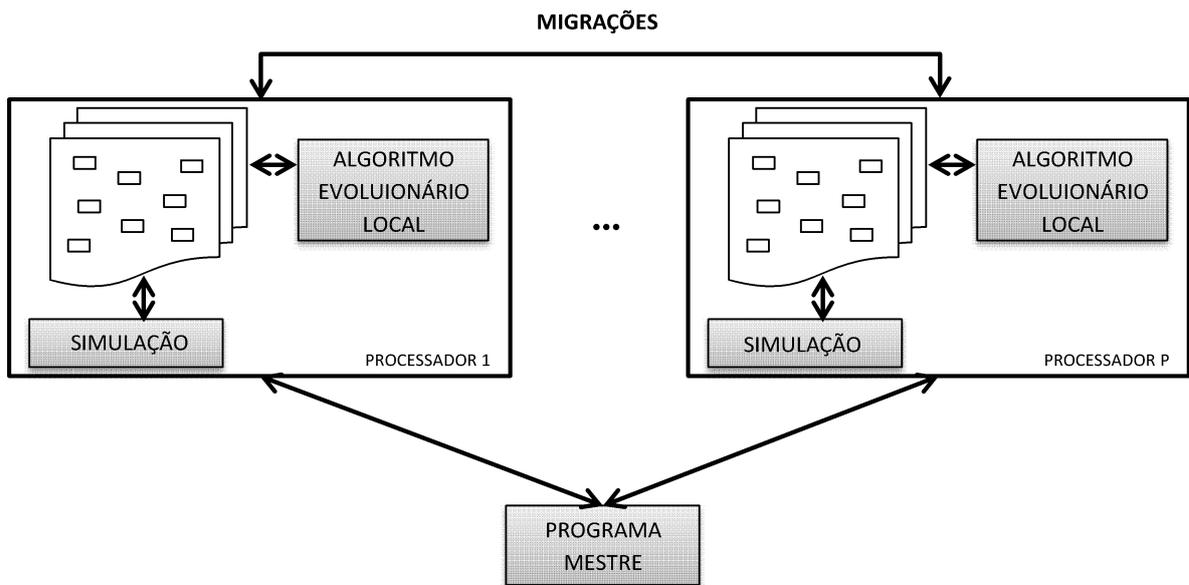


Figura 2.7: Estratégia de Granulação para computação paralela dos Algoritmos Evolucionários
 Fonte: Adaptado de Pierreal and Paris (2000).

das medidas de desempenho retornadas pela simulação, garantindo que resultados satisfatórios sejam alcançados sem ser necessário simular várias e extensas replicações.

As abordagens que promovem a iteração entre modelos de simulação e heurísticas são conhecidas como *Simheuristics*. Elas permitem que as incertezas dos problemas estocásticos de otimização combinatória sejam tratadas de maneira natural. Além disso, elas facilitam as análises de sensibilidade e confiabilidade dos critérios de otimização durante a avaliação de soluções candidatas (Juan et al., 2015).

Conforme detalhado no Capítulo 3, para cada solução a simulação retornará à otimização as taxas de recusa dos leitos, usadas para avaliar e comparar as soluções geradas. Essas taxas são calculadas via simulação, pois suas formulações matemáticas implicariam em diversas suposições irrealistas do problema estudado.

Neste caso, o algoritmo NSGA-II retornará não apenas uma, mas sim um conjunto de soluções não dominadas. Ao esforço da simulação restaria a validação dessas soluções, e o apoio na tomada de decisão quanto à escolha da solução não dominada mais adequada ao problema analisado.

Capítulo 3

Formulação Matemática

Conforme mencionado, nesta pesquisa propomos uma abordagem de Otimização via Simulação ao problema de alocação de leitos de internação. O processo de otimização é conduzido pelo Algoritmo Genético NSGA-II, e as soluções geradas são avaliadas via Simulação a Eventos Discretos. As avaliações são retornadas à otimização, que utiliza esta informação para comparar e ordenar as soluções geradas e definir novas direções de busca do algoritmo. Esse processo se repete até que um critério de parada seja alcançado.

A abordagem proposta é aplicada em uma das macro-regiões de saúde de Minas Gerais. Entretanto, o universo de aplicações não deve ser limitado ao sistema estudado.

Basicamente, investigamos se realocando os leitos já existentes ou contratando outros leitos do sistema privado seria possível atender a demanda com custos e taxas de recusa mínimos.

Desta forma, o número máximo de leitos que cada local poderá receber é dado pela soma dos leitos privados e públicos já existentes nestas localidades. O problema é, portanto, capacitado, e por isto nem sempre será possível encontrar uma solução que garanta que 100% dos pacientes alocados sejam atendidos, i.e. o paciente será recusado caso o leito em que foi alocado não esteja disponível no momento da internação.

Na sequência será detalhada a notação e a formulação matemática proposta ao problema estudado.

3.1 Conjuntos

O : pontos de origem ou residência dos pacientes, $O = \{1, \dots, no\}$.

D : pontos de destino ou atendimento dos pacientes, formado pelos locais onde já existem estru-

turas de internação operantes $D = \{1, \dots, nd\}$.

E : especialidades demandadas pelos pacientes, $E = \{1, \dots, ne\}$.

L : tipos de leito $L = \{1, \dots, nl\}$. Foram considerados os seguintes tipos de leito:

3.2 Variáveis de Decisão

A formulação matemática proposta busca otimizar a capacidade de atendimento do sistema público de internação estudado, e a alocação dos pacientes às unidades de atendimento, expressos pelas seguintes variáveis de decisão:

$x_{o,d,e}$: percentual da demanda do município o pela especialidade tipo e alocado à cidade d , onde $0 \leq x_{o,d,e} \leq 1$.

$y_{d,l}$: número de leitos do tipo l ofertados no município d , onde $y_{d,l} \in \mathbb{Z}^+$.

3.3 Parâmetros

A formulação e modelagem do problema estudado estão sujeitas ao levantamento dos seguintes parâmetros:

nl : vetor do número de leitos ofertados nos municípios (públicos + privados), $nl_d \in \mathbb{Z}^+$, $\forall d \in D$.

MC : matriz binária que representa a compatibilidade entre os leitos e especialidades, definida na Tabela 3.1, $mc_{l,e} \in [0,1]$, $\forall l \in L$, $e \in E$.

MP : matriz binária das relações de precedência entre os leitos, definida na Tabela 3.2, $mp_{l,l'} \in [0,1]$, $\forall l, l' \in L$.

DE : matriz de demanda dos municípios pelas especialidades, $de_{o,e} \geq 0$, $\forall o \in O$, $e \in E$.

DI : matriz de distância entre os pontos do sistema de internação, $di_{o,d} \geq 0$, $\forall o \in O$, $d \in D$.

dm : vetor da distância máxima permitida para deslocamento dos pacientes por especialidade demandada, $dm_e \geq 0$, $\forall e \in E$.

tp : vetor do tempo médio de internação dos pacientes por especialidade demandada, $TP_e \geq 0$, $\forall e \in E$.

$om(y_{d,l})$: ocupação máxima desejável de um *pool* de leitos.

O cálculo da função $om(y_{d,l})$ é baseado na formação de conjuntos de recursos (*pools*), definidos para cada tipo de leito hospitalar. Conforme sugerido em Jones (2011), cada *pool* deve respeitar uma ocupação máxima desejável, definida por sua capacidade, e não por valores médios ou globais do sistema.

Tabela 3.1: Compatibilidade entre os leitos de internação e especialidades

Tipo de Leito	Especialidades Compatíveis
Cirúrgico	- Cirurgia - 28 dias a 14 anos - Cirurgia - 15 a 59 anos - Cirurgia - 60 anos e mais
Clínico	- Clínica - 28 dias a 14 anos - Clínica - 15 a 59 anos - Clínica - 60 anos e mais
Obstétrico	- Obstetrícia
Pediátrico	- Clínica - 28 dias a 14 anos
Neonatal	- Neonatologia
UTI Adulto	- Cirurgia - 15 a 59 anos - Cirurgia - 60 anos e mais - Clínica - 15 a 59 anos - Clínica - 60 anos e mais - Obstetrícia
UTI Pediátrico	- Cirurgia - 28 dias a 14 anos - Clínica - 28 dias a 14
UTI Neonatal	- Neonatologia

Tabela 3.2: Relações de precedência entre os leitos

Tipo de Leito	Relações de Precedência
Cirúrgico	-
Clínico	-
Obstétrico	-
Pediátrico	-
Neonatal	Obstétrico
UTI Adulto	Cirúrgico, Clínico ou Obstétrico
UTI Pediátrico	Pediátrico
UTI Neonatal	Neonatal

Em Jones (2011), os autores também sugerem que *pools* menores mantenham taxas mais baixas de ocupação, buscando reduzir as taxas de recusa desses leitos.

Baseados nestas informações, Pinto et al. (2014) definem o conjunto de Equações (3.1) para calcular a ocupação máxima desejável nos leitos de forma que suas taxas de recusa não ultrapassem 1%.

$$\begin{aligned}
 om(y_{d,l} \leq 100) &= 18,726 \times \ln(y_{d,l}) - 9,4039 \\
 om(y_{d,l} > 100) &= 7,4508 \times \ln(y_{d,l}) + 42,484
 \end{aligned}
 \tag{3.1}$$

O conjunto de Equações (3.1) foi obtido por análise de regressão das informações amostradas em Jones (2011). Os demais parâmetros foram coletados a partir do banco de dados do SUS (DATASUS, disponível em www.datasus.gov.br). Os dados são do ano de 2017. Este período

foi utilizado para exemplificar a aplicação da metodologia proposta. Todos os dados utilizados nesta pesquisa poderão ser disponibilizados e encaminhados pela autora via e-mail ¹

3.4 Funções Objetivo

Na formulação proposta nesta pesquisa dois objetivos de otimização foram considerados: redução das taxas de recusa (f_1) e custos dos leitos (f_2).

Como f_1 e f_2 apresentam diferentes unidades de medida e distintas ordens de grandeza, os objetivos do problema foram normalizados.

f_1 : função taxa de recusa máxima, Equação (3.2), definida via simulação pela recusa máxima dos leitos que são ofertados na macro-região estudada.

$$f_1 = \max_{l \in L} (tr_l) \quad (3.2)$$

f_2 : função custo dos leitos. Conforme os princípios da economia de escala, quanto maior um *pool* de leitos, menores serão seus custos unitários. Isto se deve ao fato de que os leitos de um mesmo *pool* dividem recursos em comum. Conseqüentemente, os custos destes recursos também são divididos entre os leitos. A função definida na Equação (3.3) expressa esta lógica.

$$cl_l(n) = cl_l(1) \times n^{-\gamma} \quad (3.3)$$

E que $cl_l(n)$ é o custo do leito l em um *pool* com n leitos deste mesmo tipo, e $\gamma > 0$ é a taxa pela qual o custo decai à medida que n aumenta.

Neste caso, quanto maior o número de leitos no *pool*, menor será o custo para manter cada unidade desse recurso (cl_l).

Caberá ao tomador de decisão definir o parâmetro γ baseado na suavidade de decaimento da curva de custo de um leito. Nós assumimos que este custo seria reduzido em até 60% do custo unitário para *pools* com 100 leitos, para *pools* maiores este custo permaneceria inalterado, i.e. $cl_l(y_{d,l} \geq 100) = 0,6 \times cl_l(1)$. Após resolver esta equação, nós definimos que $\gamma = 1,11E^{-1}$. Quando disponíveis outros valores também podem ser usados sem prejudicar a aplicabilidade da metodologia proposta.

Conforme esperado, os custos diários dos leitos também dependem da tipologia do recurso.

¹Bárbara Regina Pinto e Oliveira: barbara.brpo@gmail.com

Em geral, os leitos UTI têm custos maiores que leitos padrões.

Nesta pesquisa este custo foi estimado para cada tipo de leito, dividindo-se o custo total das internações registradas em 2017 pelo total de diárias e ocupação média destes leitos, Tabela 3.3. Apesar desta estimativa estar suscetível a erros, ela nos fornece uma boa referência, pois, nenhuma informação real está disponível no banco de dados do SUS, e o valor obtido atende bem aos objetivos propostos nesta pesquisa.

Tabela 3.3: Custo médio diário de internação nos leitos

Tipo de leito (l)	$cl_l(1)$
Cirúrgico	R\$ 425,23
Clínico	R\$ 106,17
Obstétrico	R\$ 295,54
Pediátrico	R\$ 139,91
Neonatal	R\$ 154,31
UTI Adulto	R\$ 547,06
UTI Pediátrico	R\$ 577,76
UTI Neonatal	R\$ 567,05

A partir do exposto, a função f_2 é calculada conforme Equação (3.4).

$$f_2 = \sum_{d \in D, l \in L} y_{d,l} \times cl_l(y_{d,l}) \quad (3.4)$$

Definidos os conjuntos, parâmetros, e variáveis do problema estudado, foi possível propor uma modelagem matemática biobjetivo para representá-lo, detalhada na sequência.

3.5 Modelagem Matemática

$$\text{Min } \mathbf{f} = \{f_1, f_2\} \quad (3.5)$$

s.a

$$\sum_{l \in L} y_{d,l} \leq nl_d \quad \forall d \in D \quad (3.6)$$

$$y_{d,l} \leq \sum_{l' \in L, l' \neq l} mp_{l,l'} \times y_{d,l'} \quad \forall d \in D, l \in L \quad (3.7)$$

$$x_{o,d,e} \leq \sum_{l \in L} mc_{l,e} \times y_{d,l} \quad \forall d \in D, o \in O, e \in E \quad (3.8)$$

$$\left(\frac{di_{o,d}}{dm_e} - 1\right) \times x_{o,d,e} \leq 0 \quad \forall d \in D, o \in O, e \in E \quad (3.9)$$

$$\sum_{d \in D} x_{o,d,e} = 1 \quad \forall o \in O, e \in E \quad (3.10)$$

$$\sum_{\substack{o \in O, \\ e \in E}} \frac{mc_{l,e} \times de_{o,e} \times tp_e \times x_{o,d,e}}{365} \leq y_{d,l} \times om(y_{d,l}) \quad \forall d \in D, l \in L \quad (3.11)$$

$$0 \leq x_{o,d,e} \leq 1 \quad \forall o \in O, d \in D, e \in E \quad (3.12)$$

$$y_{d,l} \in \mathbb{Z}^+ \quad \forall d \in D, l \in L \quad (3.13)$$

Na formulação apresentada, a Equação (3.5) indica a função objetivo do problema, que consiste em minimizar a recusa máxima e o custo diário dos leitos à disposição dos sistema público.

As soluções propostas pelo método de otimização empregado devem respeitar um conjunto de restrições do modelo de forma a atender de maneira eficiente a demanda existente.

Conforme Equação (3.6), o número de leitos ofertados no município d deve ser menor que o total de leitos públicos e privados disponíveis nesta localidade. A Equação (3.7) indica que alguns leitos só podem ser alocados nos municípios caso haja previsão de oferta de outros leitos neste local que justifiquem esta alocação, e.g. caso não haja previsão de oferta de leitos obstétricos, nenhum leito neonatal será esperado no município. A Equação (3.8) limita a destinação de um paciente residente em o apenas a locais d que ofertem leitos compatíveis com sua especialidade.

Segundo a Equação (3.9), os pacientes residentes no município o poderão ser alocados na cidade d , somente se a distância entre esses municípios for inferior à máxima permitida para deslocamento dos pacientes com especialidade e . A Equação (3.10) garante que 100% dos pacientes com especialidade e residentes no município o sejam alocados. O que não significa que 100% dos pacientes serão admitidos, dado que esta admissão dependerá da disponibilidade dos leitos que são compatíveis com sua especialidade. Neste caso, as internações rejeitadas serão penalizadas pela função f_1 .

Segundo a Equação (3.11), o número de leitos l ofertados em um município d precisa ser suficiente para atender de maneira eficiente a demanda alocada e manter um nível de ocupação desejável (Pinto et al., 2014).

A inclusão desta restrição ao problema ajuda a refinar as soluções geradas pelo NSGA-II, evitando a avaliação e simulação de diversas soluções inviáveis ou ineficientes do ponto de vista da otimização.

Finalmente, as Equações (3.12) e (3.13) restringem o domínio das variáveis de decisão do problema.

A otimização da formulação biobjetiva proposta retorna um conjunto de soluções não dominadas. Ao esforço da simulação restaria ainda a escolha da solução mais adequada desse conjunto baseada em um critério de tomada de decisão.

Pela formulação apresentada pode-se pensar que seria mais fácil resolver o problema analisado caso ele fosse dividido em nl subproblemas, i.e. caso fosse resolvido para cada tipo de leito l . Entretanto, os parâmetros $mc_{l,e}$ e $mpl_{l,e}$ revelam que alguns leitos são compatíveis com duas ou mais especialidades, e que existem relações de precedência que devem ser respeitadas, justificando a necessidade da alocação simultânea desses recursos.

Além disso, a restrição de capacidade definida na Equação (3.6) está discriminada para cada município. Portanto, baseado no total de leitos ofertados em cada local (nl_d), o algoritmo define quantos leitos devem permanecer disponíveis para o SUS, quantos leitos extras devem ser contratados do sistema privado de saúde deste local, e como o total de leitos ofertados devem estar distribuídos entre as várias especialidades disponíveis.

Portanto, para garantir uma representação mais fidedigna destas relações e restrições do problema, é preciso analisar os leitos de forma conjunta.

Outra peculiaridade da formulação matemática proposta consiste na estratégia de alocação dos pacientes. Diferentemente das abordagens mais clássicas apresentadas na revisão de litera-

tura, neste estudo os pacientes poderão ser alocados a cidades mais distantes, desde que seja satisfeita a restrição de distância máxima permitida para atendimento.

Além disso, grande parte destas abordagens na literatura considera que cada ponto da rede seja alocado à apenas uma facilidade ou instalação, i.e. a variável de alocação $x_{o,d,e}$ é definida como sendo binária. Apesar de comum, esta suposição não representa a realidade de países pouco desenvolvidos como o Brasil, onde nem sempre é possível garantir os recursos necessários para atender a toda a demanda das cidades mais próximas. Para exemplificar este caso, a Tabela 3.4 mostra o fluxo de pacientes cirúrgicos residentes em União de Minas e atendidos dentro da macro-região de saúde.

Tabela 3.4: Atendimento de pacientes cirúrgicos de União de Minas na rede SUS

Fonte: DATASUS (2017)

Cidades	Percentual	Leitos	Distância
Uberaba	45.10%	293	317 km
Iturama	31.37%	14	39 km
União de Minas	23.53%	9	-

Apesar de União de Minas atender a cerca de 24% da necessidade local por leitos cirúrgicos, alguns pacientes residentes deste município foram encaminhados e internados em outras cidades. Verifica-se que 45% dos pacientes foram internados em Uberaba, mesmo esta cidade estando mais distante que Iturama que também recebeu pacientes de União de Minas. Portanto, os pacientes nem sempre são alocados às cidades mais próximas.

Ainda assim, pode-se questionar como esta alocação se daria na prática sendo a variável $x_{o,d,e}$ contínua, i.e. como definir quais pacientes receberão atendimento local e quais serão encaminhados para internação em outras cidades. Assim como em outros problemas de otimização, as decisões de localização e alocação de facilidades são hierarquicamente classificadas nas seguintes categorias: (i) estratégicas de longo prazo, e.g. planejamento ou design das cadeias de suprimento, (ii) táticas ou de médio prazo, e.g. alocação de capacidade operativa em instalações já existentes (iii) operacionais ou de curto prazo, e.g. sequenciamento de atividades ou processos em facilidades já operantes.

Nesta pesquisa apenas as decisões táticas foram abordadas. A principal premissa é averiguar se ao realocar leitos públicos existentes ou contratar novos leitos do sistema privado seria possível atender a demanda a níveis mínimos de recusa e custos. Desta forma, decisões de longo e curto prazo, e.g. definição de onde abrir os hospitais, e quando deve-se encaminhar ou não um paciente, estão fora do escopo deste estudo. As decisões de longo prazo já foram tomadas, dado que este

sistema já se encontra em operação, já as decisões de curto prazo deverão ser tomadas à luz dos resultados obtidos nesta pesquisa, i.e. a partir da demanda esperada e a capacidade planejada para cada município. Para informações adicionais acerca de decisões de curto prazo na área da saúde, o leitor poderá consultar os trabalhos de Marynissen and Demeulemeester (2016) e Bastos et al. (2019).

Capítulo 4

Metodologia

Este capítulo apresenta o processo de Otimização via Simulação conduzido pelo algoritmo NSGA-II, implementado no *software* MATLAB® e dividido nas seguintes etapas:

1. Representação dos indivíduos na população;
2. Geração da primeira população de indivíduos;
3. Avaliação e comparação dos indivíduos;
4. Operações de evolução e elitismo.

A escolha deste algoritmo foi baseada no fato dele (i) utilizar o princípio de elitismo, evitando que soluções eficientes sejam perdidas nos processos evolutivos, (ii) sugerir um mecanismo de preservação da diversidade dos indivíduos, (iii) aplicar o critério de dominância na avaliação e ordenação das soluções, permitindo a abordagem de múltiplos objetivos.

O fluxograma disposto na Figura 4.1 indica que o algoritmo NSGA-II segue as etapas de um Algoritmo Genético clássico, entretanto, com importantes alterações nos processos de ordenação e comparação dos indivíduos, detalhadas na sequência.

Além disso, a avaliação dos indivíduos, representada pelas linhas pontilhadas na Figura 4.1 e destacada na Figura 4.2, incorpora a metodologia de Otimização via Simulação. Esta metodologia foi escolhida para contornar a dificuldade de se representar analiticamente a recusa máxima dos leitões, i.e. f_1 , métrica usada pelo NSGA-II para avaliação, comparação e ordenação das soluções.

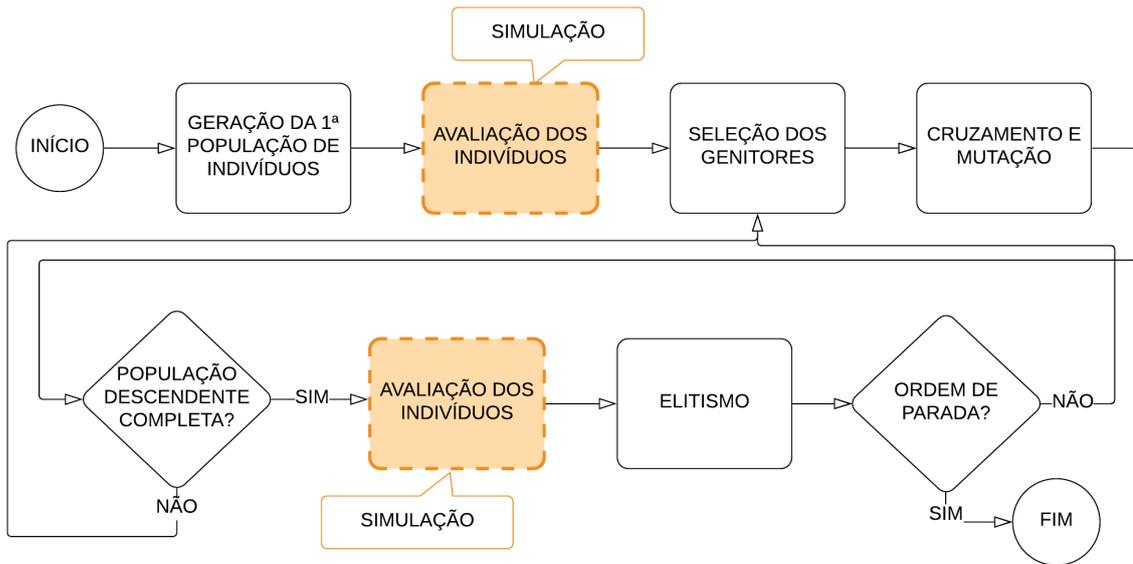


Figura 4.1: Fluxograma de implementação do Algoritmo NSGA-II

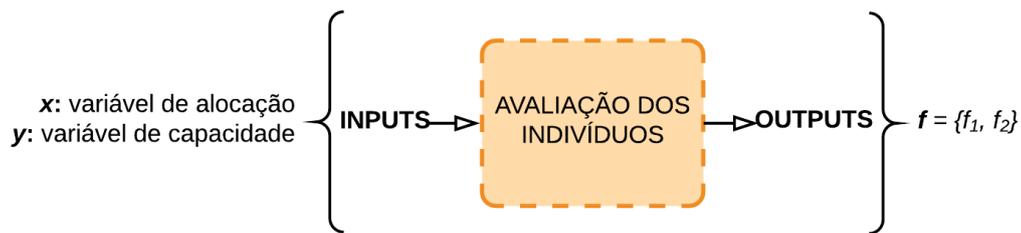


Figura 4.2: Detalhamento da avaliação dos indivíduos no NSGA-II

4.1 Representação dos Indivíduos na População

Conforme apresentado, na formulação matemática proposta nesta pesquisa, Equações (3.5) - (3.13), a variável de decisão $x_{o,d,e}$ é contínua, e representa a alocação dos pacientes nos municípios com sistemas de internação, enquanto a variável $y_{d,l}$ é discreta, e define o número de leitos disponíveis nestes municípios. Devido as diferenças de domínio e dimensão dessas variáveis torna-se difícil integrá-las na representação dos indivíduos, portanto, neste trabalho optou-se pela decomposição da otimização em duas fases. Neste caso, o número de leitos $y_{d,l}$ é otimizado primeiro, e a partir desta etapa é definida a alocação dos pacientes a esses leitos.

A Figura 4.3 indica a estrutura proposta para representar a população de indivíduos na primeira etapa descrita do processo de otimização.

Na estrutura matricial proposta à população de indivíduos, os elementos $p_{d,l,i}$ definem o

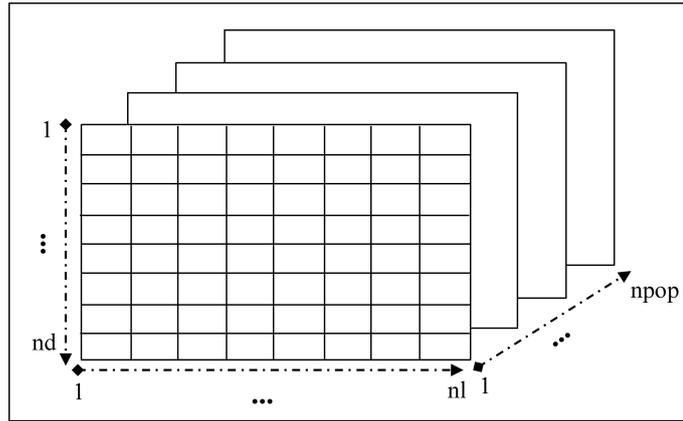


Figura 4.3: Representação dos indivíduos no NSGA-II

número de leitos l disponíveis no município d na solução proposta pelo indivíduo i , onde $d \in D = \{1, \dots, nd\}$, $l \in L = \{1, \dots, nl\}$ e $i \in I = \{1, \dots, npop\}$.

4.2 Geração da Primeira População de Indivíduos

Em Hong and Nelson (2009) os autores sugerem a adoção de diferentes estratégias para geração da primeira população de indivíduos:

- Seleção de pontos extremos do espaço de soluções viáveis.
- Seleção de pontos centrais do espaço de soluções viáveis.
- Geração aleatória de soluções ou indivíduos.
- Geração de soluções através da consulta à profissionais ou literatura relacionada.

Neste estudo, a geração da primeira população de indivíduos é feita de forma aleatória, respeitando, entretanto, a restrição do número máximo de leitos que podem ser ofertados nos municípios definida pela Equação (3.6), e a restrição de precedência entre os tipos de leito definida pela Equação (3.7).

Nesta etapa do processo são geradas $npop$ soluções da variável de otimização $y_{d,l}$ que dimensiona o número de leitos a serem ofertados nos municípios do sistema público de internação estudado.

4.3 Alocação dos Pacientes

Conforme mencionado, nesta pesquisa optou-se por uma abordagem de otimização em duas-fases. Após dimensionar a capacidade de atendimento dos municípios em um primeiro momento, i.e. definir a matriz Y de variáveis $y_{d,l}$, é preciso otimizar a alocação dos pacientes às localidades que disponham de recursos compatíveis. No Capítulo 3 desta pesquisa, essa alocação é definida pela variável $x_{o,d,e}$, que representa o percentual da demanda do município o pela especialidade tipo e alocado à cidade d .

Para otimizar a alocação dos pacientes, é criada a variável $\xi_{d,l}$. Conforme a Equação (4.1), esta variável mensura a violação da restrição de ocupação máxima desejada nos conjuntos de leitos de internação.

$$\sum_{\substack{o \in O, \\ e \in E}} \frac{mc_{l,e} \times de_{o,e} \times tp_e \times x_{o,d,e}}{365} - \hat{y}_{d,l} \times om(\hat{y}_{d,l}) \leq \xi_{d,l} \quad \forall d \in D, l \in L \quad (4.1)$$

$$\xi_{d,l} \geq 0 \quad \forall d \in D, l \in L \quad (4.2)$$

A inclusão desta variável busca tornar a formulação do problema menos restritiva, relaxando a restrição definida pela Equação (3.11) no modelo original. Segundo Coello et al. (2007), os problemas de otimização são compostos por restrições que devem ser consideradas e satisfeitas. Entretanto, restrições mais brandas podem ser relaxadas, garantindo a ampliação e exploração do espaço de busca do algoritmo, e a avaliação de algumas soluções inviáveis do problema. Isso evita, sua convergência prematura.

As soluções com $\xi_{d,l} > 0$ produzidas pela relaxação são penalizadas pela Equação (4.3), cujo objetivo é minimizar a violação da restrição de ocupação máxima desejada.

$$Min \mathbf{f} = \sum_{\substack{d \in D, \\ l \in L}} \xi_{d,l} \quad (4.3)$$

Após relaxação da restrição definida pela Equação (3.11), o modelo foi então implementado na linguagem de programação AMPL com uso do solver GUROBI. A escolha de um método exato para abordar esta segunda etapa do problema se deve ao fato da formulação proposta resultar em um modelo linear simples, de variáveis contínuas, e relaxação das restrições mais restritivas,

garantindo, desta forma, o retorno da solução ótima e viável de alocação dos pacientes no sistema com esforço computacional ínfimo.

Apesar desta escolha, outras opções para otimizar a alocação dos pacientes no sistema analisado foram avaliadas, e.g. aplicação de heurística específica, e utilização do GA implementado. Entretanto, estas alternativas foram descartadas. Na opção do NSGA-II, a principal desvantagem levantada está relacionada à dimensão e domínio da variável de alocação $x_{o,d,e}$. Conforme mencionado, os índices $o \in O = \{1, \dots, no\}$, $d \in D = \{1, \dots, nd\}$, e $e \in E = \{1, \dots, ne\}$ desta variável indicam, respectivamente, o município de residência do paciente, o município no qual este será atendido, e a especialidade por ele demandada, resultando em milhares ($no \times nd \times ne$) de variáveis desta natureza. Consequentemente, maiores populações de indivíduos seriam necessárias para garantir adequada exploração do espaço de busca do problema, reduzindo a eficiência computacional do algoritmo. Além disso, a análise simultânea de variáveis com domínios distintos, inteiras ($y_{d,l}$) e contínuas ($x_{o,d,e}$), dificulta a aplicação dos operadores evolutivos propostos.

A partir do exposto optou-se, portanto, pelo método exato para otimizar a alocação dos pacientes no sistema de internação, garantindo, desta forma, o alcance do ponto ótimo global em baixo tempo computacional.

Conforme mencionado, para cada indivíduo da população o modelo linear é executado, de forma a otimizar a variável de alocação $x_{o,d,e}$.

4.4 Avaliação dos Indivíduos via Simulação

Após dimensionar a capacidade de atendimento dos municípios e otimizar a variável de alocação dos pacientes, i.e. definir as variáveis $y_{d,l}$ e $x_{o,d,e}$, é preciso simular as soluções propostas e mensurar as taxas de recusa dos leitos SUS, que servirão para avaliar e ranquear estas soluções.

O processo de Otimização via Simulação se justifica pela incapacidade de se expressar as taxas de recusa f_1 por equações ou formulações matemáticas. Neste caso, a simulação funciona como uma caixa preta que avalia esta função para cada solução. Além disso, conforme detalhado na sequência, a dinamicidade e aleatoriedade intrínsecas ao processo de internação dos pacientes, e as especificidades do problema analisado (e.g. tempo máximo de espera por leitos), dificultam sua modelagem analítica fidedigna.

A simulação proposta nesta pesquisa (Figura 4.4) foi desenvolvida no Simulink, ferramenta para modelagem, simulação e análise de sistemas dinâmicos disponível no *software* MATLAB®.

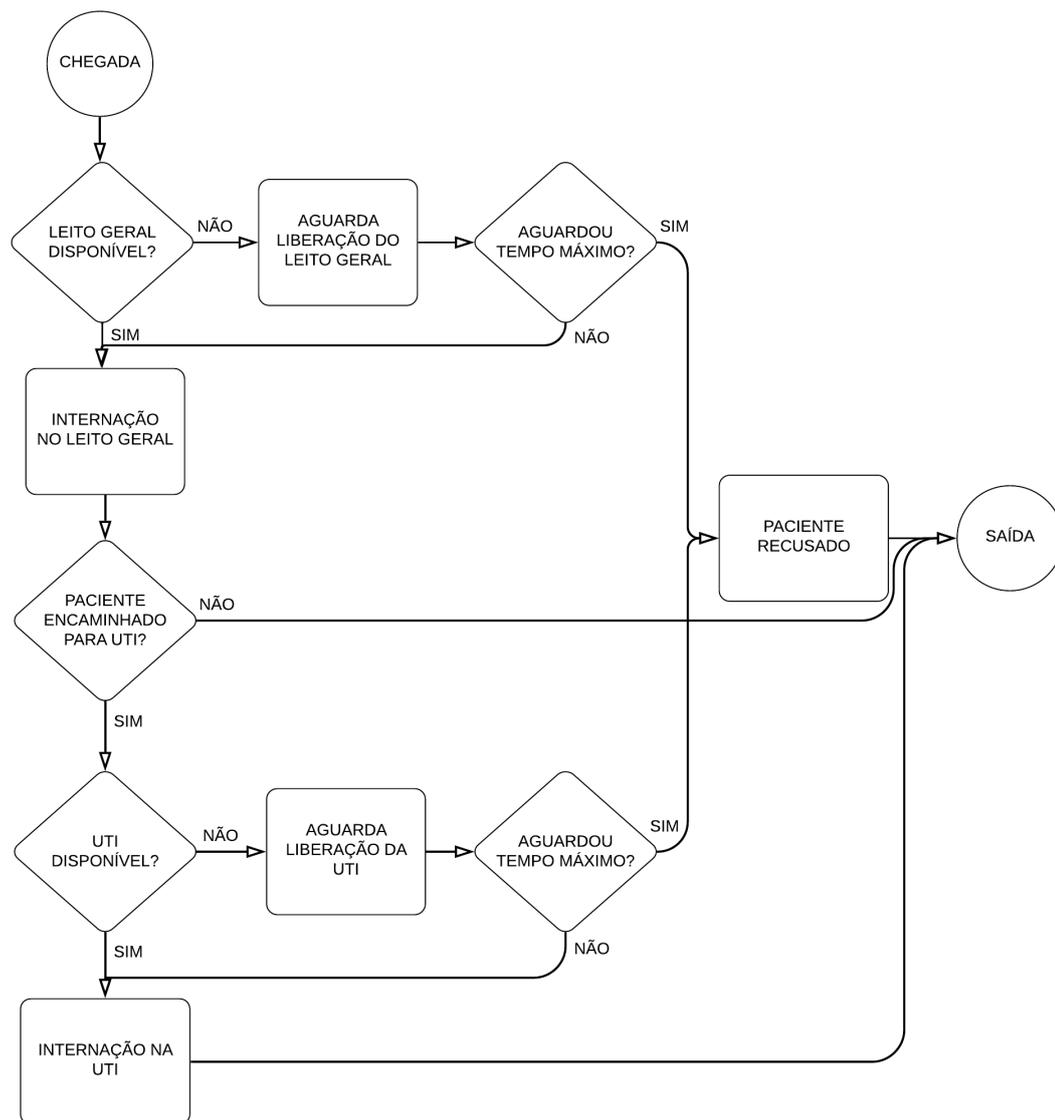


Figura 4.4: Simulação do sistema de internação SUS

Como o banco de dados do DATASUS não dispõe do momento da chegada dos pacientes no sistema, assumiu-se que estas seguem um Processo Poisson, i.e. o intervalo entre esses eventos se aproxima da Distribuição Exponencial, $EXP(1/\lambda)$, onde λ é a quantidade de pacientes que chegam por hora no sistema estudado, obtida ao se dividir o número de internações esperadas anualmente por 8.760 horas.

Ao ingressar no sistema os pacientes, entidades da simulação, recebem os seguintes atributos:

1. ESP: especialidade definida por uma função discreta baseada no percentual histórico de internações por especialidade.

2. LoS: distribuição do tempo de internação dos pacientes.
3. MUN: município de residência, definido por uma função discreta baseada no percentual histórico de internações por municípios de residência.
4. ALO: município para onde cada paciente foi alocado, indicado pela variável $x_{o,d,e}$ otimizada.

Definidos esses atributos, o paciente que ingressa no sistema será atendido se pelo menos um leito compatível com sua necessidade de especialidade estiver disponível no município para onde ele foi alocado. Conforme proposto em Akcali et al. (2006), caso todos os leitos estejam ocupados, o paciente poderá esperar a liberação; entretanto, se um leito não for liberado durante determinado intervalo de tempo, a internação requerida é computada como recusada. Considerou-se que os pacientes podem aguardar a liberação dos leitos gerais por até três horas.

Antes de deixar o sistema, os pacientes poderão ser encaminhados para a UTI com uma probabilidade p_{ICU}^e . A mesma lógica dos leitos gerais é aplicada na procura por uma UTI disponível e compatível à especialidade demandada pelo paciente. Novamente, se as UTIs estiverem ocupadas, o paciente poderá esperar sua liberação durante um tempo antes da internação requerida ser computada como recusada. Neste caso, os pacientes podem aguardar por até uma hora. Na prática, a equipe médica responsável pelo paciente que definirá quanto tempo ele poderá aguardar pelo leito, isto com base na gravidade ou urgência da internação demandada. Entretanto, é inviável analisar e simular caso a caso. Além disso, inexistem informações disponíveis para fazer qualquer inferência acerca deste parâmetro. Desta forma, definimos um tempo menor de espera para casos mais graves, i.e. casos de internações na UTI.

A partir do exposto, verifica-se que na abordagem proposta os pacientes não são encaminhados para outros hospitais quando a demanda excede a oferta dos leitos. Nestes casos, a insuficiência da capacidade de atendimento dos centros de internação é refletida através do aumento da recusa de pacientes (f_1).

Esta abordagem possibilita a modelagem mais fidedigna da operação dos sistemas de internação, pois na prática pequenos ou curtos ajustes de capacidade são viabilizados pela acomodação temporária dos pacientes nos corredores ou outros locais.

Além disso, assumiu-se que antes das UTIs, os pacientes passam pelos respectivos leitos gerais. Apesar disto nem sempre ocorrer na prática, i.e. alguns casos graves são encaminhados diretamente para as UTIs. Esta suposição é razoável, visto que após tratamento nas UTIs, os pacientes são internados em leitos gerais para continuidade do tratamento ou observação.

Na modelagem e simulação, a ordem destas interações não impacta na recusa geral dos leitos retornada ao algoritmo NSGA-II.

Cada solução candidata foi simulada por um ano (365 dias), pois, a partir deste período os resultados retornados pela simulação representam de fato as condições estacionárias do sistema estudado.

Conforme mencionado, a avaliação das soluções propostas é realizada via Método Global de computação paralela, vide Figura 2.6. Neste caso, os operadores e processos evolutivos do NSGA-II ficam a cargo do processador central ou "mestre", e apenas a etapa de simulação das soluções candidatas é paralelizada. Nesta pesquisa foram utilizadas cinco estações de trabalho quad-core com processador i5 e 8GB de memória RAM.

A Figura 4.5 indica a janela do *Admin Center*, que consiste em um ambiente do MATLAB® em que é possível controlar e monitorar os processos paralelos sendo executados pelo algoritmo. Na primeira aba são listadas as estações de trabalho, a título de exemplo incluímos apenas uma das estações utilizadas nesta pesquisa. A segunda aba indica o *MATLAB Job Scheduler* (MJS), responsável por distribuir e coordenar a execução dos processos em paralelo. No exemplo o *MJS-Simul.* é composto apenas pela estação de trabalho *PC-Simul.* com quatro núcleos (*workers*) que operam de maneira independente.

Nesta pesquisa as simulações/avaliações dos indivíduos são distribuídas aleatoriamente entre os 20 núcleos que compõem as estações de trabalho utilizadas. O pseudocódigo *Paralelização* indica como este processo acontece. O comando *parfor* consiste em um *loop* ou ciclo de comandos executados em paralelo para cada indivíduo i da população corrente. Inicialmente, as variáveis x e y dos núcleos ativos recebem os valores $x(i)$ e $y(i)$ que representam, respectivamente, o número de leitos e o fluxo de pacientes propostos pelo indivíduo i no sistema de internação estudado. Feitas estas atribuições a simulação é executada em paralelo, retornando à estação central os resultados encontrados para cada configuração analisada. Ao finalizar a simulação de um indivíduo, o núcleo recebe os *inputs* de outro indivíduo não avaliado. Este processo se repetirá até que toda a população seja avaliada.

Algorithm 1 : Paralelização

```
1: parfor each  $i \in I$ 
2:   assignin(base,x,x(i))
3:   assignin(base,y,y(i))
4:   outputs(i) ← sim(Simulação)
5: end
```

Hosts									
Add or Find...		Host			MDCE Service		MJS	Workers	
Start mdce Service...		Hostname	Reachable	Cores	Status	Up Since	Name	Count	
Stop mdce Service...		PC-Simul.	yes	4	running	2020-01-20 14:13	MJS-Simul.	4	
Test Connectivity...									

MATLAB Job Scheduler (MJS)						
Start...		Name	Hostname	Status	Up Since	Workers
Stop...		MJS-Simul.	PC-Simul.	running	2020-01-20 14:15	4
Resume						

Workers							
Start...		Worker				MJS	
Stop...		Name	Hostname	Status	Up Since	Connection	Hostname
Resume		PC-Simul_worker01	PC-Simul.	idle	2020-01-20 14:17	connected	PC-Simul.
		PC-Simul_worker02	PC-Simul.	idle	2020-01-20 14:18	connected	PC-Simul.
		PC-Simul_worker03	PC-Simul.	idle	2020-01-20 14:18	connected	PC-Simul.
		PC-Simul_worker04	PC-Simul.	idle	2020-01-20 14:18	connected	PC-Simul.

Last updated: 20/01/20 14:33 Update every 5 minutes

Figura 4.5: Simulação do sistema de internação SUS

A partir deste processo de avaliação dos indivíduos, é possível compará-los e ordená-los em relação às demais soluções da população, processo realizado em sequencial pelo processador "mestre".

4.5 Classificação dos Indivíduos

Segundo Zitzler et al. (2004), os critérios de comparação das soluções nos MOEAs se distinguem conforme a estratégia utilizada (Figura 4.6):

1. **Agregação:** combinação ponderada de múltiplos objetivos.
2. **Critério:** alternância da função objetivo usada para guiar o processo de evolução.
3. **Dominância:** ordenação dos indivíduos baseada nas fronteiras de dominância.

Nesta pesquisa optou-se pela estratégia de dominância proposta em Deb et al. (2002). Conforme o pseudocódigo *Ordenação*, nesta estratégia para cada indivíduo p da população P , são

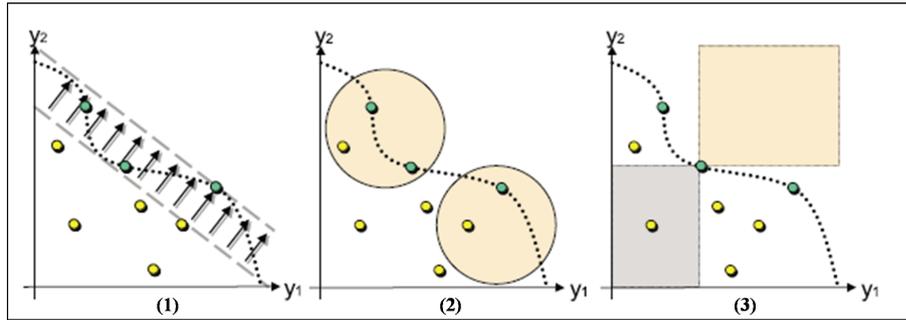


Figura 4.6: Estratégias de avaliação das soluções
 Fonte: Zitzler et al. (2004)

definidos os seguintes atributos:

1. n_p : número de soluções que dominam a solução p , sendo $n_p = 0$ para as soluções pertencentes à primeira fronteira Pareto, composta por indivíduos não dominados.
2. S_p : conjunto de soluções que a solução p domina, sendo $S_p = \emptyset$ para as soluções pertencentes à última fronteira Pareto, composta por indivíduos dominados por todos os outros das demais fronteiras.

Após inicializar estes parâmetros, cada indivíduo p da população P é comparado aos demais indivíduos q desta mesma população. Caso p domine q , então a solução q é adicionada ao conjunto de soluções dominadas por p (S_p); do contrário, se q dominar p , o contador de soluções que dominam p (n_p) é acrescido.

Realizadas essas comparações, os indivíduos p com $n_p = 0$ são incluídos na primeira fronteira (\mathcal{F}_1), e as soluções dominadas por esses indivíduos (S_p) na segunda fronteira. Esse processo continua até que todas as fronteiras sejam definidas.

Portanto, finalizada a ordenação, na primeira fronteira estarão agrupados os indivíduos que não são dominados por nenhuma outra solução. Na segunda fronteira, se encontrão aqueles indivíduos que são dominados por pelo menos uma solução da primeira fronteira. Na terceira fronteira estarão os indivíduos dominados apenas por soluções das fronteiras anteriores, e assim por diante.

Como nesta pesquisa o problema estudado é restrito, foi preciso adaptar o processo de definição das fronteiras não dominadas apresentado. Neste caso, o algoritmo implementado considera as funções objetivo e a violação das restrições das soluções comparadas.

Além desta estratégia de ordenação, Deb et al. (2002) propõem outro critério para ranqueamento e preservação da diversidade dos indivíduos, a Distância de Multidão (DM), também

Algorithm 2 : Ordenação

```
1: for each  $p \in P$  do
2:    $S_p \leftarrow \emptyset$ 
3:    $n_p \leftarrow 0$ 
4:   for each  $q \in P$  do
5:     if  $p \prec q$  then
6:        $S_p \leftarrow S_p \cup \{q\}$ 
7:     else
8:       if  $q \prec p$  then
9:          $n_p \leftarrow n_p + 1$ 
10:      end if
11:    end if
12:  end for
13:  if  $n_p == 0$  then
14:     $p_{rank} \leftarrow 1$ 
15:     $\mathcal{F}_1 \leftarrow \mathcal{F}_1 \cup \{p\}$ 
16:  end if
17: end for
18:  $i \leftarrow 1$ 
19: while  $\mathcal{F}_i \neq \emptyset$  do
20:    $Q \leftarrow \emptyset$ 
21:   for each  $p \in \mathcal{F}_i$  do
22:     for each  $q \in S_p$  do
23:        $n_q \leftarrow n_q - 1$ 
24:       if  $n_q == 0$  then
25:          $q_{rank} \leftarrow i + 1$ 
26:          $Q \leftarrow Q \cup \{q\}$ 
27:       end if
28:     end for
29:   end for
30:    $i \leftarrow i + 1$ 
31:    $\mathcal{F}_i \leftarrow Q$ 
32: end while
```

aplicado nesta pesquisa.

A DM de um indivíduo i consiste na estimativa do perímetro do cuboide formado pelas soluções vizinhas mais próximas ao indivíduo na fronteira analisada, $i-1$ e $i+1$, vide Figura 4.7. Desta forma, as soluções com menores valores de DM, estão nos locais com maior concentração de indivíduos do espaço de busca e, portanto, contribuem menos para a manutenção da diversidade da população.

Conforme o pseudocódigo do algoritmo *Distância de Multidão*, a definição das DMs dos indivíduos da fronteira \mathcal{I} depende da ordenação dessas soluções segundo as distintas funções objetivo. Após esta etapa, as DMs das soluções extremas, soluções 1 e L na Figura 4.7, são definidas como $\mathcal{I}[1]_{distance} = \mathcal{I}[L]_{distance} = \infty$. Para os outros indivíduos essa distância é calculada pelo so-

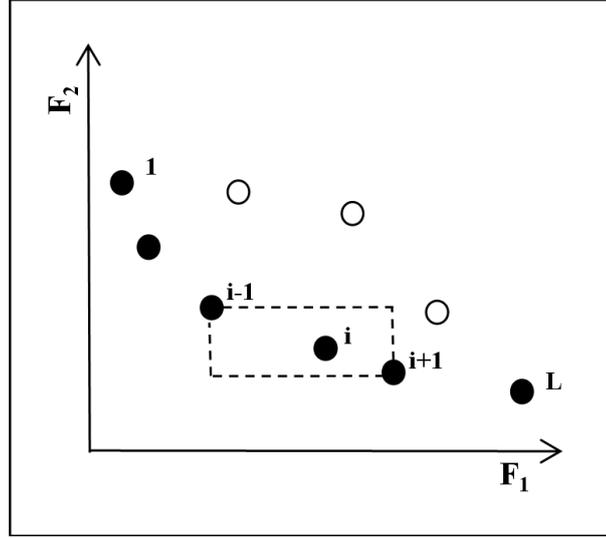


Figura 4.7: Cálculo da Distância de Multidão
 Fonte: Adaptado de Deb et al. (2002).

matório das diferenças normalizadas das funções objetivo dos indivíduos adjacentes, conforme indicado na linha 10 do pseudocódigo.

Algorithm 3 : Distância de Multidão

```

1:  $l \leftarrow |\mathcal{I}|$ 
2: for each  $i \in \mathcal{I}$  do
3:    $\mathcal{I}[i]_{distance} \leftarrow 0$ 
4: end for
5: for each objective  $m$  do
6:    $\mathcal{I} \leftarrow \text{sort}(\mathcal{I}, m)$ 
7:    $\mathcal{I}[1]_{distance} \leftarrow \infty$ 
8:    $\mathcal{I}[l]_{distance} \leftarrow \infty$ 
9:   for  $i = 2$  to  $(l - 1)$  do
10:     $\mathcal{I}[i]_{distance} \leftarrow \mathcal{I}[i]_{distance} + \frac{(\mathcal{I}[i + 1].m - \mathcal{I}[i - 1].m)}{(fmax_m - fmin_m)}$ 
11:   end for
12: end for

```

$\mathcal{I}[i].m$, $fmax_m$ e $fmin_m$ referem-se, respectivamente, ao valor da m -ésima função objetivo da solução i pertencente ao conjunto não dominado \mathcal{I} , e o valor máximo e mínimo da m -ésima função objetivo na população sendo avaliada.

A partir destes critérios, Deb et al. (2002) propõem o operador \prec_n para comparação dos indivíduos. Segundo os autores, $i \prec_n j$ se pelo menos uma das seguintes condições for satisfeita:

1. $i_{rank} < j_{rank}$: a solução i domina a solução j .
2. $i_{rank} = j_{rank}$ e $i_{DM} > j_{DM}$: as soluções i e j pertencem ao mesmo conjunto de soluções não dominadas; entretanto, a Distância de Multidão de i é maior que a de j .

Diante disto, entre indivíduos de diferentes fronteiras é considerada melhor a solução pertencente à fronteira dominante, i.e. menor *rank*. Entre indivíduos mutuamente não dominados, será considerada melhor a solução mais afastada na fronteira, i.e. solução com maior Distância de Multidão. Desta forma, garante-se a manutenção da diversidade genética da população no decorrer das gerações.

Os critérios de comparação e ordenação supracitados são utilizados nos processos evolutivos aplicados ao algoritmo NSGA-II, conforme detalhado na sequência.

4.6 Operações de Evolução e Elitismo

4.6.1 Cruzamento

Nesta tese, o processo de cruzamento se inicializa sorteando dois indivíduos genitores da população corrente, *pai1* e *pai2*. A probabilidade de ocorrência do cruzamento p_c definirá se os indivíduos serão combinados para formação de outros dois novos indivíduos descendentes, *filho1* e *filho2*, ou se serão apenas copiados na população seguinte Q_t . Esse processo é repetido até que a população Q_t esteja completa, ou seja, $|Q_t| = npop$.

A Figura 4.8 ilustra o processo de cruzamento implementado. Conforme indicado pelas setas destacadas, após definir os indivíduos genitores que participarão do cruzamento, são sorteados dois valores, *corte1* e *corte2*, que indicam a coluna e a linha onde serão feitos os cortes para troca de informação dos indivíduos.

Os quadrantes 1 e 4 do primeiro descendente (*filho1*) serão compostos pelas informações clonadas do primeiro genitor (*pai1*), enquanto os quadrantes 2 e 3 herdarão as informações do segundo genitor (*pai2*). O contrário ocorrerá na formação do segundo indivíduo descendente (*filho2*). Os indivíduos descendentes são incluídos na população Q_t , e o processo se repete até que esta população esteja completa.

Conforme mencionado, os indivíduos são definidos por matrizes Y , cujos elementos $y_{d,l}$ indicam o número de leitos do tipo l contratados no município d .

Durante o cruzamento, quaisquer indivíduos genitores poderão ser combinados, e nenhum mecanismo é aplicado para impedir a violação do número máximo de leitos que cada cidade pode receber, restrição definida pela Equação (3.6). Desta forma, após a etapa de cruzamento alguns indivíduos descendentes poderão violar a restrição de capacidade.

Ao permitir a geração e avaliação de soluções inviáveis, a abordagem proposta garante a

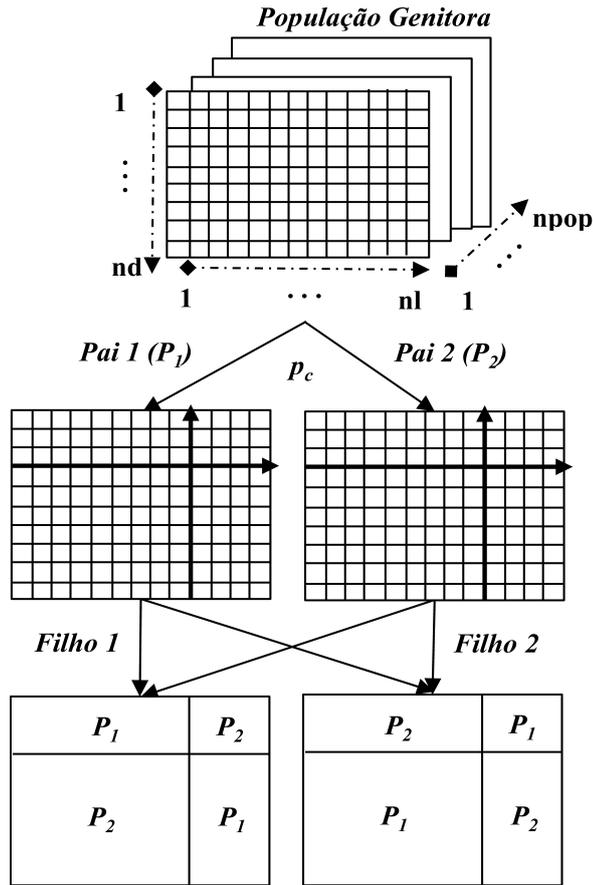


Figura 4.8: Cruzamento dos indivíduos no NSGA-II

ampliação e exploração do espaço de busca do algoritmo, evitando sua convergência prematura para regiões de ótimos locais.

Além de passar pelo processo de cruzamento, alguns indivíduos da população Q_t também poderão sofrer mutações.

4.6.2 Mutação

As mutações são processos evolutivos que ocasionam pequenas perturbações nos indivíduos. Segundo Xu et al. (2016), as pequenas alterações randômicas provocadas pela mutação aumentam as chances de que as soluções mais eficientes sejam avaliadas, e ajudam a evitar a convergência prematura do algoritmo.

A Figura 4.9 ilustra o processo de mutação considerado nesta pesquisa. Os valores l e d indicam qual tipo de leito l será redimensionado no município d .

Definido o indivíduo que passará pelo processo de mutação, a alteração do elemento hachurado é conduzida na tentativa de reduzir ou resolver a violação de restrições em soluções inviáveis, ou

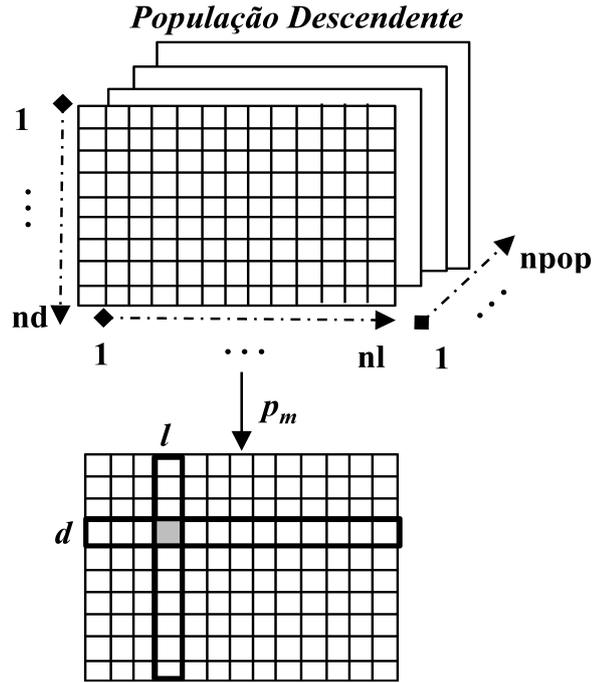


Figura 4.9: Mutação dos indivíduos no NSGA-II

melhorar as funções objetivo no caso de soluções viáveis.

4.6.3 Elitismo

Durante a evolução é possível que indivíduos bem avaliados sejam perdidos. O próprio algoritmo pode recuperar esta informação nas gerações seguintes, entretanto, não existe garantia que isto de fato aconteça, nem de quantas gerações passarão antes que esses indivíduos reapareçam na população.

Para contornar este problema optamos por utilizar o mecanismo de Elitismo proposto em Deb et al. (2002). No Elitismo as soluções mais bem avaliadas da geração corrente são copiadas na próxima população de indivíduos. Isso evita, portanto, que diversas gerações tenham que ser executadas para recuperar informações perdidas durante o processo de evolução.

Conforme Figura 4.10 e pseudocódigo *Elitismo*, inicialmente, define-se R_t , uma população temporária formada pela união das populações de indivíduos genitores P_t e descendentes Q_t .

A população R_t é ordenada segundo o critério de dominância e o cálculo da Distância de Multidão. A partir desta ordenação o algoritmo define a próxima população de indivíduos, P_{t+1} . Serão priorizados os indivíduos não dominados de R_t , i.e. pertencentes à primeira fronteira (\mathcal{F}_1). Se o número de soluções deste conjunto for inferior à $npop$, então, todas serão incluídas na população P_{t+1} , já os indivíduos restantes, i.e. $npop - |\mathcal{F}_1|$ serão obtidos das fronteiras

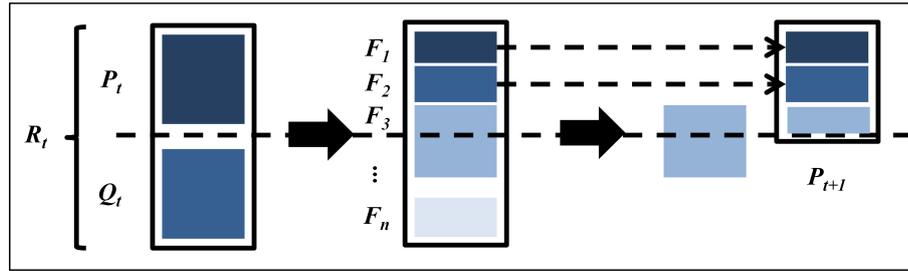


Figura 4.10: Processo de Elitismo NSGA-II
Fonte: Deb et al. (2002)

Algorithm 4 : Elitismo

- 1: $R_t \leftarrow P_t \cup Q_t$
 - 2: $\mathcal{F} \leftarrow \text{fast non dominated sort}(R_t)$
 - 3: $P_{t+1} \leftarrow \emptyset$
 - 4: $i \leftarrow 1$
 - 5: **while** $|P_{t+1}| + |\mathcal{F}_i| \leq n_{pop}$ **do**
 - 6: $P_{t+1} \leftarrow P_{t+1} \cup \mathcal{F}_i$
 - 7: $i \leftarrow i + 1$
 - 8: **end while**
 - 9: crowding distance assignment (\mathcal{F}_i)
 - 10: $\text{sort}(\mathcal{F}_i, \prec_n)$
 - 11: $P_{t+1} \leftarrow P_{t+1} \cup \mathcal{F}_i[1 : (N - |P_{t+1}|)]$
-

subsequentes.

Esse processo continua até que P_{t+1} esteja completa, i.e., $|P_{t+1}| = n_{pop}$, ou até que não seja mais possível inserir todas as soluções da fronteira l analisada na próxima geração, i.e., $|\mathcal{F}_l| + |P_{t+1}| > n_{pop}$. Neste último caso, as soluções com maior valor de Distância de Multidão serão escolhidas para completar a população P_{t+1} .

Capítulo 5

Apresentação dos Resultados

Este capítulo apresenta os principais resultados alcançados na aplicação da abordagem de Otimização via Simulação no problema de alocação de leitos de internação. Conforme mencionado, a abordagem proposta é aplicada em uma das macro-regiões de saúde de Minas Gerais.

Uma síntese da análise dos resultados levantados está apresentada no manuscrito intitulado "*A Simulation-Optimisation approach for hospital beds allocation*", submetido à *International Journal of Medical Informatics* no dia 23 de Agosto de 2019, e ainda em revisão.

5.1 Definição da macro-região de saúde

Minas Gerais é o quarto maior estado do Brasil, com cerca de 20 milhões de habitantes distribuídos em 586 mil km². O estado apresenta grande disparidade interna quanto à densidade demográfica e características socioeconômicas. Estas diferenças regionais se refletem de maneira semelhante no campo da saúde, ocorrendo maior concentração de serviços e recursos nas regiões centrais do estado.

O deslocamento da população em busca de atendimento hospitalar e a distribuição desigual dos recursos da saúde no estado de Minas Gerais, desencadearam um estudo de descentralização e regionalização dos serviços de atenção à saúde no estado. Publicado pela Secretaria de Estado de Saúde de Minas Gerais em 2011, o Plano Diretor de Regionalização da Saúde de Minas Gerais (PDRMG), propôs a divisão dos 853 municípios mineiros em 13 macro-regiões de saúde, conforme ilustrado na Figura 5.1.

A definição destas regiões de saúde baseia-se no equilíbrio entre princípios da economia de escala de recursos humanos e materiais, qualidade dos serviços prestados e acesso dos usuários.

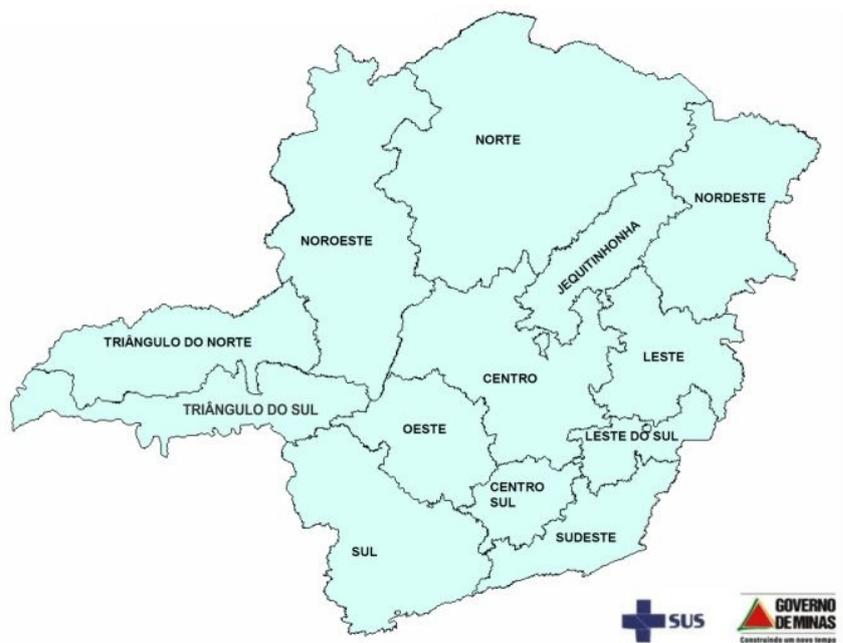


Figura 5.1: Macro-regiões de saúde em Minas Gerais
Fonte: Malachias et al. (2011)

Existem evidências concretas das relações entre qualidade e escala dos serviços prestados, e.g. hospitais que realizam muitas cirurgias apresentam taxas mais baixas de mortalidade se comparados aos que fazem poucos procedimentos cirúrgicos. Por outro lado, quando existe conflito entre escala e acesso, há de se predominar o princípio do acesso sempre. Baseado nestes fatores, o PDRMG objetiva direcionar a regionalização e descentralização dos serviços públicos de saúde com vistas ao maior e mais adequado acesso dos usuários (Malachias et al., 2011).

A distribuição e delimitação dos territórios regionais proposta no PDRMG está embasada na densidade tecnológica e complexidade dos serviços ofertados pelos diferentes níveis de atenção. Neste caso, cada macro-região delimitada deverá a longo prazo ofertar serviços de todos os níveis de atenção, i.e., atender as demandas dos pacientes residentes na própria região à nível primário, secundário e terciário.

Em concordância à autossuficiência traçada as regiões de saúde a longo prazo, esta pesquisa escolheu uma macro-região de Minas Gerais para validação e aplicação da metodologia aqui proposta.

A Tabela 5.1 indica os critérios utilizados neste processo: (i) percentual de internações registradas cujos pacientes residiam na macro-região de saúde, e.g., 98,98% das internações feitas em 2017 na região Nordeste se referem a pacientes residentes desta região; (ii) percentual de

residentes da macro-região de saúde internados em outras regiões, e.g., 14,35% dos pacientes residentes na macro-região Nordeste foram internados em outras regiões de saúde.

Tabela 5.1: Internações nas macro-regiões de saúde de Minas Gerais

Macro-região de saúde de Minas Gerais	% de internações com residentes da macro região	% de residentes internados em outras macro-regiões
Nordeste	98,98%	14,34%
Triângulo sul mineiro	98,80%	1,30%
Norte	98,65%	1,81%
Triângulo norte mineiro	98,29%	1,26%
Sul	98,26%	1,17%
Centro Sul	97,99%	13,68%
Leste do Sul	97,24%	14,02%
Noroeste	97,16%	8,75%
Leste	96,86%	6,09%
Oeste	96,54%	15,29%
Sudeste	94,54%	1,64%
Centro	92,12%	0,68%
Jequitinhonha	80,96%	9,06%

A partir dos dados da Tabela 5.1, a macro-região do triângulo sul mineiro foi escolhida para validação e aplicação da metodologia proposta nesta pesquisa.

O triângulo sul mineiro, Figura 5.2, é composto por 27 municípios distribuídos entre três micro-regiões de saúde.



Figura 5.2: Triângulo sul mineiro
Fonte: Malachias et al. (2011)

A escolha desta região se deve ao fato dela ser praticamente autossuficiente. Conforme indicado na Tabela 5.1, mais de 98% das internações registradas em 2017 neste local foram de pacientes residentes na macro-região. Além disso, menos de 1,5% dos pacientes que residem nesta região foram encaminhados para atendimento em outros locais. Isto significa dizer que esta região sofre pouca interferência das demais regiões de saúde, tornando o sistema em estudo quase independente das demais regiões de saúde.

5.2 Dados da macro-região de saúde em estudo

Conforme mencionado no Capítulo 3, as informações utilizadas nesta pesquisa foram coletadas e calculadas a partir do banco de dados DATASUS, disponível em www.datasus.gov.br, e também da Portaria 1.631/2015 (Brasil, 2015).

Segundo esta portaria, a estimativa das internações esperadas em um município o para cada especialidade e é calculada pela Equação (5.1).

$$de_{o,e} = pr_{o,e} \times ti_e \quad (5.1)$$

Em que $pr_{o,e}$ e ti_e representam, respectivamente, a população referência e a proporção de internações da especialidade e . Os cálculos de $pr_{o,e}$ e ti_e seguiram as referências disponíveis na Portaria 1.631/2015 e expostas na Tabela 5.2.

Tabela 5.2: População referência e proporção de internações em leitos gerais

Fonte: Brasil (2015).

Especialidade	População Referência	ti mínima	ti máxima
Cirurgia 15 a 59 anos	população de 15 a 59 anos \times proporção de 15 a 59 anos sem plano de saúde	2,15%	3,57%
Cirurgia 60 anos ou mais	população 60 anos e mais \times proporção 60 anos ou mais sem plano de saúde	4,40%	7,26%
Clínica 15 a 59 anos	população de 15 a 59 anos \times proporção de 15 a 59 anos sem plano de saúde	1,38%	2,46%
Clínica 60 anos ou mais	população 60 anos e mais \times proporção 60 anos ou mais sem plano de saúde	7,24%	11,68%
Obstetrícia	número de nascidos vivos \times fator de ajuste do subregistro \times proporção de mulheres de 15 a 49 anos sem plano de saúde com cobertura de obstetrícia	100%	100%
Pediatria Clínica	população de menores de 15 anos \times proporção de menores 15 anos sem plano de saúde	2,78%	4,88%
Pediatria Cirúrgica	população de menores de 15 anos \times proporção de menores 15 anos sem plano de saúde	0,84 %	1,94%
Neonatologia	número de nascidos vivos \times fator de ajuste do subregistro \times proporção de menores de 1 ano sem plano de saúde	8,24%	16,80%

A partir da Equação (5.1) e Tabela 5.2, foi possível calcular a população referência, e demanda anual mínima e máxima estimada no triângulo sul mineiro, Tabela 5.3.

Tabela 5.3: Demanda esperada para cada especialidade

Especialidade	População Referência	Demanda Mínima	Demanda Máxima
Cirurgia 15 a 59 anos	350.884	7.544	12.527
Cirurgia 60 anos ou mais	80.841	3.557	5.871
Clínica 15 a 59 anos	327.365	4.845	8.632
Clínica 60 anos ou mais	80.843	5.853	9.447
Obstetrícia	6.213	6.213	6.213
Pediatria Clínica	95.072	2.643	4.641
Pediatria Cirúrgica	95.238	800	1.846
Neonatologia	6.383	526	1.071
Total	1.042.839	31.981	50.248

Verifica-se que ao utilizar as referências mínimas da Portaria 1.631/2015 para a taxa de internação, 31.981 pacientes seriam esperados anualmente na macro-região de saúde em estudo. Ao utilizar os valores máximos, a demanda esperada subiria para 50.248 pacientes.

Como o banco de dados do DATASUS não dispõe do momento da chegada dos pacientes no sistema, na simulação assumimos que estas seguem um Processo Poisson, i.e. o intervalo entre esses eventos se aproxima da Distribuição Exponencial, $EXP(1/\lambda)$, onde λ é a quantidade de pacientes que chegam por hora no sistema estudado, obtida ao se dividir o número de internações esperadas anualmente por 8.760 horas. Desta forma, a partir da Tabela 5.3 tem-se que $\lambda_{min} = 3,65$ e $\lambda_{max} = 5,74$.

O percentual de internações em leitos padrões que são posteriormente encaminhadas para as UTIs, p_{ICU} , é definido conforme os limites de referência indicados na Portaria N° 1.631/2015 e Tabela 5.4.

Tabela 5.4: Probabilidade de encaminhamento para as UTIs

Fonte: Brasil (2015).

Especialidade	p_{ICU} mínima	p_{ICU} máxima
Cirurgia 15 a 59 anos	4,88%	6,25%
Cirurgia 60 anos ou mais	9,92%	18,46%
Clínica 15 a 59 anos	4,28%	6,24%
Clínica 60 anos ou mais	6,03%	9,85%
Obstetrícia	0,05%	0,66%
Pediatria Clínica	2,34%	5,04%
Pediatria Cirúrgica	3,10%	5,61%
Neonatologia	23,54%	39,79%

Com relação aos tempos de internação dos pacientes, utilizou-se como referência as distribui-

ções de probabilidade apresentadas na Tabela 5.5 e propostas no trabalho de Pinto et al. (2014), que também estuda um sistema público de internação brasileiro.

Tabela 5.5: Distribuição do tempo de internação dos pacientes

Fonte: Pinto et al. (2014).

Especialidade	LoS (dias)
Cirurgia 15 a 59 anos	EXP(3,63)
Cirurgia ≥ 60 anos	EXP(4,43)
Clínica 15 a 59 anos	EXP(8,63)
Clínica ≥ 60 anos	GAMMA(7,88; 1,3)
Obstetrícia	GAMMA(1,08; 2,31)
Pediatria Clínica	EXP(4,90)
Pediatria Cirúrgica	EXP(2,86)
Neonatal	LOGN(6,19; 5,34)
UTI Cirúrgica 15 a 59 anos	EXP(5,30)
UTI Cirúrgica ≥ 60 anos	97*BETA(0,411; 8,13)
UTI Clínica 15 a 59 anos	EXP(7,70)
UTI Clínica ≥ 60 anos	EXP(8,80)
UTI Obstétrica	0,5 + WEIB(2,14; 1,33)
UTI Pediatria Clínica	0,5 + 97*BETA(0,61; 4,55)
UTI Pediatria Cirúrgica	0,5 + WEIB(5,24; 0,79)
UTI Neonatal	0,5 + EXP(12,00)

Neste caso, a Portaria N° 1.631/2015 não foi utilizada como referência, pois, traz apenas valores exatos, desconsiderando a aleatoriedade desses dados. Apesar dos dados não serem relativos à macro-região estudada, entende-se que os tempos de internação dos pacientes nos leitos não dependem da região em análise. Sendo assim, as distribuições de probabilidade encontradas em Pinto et al. (2014) atendem bem aos objetivos propostos nesta pesquisa.

A partir destes dados, foi possível então analisar e definir os parâmetros do algoritmo NSGA-II implementado. As análises foram baseadas no cenário mais crítico do ponto de vista da otimização, no qual as taxas de internação equivalem aos valores máximos de referência da Portaria N° 1.631/2015. Desta forma, produz-se a simulação da demanda máxima esperada na macro-região do triângulo sul mineiro.

Além dos dados do DATASUS, também foi usada uma API do Google para se obter as distância reais entre os municípios da macro-região de saúde analisada.

5.3 Parâmetros do algoritmo NSGA-II

Assim como outras heurísticas, a eficiência dos AGs depende da escolha adequada dos parâmetros de entrada do algoritmo. As soluções geradas são sensíveis aos valores dos *inputs*; portanto, é

importante conduzir uma análise dessa relação antes de se executar o algoritmo (Jones et al., 2002).

Uma das vantagens do NSGA-II em relação aos demais AGs está no cálculo da Distância de Multidão, dado que este não depende de um parâmetro de entrada, diferente de diversos outros mecanismos também utilizados na comparação e manutenção da diversidade dos indivíduos. Além disso, conforme já mencionado, no NSGA-II a classificação dos indivíduos é baseada no critério de dominância, dependendo apenas dos valores das funções objetivo e da violação das restrições, sem depender da definição de parâmetros de penalidade para cálculo da *fitness*.

Dentre os *inputs* que impactam na eficiência desses algoritmos, destaca-se o tamanho da população de indivíduos, pois, quanto maior o número de soluções neste conjunto, melhor será a exploração do espaço de busca do problema analisado. Segundo Pierreval and Paris (2000), o número de variáveis contidas neste espaço de busca é uma boa estimativa para definição do tamanho da população de indivíduos.

Nesta tese, a estrutura matricial que representa os indivíduos do algoritmo NSGA-II é composta pelas variáveis $y_{d,l}$, onde $d \in D = \{1, \dots, nd\}$ representa os municípios da macro-região em estudo, e $l \in D = \{1, \dots, nl\}$ os tipos de leitos. O número de variáveis y sendo otimizadas pelo NSGA-II é, portanto, $nd \times nl$. Conforme já mencionado, nesta pesquisa foram considerados 8 tipos de leito e 27 municípios do triângulo sul mineiro, i.e. $nl = 8$ e $nd = 27$. Desta forma, cada indivíduo do algoritmo NSGA-II é representado por uma estrutura matricial de 216 elementos ($nd \times nl$). A partir deste escopo definiu-se que sua população conteria 200 indivíduos.

Além deste parâmetro de entrada, Xu et al. (2016) destacam a importância de se definir o número máximo de gerações ou iterações do algoritmo e a probabilidade de ocorrência dos processos evolutivos. Para determinar estes parâmetros, o algoritmo foi executado com 50, 100, 150 e 200 gerações, mantendo-se inicialmente as probabilidades dos operadores evolutivos constantes, $p_m = 0,03$ e $p_c = 0,75$, e as probabilidades de interseção fixadas nos valores máximos de referência da Portaria N° 1.631/2015.

Devido aos fatores aleatórios destes operadores, foram feitas 10 replicações do algoritmo para cada cenário em estudo, permitindo a obtenção de resultados mais confiáveis, i.e., Intervalo de Confiança (IC) de 95%.

A Figura 5.3 ilustra a primeira fronteira de indivíduos não dominados encontrada após rodar as 10 replicações dos cenários simulados.

Conforme esperado, quanto maior o número de gerações ou iterações simuladas, mais próxima

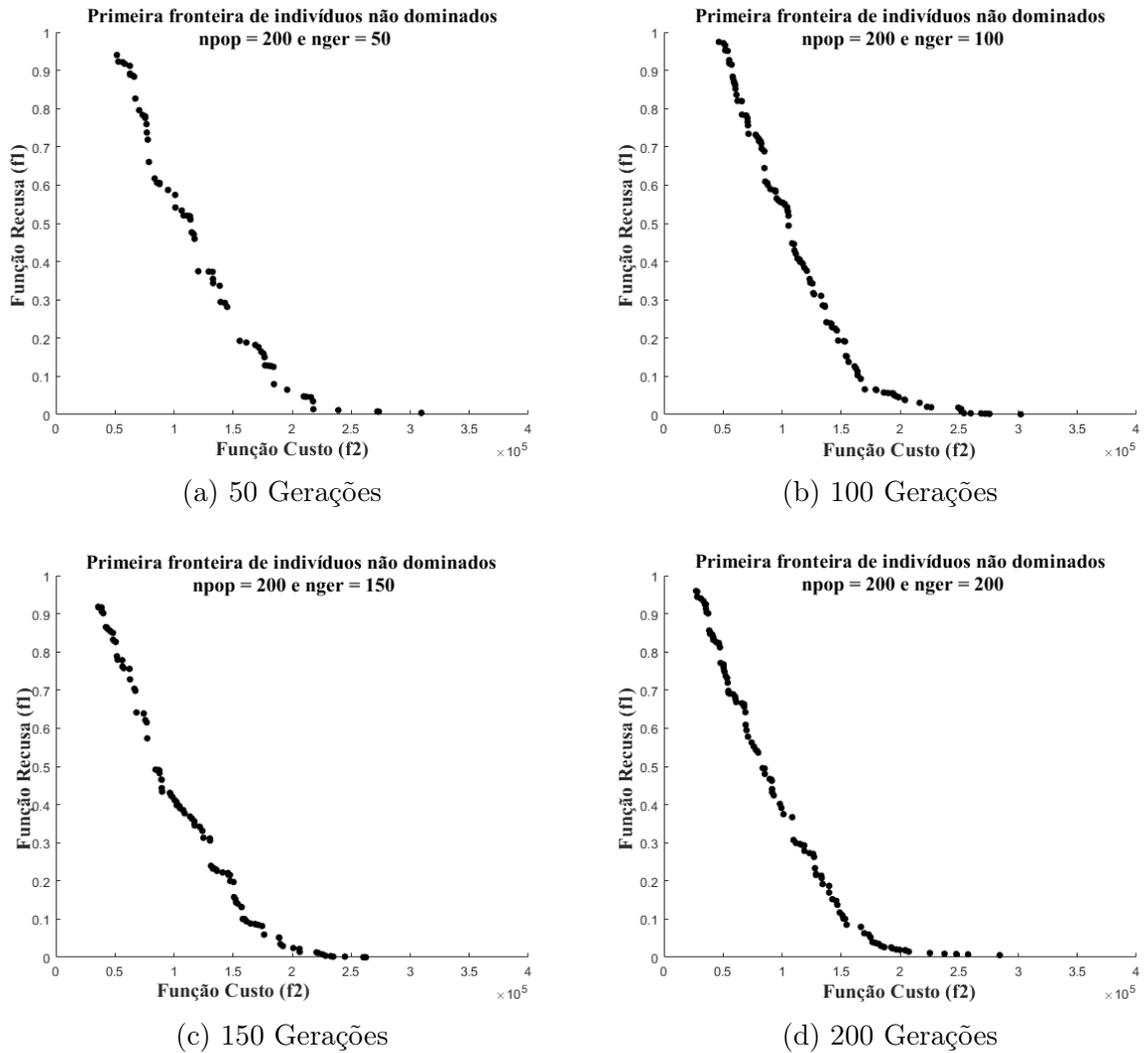


Figura 5.3: Análise de $ngger$ - Primeira fronteira de indivíduos não dominados

fica a fronteira de soluções não dominadas da região com menores taxas de recusa e custo dos leitões.

No intuito de avaliar e comparar a convergência e desempenho do algoritmo nos cenários simulados, foi calculado o Hipervolume. O Hipervolume (HV), proposto por Zitzler and Thiele (1998), mensura o espaço objetivo coberto pela área formada entre os pontos de um conjunto Pareto e certo ponto deste espaço, definido como ponto referência. Desta forma, quanto maior o valor de HV, melhor é o conjunto Pareto, pois mais disperso e diverso é o conjunto de soluções não dominadas encontrado.

A Figura 5.4 apresenta o valor médio e o intervalo de confiança (95%) da área formadas entre o ponto de referência e a primeira fronteira de indivíduos não dominados encontrada ao final de cada iteração dos cenários simulados. Neste caso, o ponto ($x = 9 \times 10^5$, $y = 1,00$) foi escolhido

como referência, pois ele se refere aos valores máximos das funções objetivo do problema.

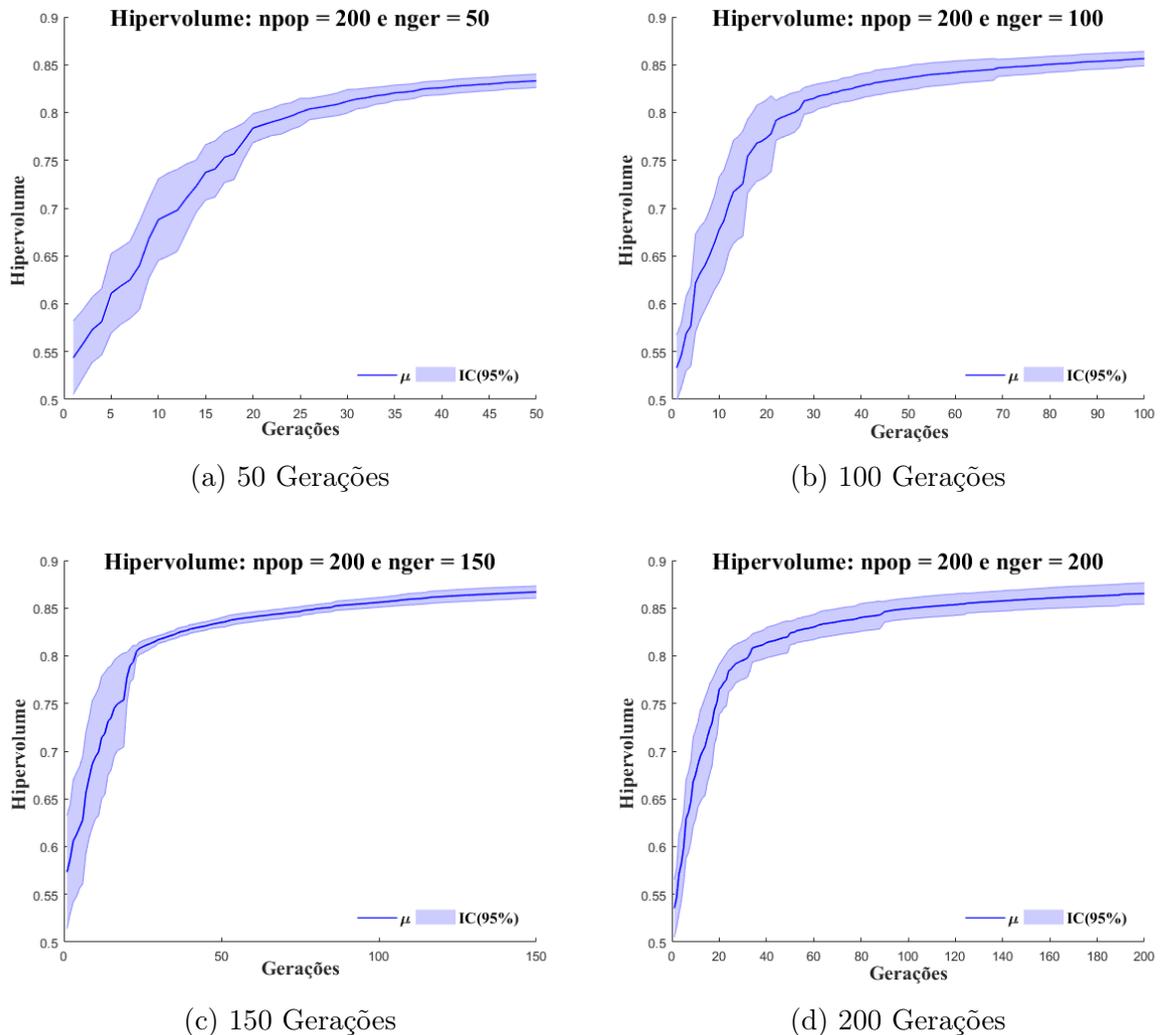


Figura 5.4: Análise de $ngger$ - Hipervolume

Ao aumentar o número máximo de gerações de 150 para 200, observou-se uma melhoria menor que 1% no valor da métrica usada para mensuração do desempenho do algoritmo. Neste caso, o esforço requerido na simulação adicional de 50 iterações foi considerado prescindível.

A partir destas análises definiu-se que para cada replicação do algoritmo, 150 iterações ou gerações de indivíduos seriam avaliadas, i.e., $ngger = 150$.

Este parâmetro foi mantido constante para determinar a probabilidade de ocorrência dos processos evolutivos. Neste caso, quatro cenários foram avaliados: (i) $p_m = 0,05$ e $p_c = 0,50$, (ii) $p_m = 0,03$ e $p_c = 0,75$, (iii) $p_m = 0,01$ e $p_c = 1,00$ e (iv) variação dinâmica de p_m e p_c conforme a geração corrente do algoritmo.

A variação dinâmica desses parâmetros no quarto cenário representa um mecanismo para

reduzir as chances do algoritmo estagnar em um ponto de ótimo local, e preservar ou restaurar níveis adequados de diversidade genética nas gerações de indivíduos.

Conforme Equações (5.2) e (5.3), nas primeiras iterações do algoritmo as chances de mutação são mais altas, já as de cruzamento entre indivíduos são mais baixas. Desta forma, é possível introduzir maior diversidade à população de indivíduos, evitando a convergência prematura do algoritmo nas primeiras iterações.

No decorrer das gerações, p_m diminui até alcançar o mínimo, e p_c aumenta até alcançar o máximo. Neste caso, o algoritmo prioriza buscas mais locais, preservando informações sobre a região de busca obtidas até então. Nesta pesquisa considerou-se $p_m^{min} = 0,01$, $p_m^{max} = 0,05$, $p_c^{min} = 0,50$ e $p_c^{max} = 1,00$.

$$p_m = p_m^{max} + (p_m^{min} - p_m^{max}) * \frac{ger - 1}{ger_{max} - 1} \quad (5.2)$$

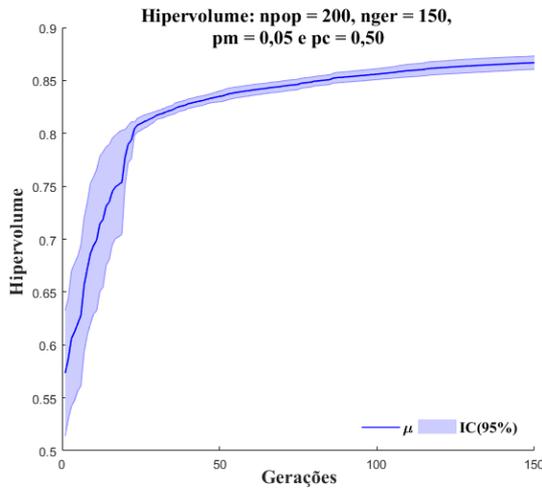
$$p_c = p_c^{min} + (p_c^{max} - p_c^{min}) * \frac{ger - 1}{ger_{max} - 1} \quad (5.3)$$

No intuito de avaliar melhor estes novos cenários, também foi plotado o Hipervolume, Figura 5.5.

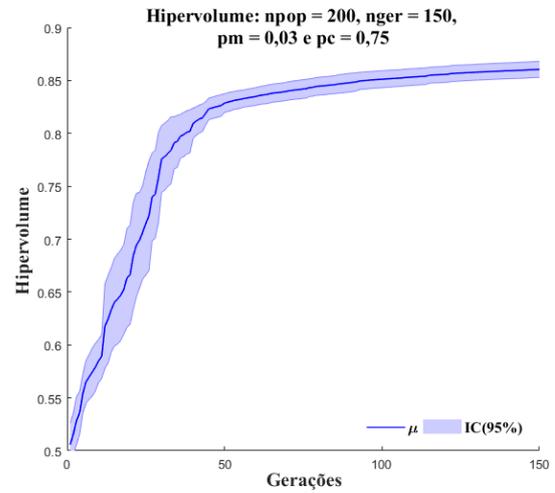
A Figura 5.5 apresenta o valor médio e o intervalo de confiança (95%) da área formadas entre o ponto de referência e a primeira fronteira de indivíduos não dominados encontrada ao final de cada iteração dos cenários simulados.

Observa-se que em relação à dispersão das soluções não dominadas e convergência do algoritmo, o quarto cenário se mostrou melhor. A partir destas análises definiu-se que p_m e p_c seriam atualizadas dinamicamente conforme a geração corrente do algoritmo e disposto nas Equações (5.2) e (5.3).

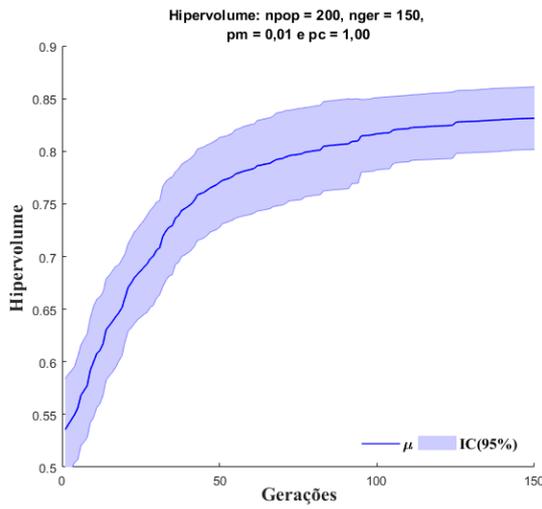
Na sequência são apresentados os resultados obtidos a partir da adoção dos parâmetros indicados na Tabela 5.6.



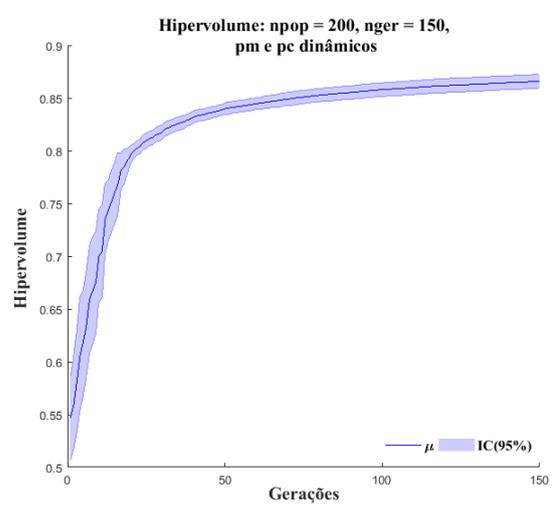
(a) $p_m = 0,05$ e $p_c = 0,50$



(b) $p_m = 0,03$ e $p_c = 0,75$



(c) $p_m = 0,01$ e $p_c = 1,00$



(d) Variação dinâmica de p_m e p_c

Figura 5.5: Análise de p_m e p_c - Hipervolume

Tabela 5.6: Parâmetros do algoritmo NSGA-II

Parâmetro	Descrição	Valores adotados
n_{pop}	Tamanho da população	200
n_{ger}	Número de gerações	150
p_m^{min}	Prob. mínima de mutação	0,01
p_m^{max}	Prob. máxima de mutação	0,05
p_c^{min}	Prob. mínima de cruzamento	0,50
p_c^{max}	Prob. máxima de cruzamento	1,00

5.4 Principais Resultados

Conforme apresentado, o objeto de estudo desta pesquisa consiste em um problema de otimização biobjetiva combinatória com variáveis inteiras e contínuas. Em problemas desta natureza, o

esforço computacional despendido na avaliação das soluções via simulação pode extrapolar o desejável na resolução de um modelo linear de médio porte, configurando uma dicotomia entre o tempo gasto na geração e análise das soluções candidatas (Fu, 2001). Além disso, a existência de um quantitativo significativo de combinações possíveis, e a diversidade do conjunto de variáveis dificulta a definição de um metamodelo confiável para avaliação das soluções propostas pela otimização.

Desta forma, nesta pesquisa foi adotada outra estratégia para redução do tempo gasto no processo de avaliação das soluções. No algoritmo proposto, os indivíduos da população são avaliados via Simulação Paralela, enquanto os operadores e processos evolutivos do NSGA-II ficam a cargo do programa central ou "mestre". Esta estratégia de paralelismo é chamada de Método Global, vide Figura 2.6. Nesta pesquisa foram utilizadas cinco estações de trabalho quad-core com processador i5 e 8GB de memória RAM. Conforme indicado na Tabela 5.7, cada replicação do algoritmo durou cerca de três horas.

Tabela 5.7: Tempo de cada replicação do algoritmo

Replicação	Duração (h)
1	03:26:38
2	03:16:14
3	03:05:31
4	03:05:46
5	03:15:41
6	03:18:36
7	02:57:44
8	03:04:16
9	03:46:35
10	03:11:03
Média	03:14:48
Desvio Padrão	00:13:57

Novamente, os resultados se basearam no cenário mais crítico do ponto de vista da otimização, no qual as taxas de internação equivalem aos valores máximos de referência da Portaria N° 1.631/2015. Produzindo, desta forma, a simulação da demanda máxima esperada na macro-região do triângulo sul mineiro. Devido aos fatores aleatórios dos operadores evolutivos, foram feitas 10 replicações do algoritmo para este cenário em estudo, permitindo a obtenção de resultados mais confiáveis.

A Figura 5.6 ilustra a primeira fronteira de indivíduos não dominados encontrada após rodar as 10 replicações do cenário simulado.

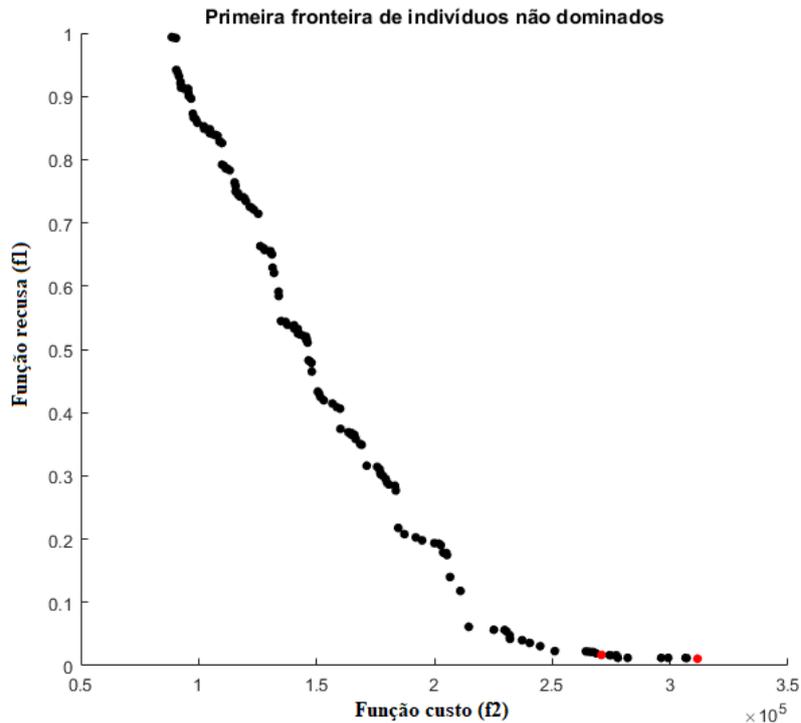


Figura 5.6: Primeira fronteira de indivíduos não dominados

Conforme esperado, inexistente uma solução que seja capaz de otimizar as funções objetivo do problema ao mesmo tempo. As soluções do extremo da fronteira por exemplo, otimizam um objetivo à custa do outro. Na ausência de qualquer informação adicional, nenhuma dessas soluções é melhor do que as outras da mesma fronteira encontrada. Desta forma, optou-se por analisar em detalhes a solução de menor recusa (Solução 1) e a solução com custo diário mais próximo ao da estrutura vigente em 2017 (Solução 2), destacadas em vermelho na Figura 5.6. A Tabela 5.8 apresenta alguns resultados obtidos a partir destas soluções. No total, 10 replicações dos cenários em estudo foram simuladas, permitindo o levantamento de resultados mais confiáveis, i.e. com IC de 95%.

Verifica-se que na Solução 1, o custo diário dos leitos, estimado pela Equação (3.4), é R\$ 311.593,44, e o intervalo de confiança da função de recusa, calculada via simulação, é $1,01\% \pm 0,08$ p.p.

Já na Solução 2, o custo diário dos leitos, estimado pela Equação (3.4), é R\$ 270.847,31, e o intervalo de confiança da função de recusa, calculada via simulação, é $1,68\% \pm 0,02$ p.p.

Para se obter um valor de referência aos resultados apresentados, o custo diário dos leitos da estrutura vigente em 2017 também foi estimado pela Equação (3.4), resultando em R\$ 270.958,97, cerca de 13% abaixo do valor obtido na Solução 1 e muito próximo ao da Solução 2. Isto foi feito

Tabela 5.8: Resultados das soluções em destaque

Resultados		Solução 1	Solução 2
Função Custo Diário		R\$ 311.593,44	R\$ 270.847,31
Função Recusa Máxima		1,01% \pm 0,08 p.p.	1,68% \pm 0,02 p.p.
Leitos	Cirúrgicos	270	234
	Clínicos	567	502
	Obstétrico	78	67
	Pediátrico	80	64
	Neonatal	44	38
	UTI Adulto	87	84
	UTI Pediatrico	22	22
	UTI Neonatal	31	30
	Total	1.179	1.041

dado que não existe hoje um banco de dados público que forneça tal informação.

O mesmo ocorre em relação à recusa. Entretanto, neste caso, não foi possível obter um valor de referência pela execução do algoritmo, pois, ao simular o contexto real de 2017, o algoritmo não consegue encontrar uma solução para o problema. Isto se deve ao fato de que em 2017, algumas cidades estavam a cerca de 400 km de distância do município mais próximo com UTI, violando, desta forma, a restrição de distância indicada pela Equação (3.9), na formulação matemática do problema. Além disso, para reproduzir o contexto de 2017 no modelo de simulação, seria preciso levantar a alocação dos pacientes para cada especialidade demandada nos municípios, tornando esta análise inviável.

Desta forma, os registros levantados em Pinto et al. (2014) foram usados como referência aos resultados apresentados. Em 2012, um centro de internações analisado por Pinto et al. (2014), registrou 224.455 pedidos de internação, o equivalente à 39% das demandas SUS na capital mineira. Entretanto, segundo os autores, 29.07%, i.e. 65.259 pacientes foram recusados. Ao comparar este registro aos resultados apresentados na Tabela 5.8, verifica-se que com a metodologia proposta é possível alcançar soluções viáveis e eficientes ao planejamento do sistema de internação estudado, i.e., soluções que procuram minimizar os custos e taxas de recusa dos leitos ao mesmo tempo.

Em relação ao número de leitos, na Solução 1 proposta pelo algoritmo, 1.179 leitos SUS seriam necessários para atendimento eficiente dos pacientes, i.e., com taxas de recusa de cerca de 1%. Já na Solução 2 seriam necessários 1.041 leitos públicos. Segundo CNES (2018), em 2017 existiam no triângulo sul mineiro 991 leitos SUS e 592 privados. Desta forma, a Solução 1 propõe contratar outros 188 leitos do sistema privado, enquanto na Solução 2 seria necessária

a contratação de outros 50 leitos. Desta forma, ambas respeitam as restrições definidas pelas Equações (3.6)-(3.13) na formulação matemática do problema.

Verifica-se que embora a Solução 1 aumente os custos diários dos leitos em 13%, ela também é capaz de reduzir as taxas de recusa em cerca de 1/30 dos valores indicados em Pinto et al. (2014). Já a Solução 2, apesar de manter os custos diários próximos ao da estrutura vigente em 2017, tem uma taxa média de recusa 66% maior que a alcançada na Solução 1. Conforme mencionado, na ausência de qualquer informação adicional, nenhuma dessas duas soluções é melhor do que a outra. Neste caso, caberia ao gestor ou tomador de decisão, baseado em critérios específicos, definir qual ponto da fronteira de soluções não dominadas seria interessante ou adequado à realidade do sistema.

Ressalta-se que no cenário em estudo as taxas de internação equivalem aos limites máximos de referência da Portaria N° 1.631/2015, produzindo, portanto, resultados ao cenário mais crítico do ponto de vista da otimização.

A seguir as Tabelas 5.9 - 5.11 e a Figura 5.7, demonstram o cumprimento das restrições impostas ao problema a partir de um estudo mais crítico dos resultados alcançados na Solução 1, i.e., de menor taxa de recusa.

A Tabela 5.9 exemplifica estes resultados na cidade de Araxá.

Tabela 5.9: Número de leitos em Araxá

Tipos de leito	Proposta (NSGA-II)	Públicos (2017)	Privada (2017)
Cirúrgico	23	22	42
Clínico	68	55	23
Obstétrico	9	12	30
Pediátrico	6	9	12
Neonatal	-	-	-
UTI Adulto	12	17	12
UTI Pediátrico	-	-	-
UTI Neonatal	-	-	-
Total	118	115	119

Conforme indicado na Tabela 5.9, na cidade de Araxá 13 leitos clínicos e 1 cirúrgico deveriam ser contratados do sistema privado. Mesmo com esta contratação extra, a Solução 1 respeita a restrição de capacidade do problema, definida pela Equação (3.6).

A configuração apresentada atende também as relações de precedência entre os leitos, definidas pela Equação (3.7), e.g. não havendo previsão de oferta de leitos neonatais na cidade de Araxá, nenhuma UTI desta mesma tipologia é esperada no município.

Outro resultado importante deste sistema se refere à ocupação dos leitos. Conforme sugerido

em Jones (2011) e considerado nesta pesquisa, os leitos devem operar próximos à ocupação máxima desejável, definida por sua capacidade, e não por valores médios ou globais do sistema. Neste caso, *pools* menores de leitos devem manter taxas mais baixas de ocupação, buscando registrar recusas semelhantes às registradas nos maiores *pools*. A Tabela 5.10 exemplifica estes resultados na cidade de Araxá.

Tabela 5.10: Ocupação média dos leitos em Araxá

Tipos de leito	Leitos	Ocupação
Cirúrgico	23	49,52% \pm 0,86 p.p.
Clínico	68	68,39% \pm 1,60 p.p.
Obstétrico	9	31,94% \pm 1,55 p.p.
Pediátrico	6	24,87% \pm 1,71 p.p.
Neonatal	-	-
UTI Adulto	12	37,48% \pm 3,44 p.p.
UTI Pediátrico	-	-
UTI Neonatal	-	-

A Tabela 5.11 indica ainda o atendimento à restrição de distância máxima permitida ao deslocamento dos pacientes na rede de saúde, definida pela Equação (3.9) na formulação matemática.

Tabela 5.11: Distância máxima percorrida pelos pacientes

Tipos de leito	Máxima simulada	Limite
Cirúrgico	289 km	300 km
Clínico	289 km	300 km
Obstétrico	289 km	300 km
Pediátrico	289 km	300 km
Neonatal	293 km	300 km
UTI Adulto	193 km	200 km
UTI Pediátrico	188 km	200 km
UTI Neonatal	193 km	200 km

A Figura 5.7 ilustra a alocação proposta as pacientes de obstétricos na macro-região de saúde do triângulo sul mineiro.

Os valores nas arestas representam o percentual de pacientes obstétricos alocados em outros locais que não os de sua residência, e.g. em Araxá, 47,58% dos pacientes obstétricos são encaminhados para Uberaba, enquanto os 52,42% restantes são atendidos no município, onde 9 leitos obstétricos foram alocados.

Nas cidades onde nenhum leito obstétrico é esperado, e.g. Veríssimo, 100% dos pacientes são encaminhados para outro município. Em alguns casos, porém, os recursos disponíveis em um

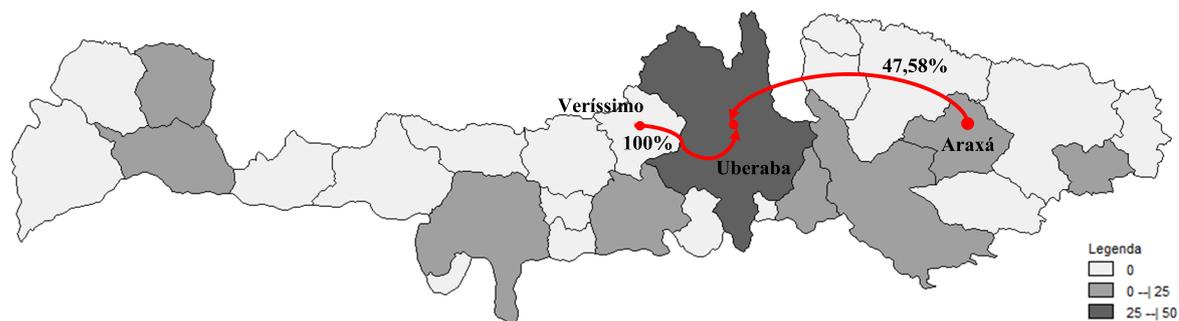


Figura 5.7: Alocação dos pacientes obstétricos

município são suficientes para atender não apenas a demanda local como também as de outros locais, e.g. em Uberaba diversos pacientes internados residem nas demais cidades da região.

Verifica-se, portanto, que pacientes residentes nos municípios com sistema de internação operante têm prioridade de atendimento, evitando assim encaminhá-los para outros locais enquanto for possível atendê-los *in loco*.

Os arcos da Figura 5.7 indicam que pacientes obstétricos são alocados apenas a locais com leitos compatíveis esperados. Além disso, o somatório do percentual de pacientes alocados no município de residência e encaminhados à outros locais com leitos de internação compatíveis, totaliza sempre 100% da demanda. Estes resultados estão alinhados às restrições de compatibilidade entre os leitos e especialidades, e a alocação de 100% dos pacientes, definidas pelas Equações (3.8) e (3.9) na formulação matemática do problema.

Em síntese, os resultados apresentados demonstram a capacidade da metodologia proposta nesta pesquisa em encontrar soluções viáveis e eficientes ao planejamento da capacidade e alocação dos pacientes no sistema público de internação estudado, i.e. soluções que procuram minimizar os custos e taxas de recusa dos leitos.

Capítulo 6

Conclusões

Nesta pesquisa foi proposta uma metodologia de Otimização via Simulação para dimensionar a capacidade de atendimento e a alocação dos pacientes em um sistema público de internação em Minas Gerais.

Apesar de estar entre os problemas mais debatidos na área da saúde, até o ano de 2015 as decisões acerca do sistema público de internação brasileiro eram pautados na Portaria nº 1.101/2002, que utilizava como referência valores médios e globais de internação.

Em conformidade aos parâmetros propostos na Portaria nº 1.631/2015, que revogou a Portaria nº 1.101/2002, na modelagem matemática apresentada no Capítulo 3 desta pesquisa, os parâmetros e variáveis do problema foram discriminados por especialidade, município de residência/atendimento dos pacientes, e tipo de leito. Desta forma, garantiu-se a abordagem de subsistemas mais homogêneos, com redução de erros na estimativa da necessidade real de leitos, e.g. altas taxas de recusa ou desperdício dos recursos públicos.

Além disso, na formulação do problema duas funções objetivo conflitantes foram consideradas, a minimização das taxas de recusa e o custo dos leitos em operação no sistema público. A função custo diário dos leitos não foi representada por uma constante, mas sim por uma função que depende do número e tipo de leito em questão, de forma à considerar na otimização a lógica básica da economia de escala.

Levou-se em conta, também, a relação entre as taxas de ocupação e recusa dos leitos, pois, conforme sugerido em Jones (2011), para pequenos *pools* de leitos operarem com taxas de recusa semelhantes às observadas nos grandes centros, é preciso garantir menores níveis de ocupação.

Basicamente, três alternativas poderiam viabilizar a otimização do sistema estudado: (i) aquisição de novos leitos públicos; (ii) contratação de leitos do sistema privado ou (iii) realoca-

ção dos leitos SUS existentes. Diante das restrições financeiras, a primeira alternativa não foi considerada, i.e., investigamos se realocando os leitos e fazendo novas contratações do sistema privado seria possível atender a demanda existente com custos e taxas de recusa mínimos.

A escolha do Algoritmo Genético NSGA-II para condução da otimização, baseou-se em alguns aspectos desse algoritmo adequados à tratativa do problema analisado, como:

1. Capacidade de analisar problemas de otimização sem informação preliminar do gradiente das funções objetivo, permitindo sua aplicação em problemas menos estruturados.
2. Avaliação de um conjunto de soluções por iteração, garantindo a abordagem de problemas com múltiplos objetivos.
3. Utilização de operadores estocásticos na condução da otimização, favorecendo a continuidade de soluções desejáveis no decorrer dos processos evolutivos.

Entretanto a dinamicidade, aleatoriedade e especificidades do problema analisado, inviabilizaram a formulação analítica de uma de suas funções objetivo, nos levando à integrar a Simulação à Eventos Discretos na avaliação das soluções geradas pelo NSGA-II, em um processo conhecido como Otimização via Simulação. Os resultados da simulação são utilizados pelo NSGA-II para comparar as soluções encontradas e estabelecer novas direções de busca do algoritmo.

Para reduzir o esforço computacional requisitado no processo de avaliação das soluções, nesta pesquisa foi utilizado o Método Global de computação paralela. No algoritmo proposto, os indivíduos da população são avaliados via Simulação Paralela, enquanto os operadores e processos evolutivos do NSGA-II ficam a cargo do programa central ou "mestre".

Como as duas funções objetivo do problema analisado são conflitantes, nenhuma das soluções não dominadas obtidas pelo NSGA-II foi capaz de otimizá-las ao mesmo tempo. Desta forma, optou-se por analisar em detalhes a solução de menor recusa (Solução 1) e a solução com custo diário mais próximo ao da estrutura vigente em 2017 (Solução 2).

Os resultados da Solução 1 mostraram que 1.179 leitos SUS seriam necessários para atendimento eficiente dos pacientes na região em estudo, i.e., com taxas de recusa de cerca de 1%. E que 188 leitos extras devem ser contratados do sistema privado para alcançar melhores níveis de serviço. Já na Solução 2, cujos custos diários são mais próximos ao da estrutura vigente em 2017, os resultados mostraram que 1.041 leitos SUS seriam necessários, sendo 50 leitos extras contratados do sistema privado.

Verificou-se que embora a Solução 1 aumente os custos diários dos leitos em 13%, ela também é capaz de reduzir as taxas de recusa em cerca de 1/30 dos valores indicados em Pinto et al. (2014). Já a Solução 2, apesar de manter os custos diários próximos ao da estrutura vigente em 2017, tem uma taxa média de recusa 66% maior que a alcançada na Solução 1. Conforme mencionado, na ausência de qualquer informação adicional, nenhuma dessas duas soluções é melhor do que a outra. Neste caso, caberia ao gestor ou tomador de decisão, baseado em critérios específicos, definir qual ponto da fronteira de soluções não dominadas seria interessante ou adequado à realidade do sistema.

Após análise destas soluções, concluiu-se que os leitos disponíveis nos municípios e a alocação dos pacientes afeta diretamente a performance do sistema de internação estudado, i.e., nas taxas de recusa e ocupação praticadas. Além disso, verifica-se que ao realocar alguns leitos já instalados na macro-região estudada, e contratar outros leitos do sistema privado, é possível minimizar as taxas de recusa e custos de operação desse sistema.

Os resultados alcançados demonstram a capacidade da metodologia proposta em encontrar soluções eficientes ao sistema estudado, i.e., as restrições são completamente respeitadas pelas soluções não dominadas encontradas, que buscam otimizar as taxas de recusa e o custo dos leitos, e também de reduzir os problemas agravados pelo planejamento ineficiente deste sistema como: (i) altos riscos de infecção dos pacientes alocados em condições precárias, (ii) longas filas de espera por leitos, e (iii) desperdício de investimentos públicos em recursos mal estimados.

A pesquisa também contribuiu para atender algumas questões ainda latentes na Otimização via Simulação levantadas em Amaran et al. (2014), como:

1. Capacidade de representar problemas NP-Difíceis, com múltiplos objetivos e variáveis contínuas e discretas trabalhadas conjuntamente.
2. Análise de diferentes restrições estocásticas e determinísticas;
3. Tratamento adequado de múltiplos *outputs* na simulação;
4. Consolidação e cooperação de diversas técnicas de simulação e otimização em algoritmos híbridos.

Desta forma, a metodologia apresentada configura uma alternativa eficiente para determinar a capacidade e a alocação dos leitos no Brasil, país onde na ausência de diretrizes adequadas, decisões estratégicas de planejamento público são frequentemente fundamentadas em perspectivas simplistas, puramente políticas ou pragmáticas.

Em trabalhos futuros, a metodologia poderá ser aplicada a outras regiões de saúde, e dados reais mais atuais poderão ser usados para apoiar os gestores nos processos de tomada de decisão.

Além disso, pesquisas futuras poderão explorar outras direções, como:

1. Aplicação de técnicas de Planejamento de Experimentos para calibração dos parâmetros de entrada do algoritmo.
2. Estudo de outros métodos de apoio à tomada de decisão para escolha da solução não dominada mais adequada ao problema analisado.
3. Consideração da sazonalidade e variabilidade das demandas para analisar os impactos na ocupação de leitos e taxas de recusa.
4. Proposição de metodologias e estratégias para apoiar a tomada de decisões de curto prazo, e.g., definir quais pacientes receberão atendimento local e quais serão encaminhados para internação em outras cidades à luz dos resultados obtidos nesta pesquisa.

Referências Bibliográficas

- Abraham, A. and Jain, L. (2005). Evolutionary multiobjective optimization. In *Evolutionary Multiobjective Optimization*, pages 1–6. Springer.
- Adewumi, A. O. and Ali, M. M. (2010). A multi-level genetic algorithm for a multi-stage space allocation problem. *Mathematical and Computer Modelling*, 51(1):109–126.
- Afshari, H. and Peng, Q. (2014). Challenges and solutions for location of healthcare facilities. *Industrial Engineering and Management*, 3(2):1–12.
- Ahmadi-Javid, A., Seyedi, P., and Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, 79:223–263.
- Akcali, E., Coˆt e, M. J., and Lin, C. (2006). A network flow approach to optimizing hospital bed capacity decisions. *Health Care Management Science*, 9(4):391–404.
- Alizadeh, M., Eskandari, H., Sajadifar, S. M., and Geiger, C. D. (2011). Analyzing a stochastic inventory system for deteriorating items with stochastic lead time using simulation modeling. In *Proceedings of the Winter Simulation Conference*, pages 1650–1662. Winter Simulation Conference.
- Amaran, S., Sahinidis, N. V., Sharda, B., and Bury, S. J. (2014). Simulation optimization: a review of algorithms and applications. *4OR*, 12(4):301–333.
- ANS (2018). Ag ncia Nacional de Sa de Suplementar. <http://www.ans.gov.br/>. Acessado: 2019-02-14.
- Aringhieri, R., Bruni, M. E., Khodaparasti, S., and Van Essen, J. (2017). Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368.

- Bachouch, R. B., Guinet, A., and Hajri-Gabouj, S. (2012). An integer linear model for hospital bed planning. *International Journal of Production Economics*, 140(2):833–843.
- Bäck, T. (1963). *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithm*. Oxford University Press.
- Bair, A. E., Song, W. T., Chen, Y.-C., and Morris, B. A. (2010). The impact of inpatient boarding on ed efficiency: a discrete-event simulation study. *Journal of medical systems*, 34(5):919–929.
- Banks, J., Carson, J., and Nelson, B. (2000). *Discrete-Event System Simulation*. Prentice hall Englewood Cliffs, NJ, USA.
- Barton, R. R. (1992). Metamodels for simulation input-output relations. In *Proceedings of the 24th conference on Winter simulation*, pages 289–299. ACM.
- Bastos, L. S., Marchesi, J. F., Hamacher, S., and Fleck, J. L. (2019). A mixed integer programming approach to the patient admission scheduling problem. *European Journal of Operational Research*, 273(3):831–840.
- Belciug, S. and Gorunescu, F. (2016). A hybrid genetic algorithm-queuing multi-compartment model for optimizing inpatient bed occupancy and associated costs. *Artificial intelligence in medicine*, 68:59–69.
- Bidhandi, H. M., Patrick, J., Noghani, P., and Varshoei, P. (2019). Capacity planning for a network of community health services. *European Journal of Operational Research*, 275(1):266–279.
- Brasil (2002). Portaria n° 1.101/2002: estabelece parâmetros assistenciais no sus. <http://bvsms.saude.gov.br/bvs/saudelegis.html>. Acessado em 7 Março de 2016.
- Brasil (2015). Portaria n° 1.631/2015: aprova critérios e parâmetros para o planejamento e programação de ações e serviços de saúde no âmbito do sus. <http://bvsms.saude.gov.br/bvs/saudelegis.html>. Acessado em 7 Março de 2016.
- Bruno, G., Esposito, E., Genovese, A., and Piccolo, C. (2016). Institutions and facility mergers in the italian education system: Models and case studies. *Socio-Economic Planning Sciences*, 53:23–32.

- Cao, J. X., Xue, C. C., Jian, M. Y., and Yao, X. R. (2018). Research on the station location problem for public bicycle systems under dynamic demand. *Computers & Industrial Engineering*.
- Ceschia, S. and Schaerf, A. (2011). Local search and lower bounds for the patient admission scheduling problem. *Computers & Operations Research*, 38(10):1452–1463.
- Chen, N., Zhan, Z.-h., Zhang, J., Liu, O., and Liu, H.-l. (2010). A genetic algorithm for the optimization of admission scheduling strategy in hospitals. In *IEEE Congress on Evolutionary Computation*, pages 1–5. IEEE.
- Chen, P.-S. and Lin, M.-H. (2017). Development of simulation optimization methods for solving patient referral problems in the hospital-collaboration environment. *Journal of biomedical informatics*, 73:148–158.
- Cheng, R., Jin, Y., Olhofer, M., and Sendhoff, B. (2016). A reference vector guided evolutionary algorithm for many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791.
- CNES (2018). Cadastro Nacional de Estabelecimentos de Saúde. <http://cnes.datasus.gov.br/br>. Acessado em 5 de Fevereiro de 2019.
- Coello, C. A. C. and Sierra, M. R. (2004). A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. In *Mexican International Conference on Artificial Intelligence*, pages 688–697. Springer.
- Coello, C. C. (2006). Evolutionary multi-objective optimization: a historical view of the field. *IEEE computational intelligence magazine*, 1(1):28–36.
- Coello, C. C., Lamont, G. B., and Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer Science & Business Media.
- Das, I. and Dennis, J. E. (1997). A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural optimization*, 14(1):63–69.
- DATASUS (2017). Sistema de Informação Hospitalar. <http://datasus.saude.gov.br>. Acessado: 2019-01-18.

- Deb, K. (2011). Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*, pages 3–34. Springer.
- Deb, K. and Jain, H. (2014). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197.
- Diefenbach, M. and Kozan, E. (2011). Effects of bed configurations at a hospital emergency department. *Journal of Simulation*, 5(1):44–57.
- Eskandari-Khanghahi, M., Tavakkoli-Moghaddam, R., Taleizadeh, A. A., and Amin, S. H. (2018). Designing and optimizing a sustainable supply chain network for a blood platelet bank under uncertainty. *Engineering Applications of Artificial Intelligence*, 71:236–250.
- Farahani, R. Z., SteadieSeifi, M., and Asgari, N. (2010). Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, 34(7):1689–1709.
- Fonseca, C. M. and Fleming, P. J. (1993). Multiobjective genetic algorithms. In *Genetic algorithms for control systems engineering, IEE colloquium on*, pages 6–1. IET.
- Fonseca, C. M., Knowles, J. D., Thiele, L., and Zitzler, E. (2005). A tutorial on the performance assessment of stochastic multiobjective optimizers. In *Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 216, page 240.
- Fu, M. C. (2001). Simulation optimization. In *Proceedings of the 33rd conference on Winter simulation*, pages 53–61. IEEE Computer Society.
- Fu, M. C. (2002). Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing*, 14(3):192–215.
- Fu, M. C. (2013). Simulation optimization. In *Encyclopedia of Operations Research and Management Science*, pages 1418–1423. Springer.
- Fu, M. C. et al. (2015). *Handbook of simulation optimization*, volume 216. Springer.

- Gansterer, M., Almeder, C., and Hartl, R. F. (2014). Simulation-based optimization methods for setting production planning parameters. *International Journal of Production Economics*, 151:206–213.
- Guerriero, F., Miglionico, G., and Olivito, F. (2016). Location and reorganization problems: The calabrian health care system case. *European Journal of Operational Research*, 250(3):939–954.
- Günel, M. M. and Pidd, M. (2010). Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1):42–51.
- Gunantara, N. (2018). A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1):1–16.
- Holm, L. B., Lurås, H., and Dahl, F. A. (2013). Improving hospital bed utilisation through simulation and optimisation: with application to a 40% increase in patient volume in a norwegian general hospital. *International journal of medical informatics*, 82(2):80–89.
- Hong, L. J. and Nelson, B. L. (2009). A brief introduction to optimization via simulation. In *Winter Simulation Conference*, pages 75–85. Winter Simulation Conference.
- IBGE (2015). Instituto Brasileiro de Geografia e Estatística. <https://www.ibge.gov.br>. Acessado: 2018-03-10.
- Jacobson, S. H., Hall, S. N., and Swisher, J. R. (2006). Discrete-event simulation of health care systems. In *Patient flow: Reducing delay in healthcare delivery*, pages 211–252. Springer.
- Jones, D. F., Mirrazavi, S. K., and Tamiz, M. (2002). Multi-objective meta-heuristics: An overview of the current state-of-the-art. *European journal of operational research*, 137(1):1–9.
- Jones, R. (2011). Hospital bed occupancy demystified. *British Journal of Healthcare Management*, 17(6):242–248.
- Juan, A. A., Faulin, J., Grasman, S. E., Rabe, M., and Figueira, G. (2015). A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives*, 2:62–72.
- Jun, J., Jacobson, S. H., and Swisher, J. R. (1999). Application of discrete-event simulation in health care clinics: A survey. *Journal of the operational research society*, 50(2):109–123.

- Karmellos, M., Kiprakis, A., and Mavrotas, G. (2015). A multi-objective approach for optimal prioritization of energy efficiency measures in buildings: Model, software and case studies. *Applied Energy*, 139:131–150.
- Kasaie, P. and Kelton, W. D. (2013). Simulation optimization for allocation of epidemic-control resources. *IIE Transactions on Healthcare Systems Engineering*, 3(2):78–93.
- Kelton, W. D. and Law, A. M. (2000). *Simulation modeling and analysis*. McGraw Hill Boston.
- Knowles, J. and Corne, D. (1999). The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation. In *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, volume 1. IEEE.
- Konak, A., Coit, D. W., and Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 91(9):992–1007.
- Kozeny, V. (2015). Genetic algorithms for credit scoring: Alternative fitness function performance comparison. *Expert Systems with Applications*, 42(6):2998–3004.
- Lee, C. W. and Kwak, N. (2011). Strategic enterprise resource planning in a health-care system using a multicriteria decision-making model. *Journal of medical systems*, 35(2):265–275.
- Lee, L. H., Chew, E. P., Teng, S., and Chen, Y. (2008). Multi-objective simulation-based evolutionary algorithm for an aircraft spare parts allocation problem. *European Journal of Operational Research*, 189(2):476–491.
- Li, J., Huang, Y., and Niu, X. (2016). A branch population genetic algorithm for dual-resource constrained job shop scheduling problem. *Computers & Industrial Engineering*, 102:113–131.
- Lin, R.-C., Sir, M. Y., and Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega*, 41(5):881–892.
- Malachias, I., Leles, F. A. G., Pinto, M. A. S., de Andrade, L. C. F., de Alencar, F. B., da Silva, A. E., de Almeida, H. D., and de Siqueira, M. (2011). Plano diretor de regionalização da saúde de minas gerais (pdr-mg). *Belo Horizonte*.
- Marynissen, J. and Demeulemeester, E. (2016). Literature review on integrated hospital scheduling problems (technical report). *Faculty of Economics and Business, KU Leuven*.

- Nogueira, L. C., Pinto, L. R., and Silva, P. M. S. (2016). Reducing emergency medical service response time via the reallocation of ambulance bases. *Health Care Management Science*, 19(1):31–42.
- Pierreval, H. and Paris, J.-L. (2000). Distributed evolutionary algorithms for simulation optimization. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(1):15–24.
- Pinto, L. R., Campos, F. C. C., Oliva Perpetuo, I. H., Ribeiro, N. M. B., and Cristina, Y. (2014). Analysis of hospital bed capacity via queuing theory and simulation. In *Simulation Conference (WSC), 2014 Winter*, pages 1281–1292. IEEE.
- Pitaksringkarn, L. and Taylor, M. A. (2005). Grouping genetic algorithm in gis: A facility location modelling. *Journal of the Eastern Asia Society for Transportation Studies*, 6:2908–2920.
- Rahman, S.-u. and Smith, D. K. (2000). Use of location-allocation models in health service development planning in developing nations.
- Range, T. M., Lusby, R. M., and Larsen, J. (2014). A column generation approach for solving the patient admission scheduling problem. *European Journal of Operational Research*, 235(1):252–264.
- reynolds, J., Nafpliotis, N., and Goldberg, D. E. (1994). A niched pareto genetic algorithm for multiobjective optimization. In *Proceedings of the first IEEE conference on evolutionary computation, IEEE world congress on computational intelligence*, volume 1, pages 82–87. Citeseer.
- Rich, N. and Piercy, N. (2013). Losing patients: a systems view on healthcare improvement. *Production Planning & Control*, 24(10-11):962–975.
- Schott, J. R. (1995). Fault tolerant design using single and multicriteria genetic algorithm optimization. Technical report, DTIC Document.
- Shi, Y., Boudouh, T., and Grunder, O. (2017). A hybrid genetic algorithm for a home health care routing problem with time window and fuzzy demand. *Expert Systems with Applications*, 72:160–176.
- Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F. (2001). Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA journal*, 39(12):2233–2241.

- Soleimani, H. and Kannan, G. (2015). A hybrid particle swarm optimization and genetic algorithm for closed-loop supply chain network design in large-scale networks. *Applied Mathematical Modelling*, 39(14):3990–4012.
- Srinivas, N. and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248.
- Stummer, C., Doerner, K., Focke, A., and Heidenberger, K. (2004). Determining location and size of medical departments in a hospital network: A multiobjective decision support approach. *Health Care Management Science*, 7(1):63–71.
- Swisher, J. R., Hyden, P. D., Jacobson, S. H., and Schruben, L. W. (2000). A survey of simulation optimization techniques and procedures. In *Simulation Conference, 2000. Proceedings. Winter*, volume 1, pages 119–128. IEEE.
- Tavana, M., Abtahi, A.-R., Di Caprio, D., Hashemi, R., and Yousefi-Zenouz, R. (2018). An integrated location-inventory-routing humanitarian supply chain network with pre-and post-disaster management considerations. *Socio-Economic Planning Sciences*, 64:21–37.
- Van Veldhuizen, D. A. (1999). Multiobjective evolutionary algorithms: classifications, analyses, and new innovations. Technical report, DTIC Document.
- Xu, W., Chong, A., Karaguzel, O. T., and Lam, K. P. (2016). Improving evolutionary algorithm performance for integer type multi-objective building system design optimization. *Energy and Buildings*, 127:714–729.
- Zarrinpoor, N., Fallahnezhad, M. S., and Pishvaei, M. S. (2017). Design of a reliable hierarchical location-allocation model under disruptions for health service networks: A two-stage robust approach. *Computers & Industrial Engineering*, 109:130–150.
- Zhang, J. and Xing, L. (2017). A survey of multiobjective evolutionary algorithms. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, volume 1, pages 93–100. IEEE.
- Zhang, Q. and Li, H. (2007). Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6):712–731.

- Zhang, W., Cao, K., Liu, S., and Huang, B. (2016). A multi-objective optimization approach for health-care facility location-allocation problems in highly developed cities such as hong kong. *Computers, Environment and Urban Systems*, 59:220–230.
- Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P. N., and Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1):32–49.
- Zhou, L., Geng, N., Jiang, Z., and Wang, X. (2018). Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega*, 81:220–233.
- Zitzler, E., Laumanns, M., and Bleuler, S. (2004). A tutorial on evolutionary multiobjective optimization. In *Metaheuristics for multiobjective optimisation*, pages 3–37. Springer.
- Zitzler, E., Laumanns, M., and Thiele, L. (2001). Spea2: Improving the strength pareto evolutionary algorithm. *TIK-report*, 103.
- Zitzler, E. and Thiele, L. (1998). Multiobjective optimization using evolutionary algorithms—a comparative case study. In *International Conference on Parallel Problem Solving from Nature*, pages 292–301. Springer.
- Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Da Fonseca, V. G. (2003). Performance assessment of multiobjective optimizers: an analysis and review. *IEEE transactions on evolutionary computation*, 7(2):117–132.