

CAMILA LAURA PAREJA YALE

**PREVISÃO DA DEMANDA DE SUPRIMENTOS
DE SOCORRO COM APLICAÇÃO NA
LOCALIZAÇÃO DE DEPÓSITOS DA DEFESA
CIVIL DO ESTADO DE SÃO PAULO**

São Paulo
2022

CAMILA LAURA PAREJA YALE

**PREVISÃO DA DEMANDA DE SUPRIMENTOS
DE SOCORRO COM APLICAÇÃO NA
LOCALIZAÇÃO DE DEPÓSITOS DA DEFESA
CIVIL DO ESTADO DE SÃO PAULO**

Versão Corrigida

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do Título de Mestre em
Ciências.

São Paulo
2022

CAMILA LAURA PAREJA YALE

**PREVISÃO DA DEMANDA DE SUPRIMENTOS
DE SOCORRO COM APLICAÇÃO NA
LOCALIZAÇÃO DE DEPÓSITOS DA DEFESA
CIVIL DO ESTADO DE SÃO PAULO**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do Título de Mestre em
Ciências.

Área de Concentração:
Engenharia de Produção

Orientador:
Prof. Dr. Hugo Tsugunobu Yoshida
Yoshizaki

São Paulo
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

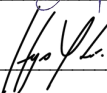
Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, __17__ de __dezembro__ de __2022__

Assinatura do autor:



Assinatura do orientador:



Catálogo-na-publicação

Yale, Camila Laura

Previsão da demanda de suprimentos de socorro com aplicação na localização de depósitos da Defesa Civil do Estado de São Paulo / C. L. Yale -- versão corr. -- São Paulo, 2022.

88 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Logística humanitária I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

AGRADECIMENTOS

Ao meu esposo pela paciência, disposição e apoio incondicional.

À minha família e amigos pela força e constante motivação.

Ao meu orientador Prof. Hugo e toda a equipe de pesquisa pelas ricas discussões e o contínuo intercâmbio de conhecimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos e o suporte econômico.

RESUMO

No Estado de São Paulo desastres naturais de baixa e media magnitude acontecem frequentemente no verão em consequência das chuvas. A Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC-SP) é responsável pelo abastecimento de ajuda humanitária aos 645 municípios. Portanto, estimativas de demanda por suprimentos de socorro são de vital importância para auxiliar na resposta ágil e eficiente após a ocorrência de sinistros. A previsão de suprimentos de emergência é uma tarefa complexa pela natureza inerente aos desastres; informações limitadas, quantidade excessiva de zeros na variável de resposta e fatores preditores não lineares são os principais desafios no desenvolvimento do modelo de previsão. Nessa perspectiva, os Generalized Linear Models (GLM) foram explorados para a modelagem da previsão da demanda de materiais de socorro. Os preditores do modelo incluem indicadores de vulnerabilidade, aspectos demográficos e anomalias climáticas relacionadas ao fenômeno El Niño. Dada a existência de vulnerabilidades diferenciadas nos municípios, é proposta a estrutura de modelo multinível Generalized Linear Mixed Models (GLMM) que leva em consideração esses agrupamentos hierárquicos evidenciados no conjunto de dados. Assim, o Zero-Inflated Negative Binomial Multilevel (ZINBM) foi selecionado como o modelo final para estimar a demanda de suprimentos de socorro.

Ao reduzir a incerteza na demanda, soluções adequadas para o problema de localização de depósitos da CEPDEC-SP podem ser atingidas, com esse propósito foi desenvolvido um modelo de otimização que busca minimizar os custos totais e maximizar o nível de serviço, assim como um Sistema de Apoio à Decisão (SAD) que possibilita à equipe da CEPDEC-SP avaliar diversas alternativas de localização para optar pela solução mais conveniente.

Palavras-Chave – Sistema de Apoio a Decisão, Modelo de Previsão de Demanda, Logística Humanitária, Zero-Inflated, Modelo Hierárquico.

ABSTRACT

In the State of São Paulo, natural disasters of low and medium magnitude often occur in the summer as a result of the rains. The Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC-SP) is responsible for supplying humanitarian aid to the 645 municipalities. Therefore, estimates of demand for relief supplies are of vital importance to assist in an agile and efficient response after the occurrence of accidents. Forecasting emergency supplies is a complex task due to the inert nature of the disasters; limited information, excessive amount of zeros in the response variable and non-linear predictors are the main challenges in developing the forecasting model. From this perspective, Generalized Linear Models (GLM) were explored to model the forecast of demand for relief materials. The model's predictors include vulnerability indicators, demographic aspects and climatic anomalies related to the El Niño phenomenon. Given the existence of differentiated vulnerabilities in the municipalities, the structure of a multilevel model GLMM is proposed, which takes in account these hierarchical groups evidenced in the dataset. Thus, a Zero-Inflated Negative Binomial Multilevel (ZINBM) model was selected as the final model to estimate demand for relief supplies.

By reducing demand uncertainty, adequate solutions to the CEPDEC-SP deposit location problem can be achieved, with this purpose an optimization model was developed that aims to minimize the total cost and maximize the level of service, as well as a Support Decision System (SDS) that allows the staff at CEPDEC-SP to evaluate different localization alternatives to choose the most convenient solution.

Keywords – Decision Support System, Demand Forecasting Model, Humanitarian Logistics, Zero-Inflated, Hierarchical Model.

LISTA DE FIGURAS

1	Fatores que influem nos desastres naturais	26
2	Ciclo de vida dos desastres	27
3	Clusterização de co-ocorrência de palavras-chaves	43
4	Composição da demanda por tipo de suprimento de socorro	46
5	Cesta básica e parte do kit de limpeza estocado no armazém central da CEPDEC-SP	46
6	Histograma da demanda média de suprimentos de socorro	47
7	Demanda de suprimentos de socorro por ano	48
8	Boxplot da demanda de suprimentos de socorro	51
9	Coeficiente silhueta por número de clusters	55
10	Agrupamento dos municípios do Estado de São Paulo.	56
11	Resultado do teste para verificação de existência de superdispersão	56
12	Histograma da demanda real x demanda estimada	60
13	Visão geográfica da demanda real x demanda estimada	61
14	Aba inicial “Explicações” do SAD	86
15	Aba “Modelo Iterativo” do SAD	87
16	Gráfico de Curva de Pareto	87
17	Informação de uma solução específica	88

LISTA DE TABELAS

1	Exemplos de desastres por modo de início	27
2	Artigos de modelos de previsão de desastres com aplicações humanitárias .	38
3	Resumo de artigos que desenvolveram modelos de previsão de demanda humanitária	39
4	Fatores preditores de previsão da demanda	49
5	Estatísticas descritivas ou frequência das variáveis envolvidas no modelo .	50
6	Resultados do teste de Voung do modelo de regressão BN x ZINB	57
7	Estimativas do modelo BN, ZINB e ZINBM	58
8	Comparação de medidas de qualidade de ajuste dos modelos	59
9	Modelos de localização de instalações	80

GLOSSÁRIO

- AF** Accuracy factor. 40
- AIC** Akaike Information Criterion. 32, 36, 53, 57, 59, 62, 63
- AM** Aprendizagem de Máquinas. 41
- ARIMA** Autoregressive integrated moving average. 39–41
- AUSPIN** Agência USP de inovação. 88
- BIC** Bayesian Information Criterion. 32, 36, 57
- BN** Binomial Negativo. 31–33, 44, 52, 53, 56–59, 63
- BP** back propagation. 39
- CBR** Case-Based Reasoning. 39–41
- CEPDEC-SP** Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo. 20–23, 45–47, 52, 55, 83, 85
- DAEE** Departamento de Águas e Energia Elétrica. 47
- EM** Ensemble Model. 40
- EMD** Empirical Mode Decomposition. 39
- ENSO** El Niño Southern Oscillation. 49, 64
- ETS** Exponential Smoothing. 40
- EWMA** Exponentially Weighted Moving Average. 40
- FSG** Forecasting for Social Good. 37
- GLM** Generalized Linear Models. 28, 29, 35, 36, 52, 59
- GLMM** Generalized Linear Mixed Models. 35, 52

HM Hurdle Models. 43, 44, 64

IA Inteligência Artificial. 41

IBGE Instituto Brasileiro de Geografia e Estatística. 49, 50

IDH Índice de Desenvolvimento Humano. 39, 47

IF Intuitionistic Fuzzy. 40

IIF International Institute of Forecasters. 37

IPVS Índice Paulista de Vulnerabilidade Social. 49

LSDV Least Squared Dummy Variable. 39

LSTM Long Short-Term Memory. 40

MA Moving average. 40

MAE Mean absolute Error. 39–41

MAPE Mean Absolute Percentage Error. 39, 40

MRA Multivariate Regression Analysis. 39

NN Neural Network. 39, 41

NSE Nash-Sutcliffe efficiency. 40

NTC&Logística Associação Nacional do Transporte de Cargas e Logística. 83

PB Percentage Bias. 40

PCA Principal Component Analysis. 39, 41

PCR Principal Component Regression. 40

RBF Radial Basis Function. 39

RMSE Root Mean Squared Error. 39–41

RPCA Robust Principal Component Analysis. 40

SAD Sistema de Apoio à Decisão. 22, 23, 64, 83, 85–88

SVM Support Vector Machine. 39, 41

SVR Support Vector Regression. 39, 40

TRC Transporte de Cargas Rodoviárias. 83

ZI Zero-Inflated. 32, 33, 35, 36, 43, 44, 64

ZINB Zero-Inflated Negative Binomial. 22, 34, 44, 52, 53, 56–60, 63

ZINBM Zero-Inflated Negative Binomial Multilevel. 44, 50, 52, 53, 55, 57–60, 62–64

ZIP Zero-Inflated Poisson. 33, 34, 52, 59, 63, 64

ZIPM Zero-Inflated Poisson Multilevel. 64

SUMÁRIO

1	Introdução	19
1.1	Contexto e Motivação	20
1.2	Objetivos	23
1.2.1	Perguntas da pesquisa	23
2	Contextualização Teórica	25
2.1	Desastres e Logística Humanitária	25
2.2	Modelos de previsão de demanda	28
2.2.1	Generalized Linear Models (GLM)	28
2.2.2	Modelos de regressão para dados de contagem	29
2.2.2.1	Modelo de regressão de Poisson	30
2.2.2.2	Modelo de regressão Binomial Negativo (BN)	31
2.3	Modelo de regressão Zero-Inflated (ZI)	32
2.3.1	Modelo de regressão Zero-Inflated Poisson (ZIP)	33
2.3.2	Modelo de regressão Zero-Inflated Negative Binomial (ZINB)	34
2.4	Generalized Linear Mixed Models (GLMM)	35
2.5	Métricas de avaliação	36
3	Revisão Bibliográfica	37
3.1	Revisão da literatura de previsão de demanda em operações humanitárias	37
3.2	Análise bibliométrica de modelos de regressão de dados de contagem com estimações inflacionadas de zeros	42
4	Metodologia	45
4.1	Materiais	45

4.2	Método	52
5	Resultados	55
5.1	Resultados e Discussão	55
6	Conclusões	63
6.1	Conclusões	63
6.2	Limitações e Trabalhos Futuros	64
	Referências	65
	Apêndice A – Informações da demanda e dos fatores preditores	73
	Anexo A – Modelo matemático de otimização do problema de localização com múltiplos critérios	79
A.1	Formulação do modelo matemática	81
A.2	Determinação dos parâmetros	83
	Anexo B – Sistema de Apoio à Decisão	85
B.1	Conceito	85
B.2	Desenvolvimento	85
B.3	Estrutura	86

1 INTRODUÇÃO

Os desastres estão perdendo seu caráter de fenômeno extraordinário. Nos últimos 20 anos, o número de desastres dobrou sendo 90% deles relacionados com mudanças climáticas. A frequência de fenômenos naturais como furacões, secas e inundações têm uma tendência de crescimento e, conseqüentemente, um aumento na quantidade de pessoas afetadas por tais eventos (ONU, 2019).

A vulnerabilidade das sociedades atuais tem demandado maiores esforços dos Estados e de organizações de ajuda humanitária. Os processos e sistemas envolvidos na mobilização das pessoas, recursos, habilidades e conhecimento para ajudar indivíduos vulneráveis afetados por um desastre são definidas como operações humanitárias (WASSENHOVE, 2006).

Proporcionar suprimentos de socorro após a ocorrência de um desastre é uma atividade de vital importância para o auxílio das pessoas atingidas. A agilidade na distribuição é condicionada pela disponibilidade desses itens no lugar do desastre. Assim, a localização de instalações para armazenagem de suprimentos de socorro em uma fase de preparação para um eventual desastre desempenha um papel fundamental nas operações humanitárias de resposta. No que se refere ao problema de localização de instalações para pré-posicionamento de suprimentos de alívio, a demanda é um parâmetro relevante a ser considerado, dado que influencia diretamente nos custos e no nível de atendimento (HOLGUÍN-VERAS; JALLER, 2012; LAAN et al., 2016).

A previsão da demanda desempenha um importante papel nas fases de mitigação, preparação, resposta e recuperação da gestão de desastres. A mitigação e a preparação exigem um planejamento cuidadoso e orçamentário de aquisições, onde a previsão desempenha um papel integral. Durante a fase de resposta, a execução do plano pode ser aprimorado com estimativas de demanda em “tempo real”. A fase de recuperação, geralmente apresenta restrições orçamentárias, onde as previsões voltarão a desempenhar um papel essencial no processo orçamentário. Claramente, a previsão deve ser considerada nesse planejamento e preparação.

A escassez de materiais ou uma gestão ineficiente de recursos pode comprometer a resposta de emergência. A previsão da demanda tem um papel crucial nas operações humanitárias e na redução de perdas. Segundo Wassenhove e Martinez (2012), a previsão da demanda pode gerar mais de 7% nos custos operacionais anuais. No entanto, a previsão da demanda no contexto humanitário é ainda uma área sub explorada, pois não existem métodos de previsão abrangentes ou abordagens autorizadas de inferência de recursos de emergência após um evento de desastre (ZHU; ZHANG; SUN, 2019).

O abastecimento de ajuda humanitária aos 645 municípios do Estado de São Paulo é uma responsabilidade da Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC-SP). A maior parte da demanda por suprimentos de socorro é gerada por inundações e deslizamentos de terra, desastres comuns no Estado. Um registro foi elaborado com base nas solicitações por materiais de socorro atendidas pela CEPDEC-SP nos anos compreendidos entre 2015 e 2020. Tal relatório contém informações sobre as requisições de 161 municípios; os outros 484 não aparecem no registro, infere-se que não houve necessidade de suprimentos de socorro por parte desses municípios no espaço de tempo analisado.

A previsão de demanda de materiais de socorro é uma questão complexa. Estratégias devem basear-se na compreensão das instâncias derivadas da vulnerabilidade, tanto natural quanto humana. Obter informações explicativas do fenômeno e das necessidades em eventos de desastres é um dos principais desafios. Diferentes adversidades que dificultam a coleta de dados de campo relevantes acontecem nesses cenários.

Neste sentido, o presente estudo pretende desenvolver um modelo de previsão de demanda de suprimentos de socorro que seja capaz de lidar com a complexidade das variáveis preditoras, a informação limitada e a incerteza própria dos desastres.

1.1 Contexto e Motivação

O Estado de São Paulo não é suscetível a desastres de grande magnitude. Contudo, aqueles de menor escala são frequentes e estão associados predominantemente a deslizamentos de encostas, inundações, erosão acelerada e tempestades. No período de 2000 a 2008 houve um registro de 1.861 acidentes, sendo: em torno de 50% (944) de inundações (incluindo enchentes e alagamentos), 19% (367) de deslizamentos, 4% (65) de raios e 27% (485) de acidentes diversos (chuvas fortes, vendavais, desabamentos de casas e muros, etc.) (BROLLO; FERREIRA, 2009).

No Estado de São Paulo, algumas regiões são mais vulneráveis a desastres naturais do que outras; especialmente nos grandes centros urbanos, que apresentam extremas desigualdades sociais, com áreas de alto padrão de qualidade de vida e outras de extrema miséria (SEADE, 2010).

O Sistema Estadual de Defesa Civil foi criado pelo Decreto n° 7.550, de 09 de fevereiro de 1976 e reorganizado em 16 de junho de 1995 pelo então Governador Mario Covas, mediante o Decreto n° 40.151, com uma Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC-SP), subordinada diretamente ao Governador do Estado e dirigida pelo Coordenador Estadual de Proteção e Defesa Civil, constituindo-se o órgão central do Sistema de Proteção e Defesa Civil do Estado de São Paulo (CEPDEC, 1995).

A CEPDEC-SP é uma organização voltada à proteção civil, que por meio de ações de prevenção, mitigação, preparação, resposta e recuperação pretende evitar a perda de vidas, reduzir os prejuízos, e diminuir o sofrimento das populações assoladas por catástrofes no Estado de São Paulo. Além disso, a instituição também é responsável pela cadeia de abastecimento de ajuda humanitária a municípios em situações de anormalidade, geralmente declaradas após desastres naturais. Assim, um dos principais desafios da CEPDEC-SP é fornecer suprimentos de alívio, de forma ágil e eficiente, para socorrer pessoas atingidas por desastres. A previsão da demanda de materiais de alívio e a localização de depósitos para seu armazenamento são fatores determinantes para o sucesso deste desafio.

A CEPDEC-SP dispõe de informações referentes à demanda de suprimentos de socorro. Registros da quantidade e tipo de materiais foram levantados durante um período de 5 anos. Os relatórios possuem registros detalhados de todas as solicitações dos municípios do Estado de São Paulo; porém, muitos deles não foram listados nenhuma vez no espaço de tempo analisado. Inferindo a demanda nula para esses municípios, a variável de resposta é composta por uma quantidade excessiva de zeros. Isso, junto à complexidade dos fatores preditores, dificultam a modelagem preditiva da demanda de materiais de socorro.

Alguns trabalhos foram desenvolvidos com diferentes abordagens para auxiliar nas decisões de localização de depósitos de suprimentos de alívio da CEPDEC-SP. Esses trabalhos são apresentados a seguir.

Brito (2015) formulou um modelo estocástico para a localização de depósitos de suprimentos de alívio aplicado na região do Vale do Paraíba, no Sudeste do Estado de São Paulo. O modelo visa minimizar os custos, considerando restrições de capacidade, mate-

riais disponíveis e nível mínimo de atendimento. Os dados de demanda foram obtidos de mapeamentos de risco e dados de população.

Consenza (2015) por sua vez, apresenta um modelo para localização de depósitos similar ao proposto por Brito (2015), mas estendível para todos os municípios do Estado de São Paulo. Consenza (2015) propôs um modelo de previsão do número de pessoas vivendo em condições de risco considerando as seguintes variáveis explicativas: número de intervenções da CEPDEC-SP, população, densidade demográfica e IDH. Ajustes com distribuição de Poisson e distribuição Binomial negativa foram testados visualmente nos dados, comparando a função de distribuição cumulativa empírica com o teórico. A partir dos gráficos, Consenza (2015) concluiu que a distribuição binomial negativa tem o melhor ajuste para a distribuição empírica. O método que conduz ao modelo final consistiu em um processo iterativo de desenvolvimento de modelos de previsão e avaliações comparativas. Finalmente o modelo Zero-Inflated Negative Binomial (ZINB) foi selecionado para estimar o número de pessoas que vivem em condições de risco; isso e informações das necessidades por pessoa compiladas de outras fontes, foram a base para calcular uma aproximação da demanda total de suprimentos de socorro do Estado de São Paulo.

O foco do trabalho desenvolvido por Okane (2019) foi gerar uma ferramenta visual interativa para dar suporte nas decisões relacionadas à localização de depósitos de suprimentos de socorro da CEPDEC-SP. O modelo de localização proposto é muito mais simples do que o processo multi-objetivo de dois estágios definido por Consenza (2015). Neste caso, os locais para os depósitos são parâmetros fornecidos pelo usuário, então o objetivo é atribuir a cada município o armazém mais próximo. A simplicidade do modelo de localização foi importante nesta etapa devido à necessidade de criar interações em tempo real com os usuários durante o cálculo das respostas do modelo. Para estimar a demanda foi aplicado o modelo de previsão estabelecido anteriormente por Consenza (2015), com acréscimo da variável preditora que contempla o nível de chuvas. O Sistema de Apoio à Decisão (SAD) desenvolvido por Okane (2019) surgiu como uma solução de independência e interatividade para a CEPDEC-SP. A visualização dos resultados também foi outro ganho do sistema implementado.

O presente trabalho pretende dar continuidade às pesquisas passadas e aproveitar algumas oportunidades de melhoria relacionadas à previsão da demanda; tais como, partir de uma abordagem focada nas necessidades de suprimentos de socorro, considerar variáveis explicativas relacionadas a vulnerabilidade natural e humana, e incluir uma estrutura hierárquica na modelagem.

Por outro lado, o SAD proposto por Okane (2019) para a CEPDEC-SP localizar seus depósitos de suprimentos de socorro foi mantido nesta dissertação, sofisticando o modelo de otimização e aprimorando a interface visual interativa para facilitar a personalização de cenários com diferentes locais de instalação de depósitos. Oferecendo ao tomador de decisão autonomia para gerar simulações e rapidamente verificar os resultados em relação ao custo e nível de serviço.

Mais detalhes do modelo de otimização e o SAD são apresentados no Anexo A e B, respectivamente.

1.2 Objetivos

O objetivo principal do projeto é desenvolver um modelo de previsão da demanda de suprimentos de socorro a fim de obter estimativas anuais das necessidades dos 645 municípios que constituem o Estado de São Paulo.

A pesquisa é desenvolvida com dados proporcionados pela CEPDEC-SP somadas a outras informações geográficas e estatísticas do Brasil obtidos de fontes oficiais. Embora o modelo seja aplicado a um evento e local específico, deseja-se proporcionar uma abordagem para ser aplicada em situações semelhantes.

Para atingir o objetivo principal, o projeto tem os seguintes objetivos específicos:

- Identificar fatores preditores da demanda de suprimentos de socorro;
- Desenvolver o modelo de previsão;
- Avaliar o desempenho do modelo de previsão;

1.2.1 Perguntas da pesquisa

A pergunta principal da pesquisa é:

- Qual é a demanda anual estimada no que se refere aos suprimentos de socorro por município do Estado de São Paulo?

A partir da pergunta principal de pesquisa alguns questionamentos secundários são gerados:

- Quais fatores preditores influenciam na demanda?

- Como avaliar o desempenho do modelo de previsão?

A dissertação está organizada em três fases, a saber: planejamento, execução e apresentação dos resultados. Na fase de planejamento, definem-se o problema, os objetivos e as questões da pesquisa (Capítulo 1). A fase de execução apresenta, por sua vez, a contextualização teórica e revisão da literatura sobre tópicos relevantes ao tema de pesquisa (Capítulo 2 e 3), a metodologia utilizada no desenvolvimento do estudo (Capítulo 4). Por fim, será apresentada a fase de validação e divulgação dos resultados, conclusões, e trabalhos futuros (Capítulo 5 e 6).

2 CONTEXTUALIZAÇÃO TEÓRICA

Nesse capítulo são apresentados conceitos e definições sobre desastres e operações humanitárias além de modelos de previsão de demanda. O objetivo principal desta revisão da literatura é fornecer referências metodológicas, gerar fundamentos para direcionar a dissertação, além de garantir sua relevância acadêmica.

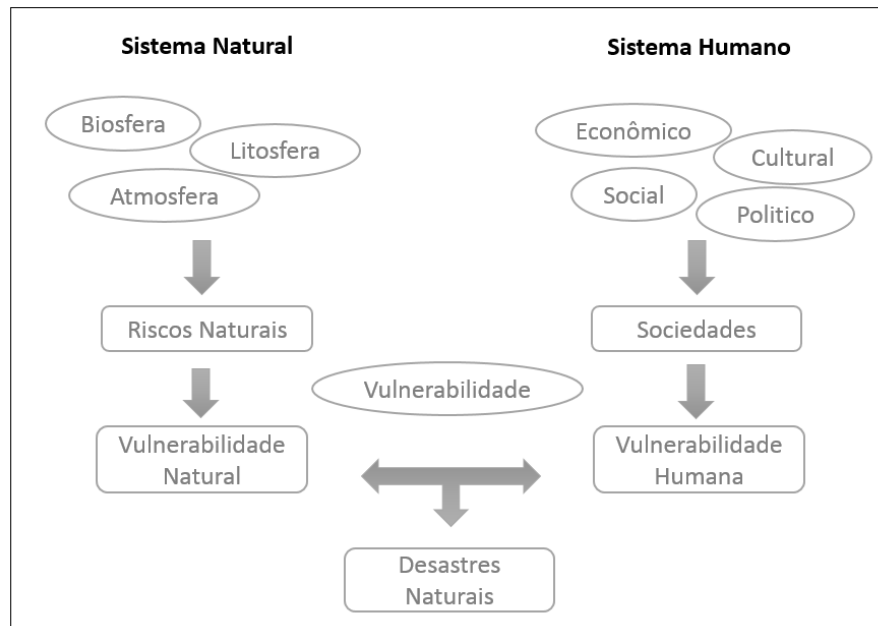
2.1 Desastres e Logística Humanitária

De acordo com IFRC (2021b), desastre é um evento repentino e calamitoso que perturba gravemente o funcionamento de uma comunidade ou sociedade e causa perdas humanas, materiais e econômicas ou ambientais que excedem a capacidade da comunidade ou sociedade de lidar com seus próprios recursos. Tais danos são produzidos não apenas no momento da sua ocorrência, mas a longo prazo (ALCÁNTARA-AYALA, 2002).

Os desastres são classificados quanto sua natureza de origem ou causa primária, podendo ser naturais ou humanos (antropogênicos) (IFRC, 2021b). Considera-se desastre natural todo aquele causado por um fenômeno ou desequilíbrio da natureza, agravado ou não pela atividade das pessoas, a saber: chuvas intensas provocando inundação, erosão e deslizamentos; ventos fortes formando vendaval, tornado, furacão entre outros. Desastres de natureza humana são, por sua vez, aqueles resultantes de ações ou omissões humanas e estão diretamente relacionados com atividades do homem, seja como agente ou autor. Podemos citar, por exemplo, acidentes de trânsito, incêndios urbanos, contaminação de rios, rompimento de barragens, etc. (SOUZA et al., 2009).

Westgate (1976) destaca-se entre os primeiros em reconhecer a importância da vulnerabilidade natural e humana, definindo desastre como a interação entre fenômenos físicos ou naturais extremos e um grupo humano vulnerável. Segundo Alcántara-Ayala (2002), a vulnerabilidade é dada pela combinação dos sistemas natural e humano; fatores que influem nos desastres são apresentados na Figura 1.

Figura 1: Fatores que influem nos desastres naturais



Fonte: Alcántara-Ayala (2002)

Ainda que desastres naturais ocorrem em todo o mundo, seu impacto nos países em desenvolvimento é maior devido à localização geográfica em zonas altamente suscetíveis a desastres naturais (vulnerabilidade natural), e também devido aos diferentes tipos de vulnerabilidades econômicas, sociais, políticas e culturais existentes (vulnerabilidade humana) (ALCÁNTARA-AYALA, 2002).

Além da classificação quanto a sua natureza de origem, Wassenhove (2006) classifica os desastres conforme o modo de início. Desastres que apresentam início súbito são aqueles que tem um impacto abrupto e repentino na sociedade, exigindo socorro rápido e coordenado. Nos de início lento, por sua vez, organizações e governos são capazes de adotar medidas de mitigação para o momento em que o evento ocorre. A Tabela 1 apresenta exemplos de desastres de acordo com o modo de início e a origem.

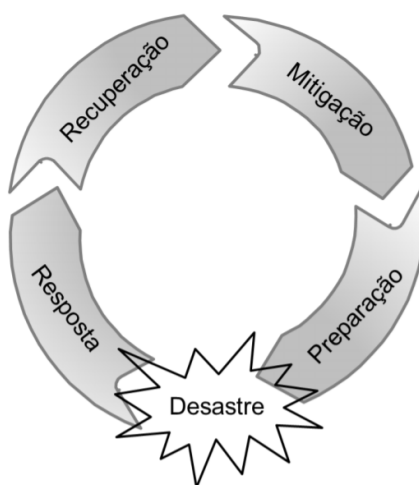
Tabela 1: Exemplos de desastres por modo de início

	Natural	Antropogênico
Início súbito	Terremoto	Ataque terrorista
	Furacão	Golpe de Estado
	Tornado	Acidente químico
Início lento	Seca	Crise política
	Degradação ambiental	Crise de refugiados

Fonte: Wassenhove (2006)

A gestão de desastres pode ser definida como a organização e gestão de recursos e responsabilidades para lidar com todos os aspectos humanitários de emergências (IFRC, 2021a). O ciclo de vida dos desastres é dividido em quatro fases: mitigação, preparação, resposta e recuperação (FEMA, 2019). A Figura 2 ilustra a relação das quatro fases da gestão de emergência. A Mitigação consiste na aplicação de medidas para evitar a ocorrência do desastre ou minimizar os danos no caso de sua ocorrência. A preparação está relacionada com os planos para a ação apropriada em caso de desastre. A resposta, por sua vez, envolve o uso de recursos e procedimentos de emergência realizados de acordo com o planejamento da fase de preparação, enquanto a recuperação refere-se a atividades que abrangem efeitos a longo prazo (PÉREZ-GALARCE et al., 2017).

Figura 2: Ciclo de vida dos desastres



Fonte: FEMA (2019)

Na fase de preparação a desastres, as informações de estimativas da demanda são chave para definir estratégias ótimas de pré-posicionamento de materiais e armazenagem

de suprimentos de socorro e deste modo, garantir uma intervenção ágil e eficiente pós-desastre. Contudo, a previsão da demanda é uma questão complexa devido à diversidade de fatores particulares da entidade ameaçada. Enquanto fatores endógenos podem complicar a coleta de dados de campo relevantes e o uso de técnicas de previsão, fatores exógenos, por sua vez, podem influenciar negativamente à precisão da previsão.

2.2 Modelos de previsão de demanda

Análises de dados consistem em abordagens descritivas, análises preditivas e prescritivas. A análise descritiva é definida como a descrição e categorização do que aconteceu no passado, enquanto a preditiva é usada para prever eventos futuros e descobrir padrões preditivos nos dados usando algoritmos matemáticos. A análise prescritiva aplica dados e algoritmos matemáticos para a tomada de decisões (SEYEDAN; MAFAKHERI, 2020). A análise preditiva é a mais utilizada na demanda da cadeia de suprimento e na previsão de compras (WANG et al., 2016). Assim, nas seções a seguir, o foco da abordagem será orientado a análise preditiva.

Para definir o modelo a ser usado, é importante identificar previamente o tipo de dado e compreender seu comportamento. No contexto da demanda de suprimentos de socorro, modelos de regressão para dados de contagem são os mais apropriados. Posto que, para explicar uma variável dependente que assume um limitado conjunto de valores estritamente positivos, os modelos de regressão por mínimos quadrados sofrem várias limitações e propriedades estatísticas (MIAOU; LUM, 1993; SHANKAR; MANNERING; BARFIELD, 1995; YAACOB; LAZIM; WAH, 2010).

2.2.1 Generalized Linear Models (GLM)

A classe de modelos conhecidos como Generalized Linear Models (GLM) junto à teoria de quase-verossimilhança, foram introduzidos formalmente por Nelder e Wedderburn (1972), que consolidaram nessa categoria vários modelos estatísticos, como: modelos logísticos envolvendo Bernoulli e distribuições binomiais, modelos de dados de contagem envolvendo distribuição de Poisson, e modelos polinomiais envolvendo distribuição Gama, entre outros.

Componente aleatório, componente sistemático e função de ligação são os três principais componentes de qualquer GLM (AGRESTI, 2015; FOX, 2015):

- Componente aleatório, que especifica a distribuição de probabilidade da variável de resposta Y .
- Componente sistemático, que especifica as variáveis explicativas do modelo, relacionadas em um único *preditor linear*: $\eta = \sum_{j=1}^p x_j \beta_j$.
- Função de ligação canônica representada por η ou $g(\mu)$, que especifica a ligação entre os componentes aleatórios e sistemáticos.

Os GLM oferecem considerável flexibilidade na modelagem; através da função de ligação, permitem especificar uma conexão linear entre a variável de resposta e os preditores, mesmo que sua relação subjacente não seja linear (VENABLES; DICHMONT, 2004).

Cada caso particular dos GLM envolve as características da variável dependente, sua distribuição e a função de ligação canônica, assim como a função de verossimilhança associada. A estimação por máxima verossimilhança objetiva, através de ferramentas de programação linear, determinar os valores dos parâmetros que maximizam o valor do logaritmo da função de verossimilhança (FÁVERO; BELFIORE, 2017).

O método de máxima verossimilhança fornece além das estimativas dos coeficientes de regressão, seus respectivos erros padrão. A Equação (2.1) apresenta o desvio residual para os GLM, na forma geral:

$$D_m = 2(\log L_s - \log L_m), \quad (2.1)$$

Onde L_m é a verossimilhança maximizada sob o modelo proposto e L_s é a verossimilhança maximizada sob um modelo saturado, que dedica um parâmetro para cada observação e, conseqüentemente, ajusta os dados o mais próximo possível. O desvio residual é uma generalização da soma dos quadrados dos resíduos em modelos lineares (FOX, 2015).

2.2.2 Modelos de regressão para dados de contagem

Como parte dos GLM, os modelos de regressão para dados de contagem são utilizados no contexto de modelagem de variáveis dependentes de natureza inteira, discreta e não negativa. Wooldridge (2010) descreve de forma geral os modelos de dados de contagem na Equação 2.2.

$$\widehat{Y}_i = \lambda_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}), \quad (2.2)$$

Onde λ_i representa o número esperado de ocorrências do fenômeno em estudo para uma determinada exposição; β_0 é o intercepto; $\beta_j (j = 1, 2, \dots, k)$ são os coeficientes estimados para cada variável preditora x_i ; k é o número de variáveis preditoras e $i (i = 1, 2, \dots, n)$ representa cada observação em uma amostra de tamanho n .

2.2.2.1 Modelo de regressão de Poisson

O modelo de regressão de Poisson é o mais utilizado para a análise de dados de contagem, dada sua simplicidade é um bom ponto de partida (CAMERON; TRIVEDI, 1990). A variável resposta que segue uma distribuição Poisson é assumida como independente com probabilidade de ocorrência do valor de contagem m considerando-se determinada exposição (período de tempo, área, região, e outros) é apresentada na Equação 2.3.

$$p(Y_i = m) = \frac{e^{-\lambda_i} \lambda_i^m}{m!}, \quad m = 0, 1, 2, \dots \quad (2.3)$$

Onde, λ representa o número esperado de ocorrências do fenômeno em estudo para determinada exposição; λ é o único parâmetro livre da distribuição Poisson, não permitindo que a variância seja ajustada à média. Portanto, sua aplicação é restrita para dados de contagem que assumem a característica de equidispersão, ou seja, igualdade de variância e média, $\mu_i = E(Y_i) = Var(Y_i) = \lambda_i$. Se esta suposição não for válida, o estimador obtido da máxima verossimilhança de Poisson será incorreto (CAMERON; TRIVEDI, 2013).

Os coeficientes da regressão Poisson β_j podem ser estimados através do método de máxima verossimilhança. A função de verossimilhança é dada pela Equação 2.4, a partir da qual pode ser deduzido o logaritmo da função de verossimilhança na Equação 2.5.

$$L = \prod_{i=1}^n \frac{e^{-\lambda_i} (\lambda_i)^{Y_i}}{Y_i!} \quad (2.4)$$

$$LL = \sum_{i=1}^n [-\lambda_i + (Y_i) \ln(\lambda_i) - \ln(Y_i!)] \quad (2.5)$$

2.2.2.2 Modelo de regressão Binomial Negativo (BN)

O modelo Binomial Negativo (BN) está diretamente ligado à presença de superdispersão no conjunto de dados de contagem. A superdispersão ocorre quando a variância dos dados de contagem é estatisticamente maior que a média, $\mu_i = E(Y_i) < Var(Y_i) = \lambda_i$. As expressões 2.6 e 2.7 representam a média e a variância, respectivamente, para um modelo binomial negativo.

$$E(Y_i) = \mu_i \quad (2.6)$$

$$Var(Y_i) = \mu_i + \phi\mu_i^2 \quad (2.7)$$

Onde o parâmetro ϕ representa a superdispersão dos dados de contagem. Cameron e Trivedi (1990) estabeleceram um teste para detectar superdispersão de dados de contagem através da Expressão 2.8, onde a hipótese nula H_0 é a equidispersão e o fenômeno de superdispersão é associado à hipótese alternativa H_i que ocorreria para o caso em que ϕ é estatisticamente maior que zero. Para a detecção de superdispersão nos dados de contagem num certo nível de significância, um modelo de Poisson deve ser estimado a priori (CAMERON; TRIVEDI, 1990).

$$Var(Y|X) = E(Y|X) + \phi[E(Y|X)]^2 \quad (2.8)$$

O número esperado de ocorrências de uma distribuição BN é apresentado na Equação 2.9, em que o primeiro termo μ_i representa o valor esperado de ocorrências e possui uma distribuição Poisson. O segundo termo ν_i corresponde à parcela aleatória do número de ocorrências da variável dependente e possui uma distribuição Gama.

$$\lambda_i = \mu_i \cdot \nu_i \quad (2.9)$$

A função de distribuição de probabilidade da variável ν_i para determinada observação i é dada pela expressão 2.10.

$$p(\nu_i) = \frac{\delta^\psi \cdot \nu_i^{\psi-1} \cdot e^{-\nu_i \psi}}{\Gamma(\psi)} \quad (2.10)$$

Em que ψ é o parâmetro de forma, δ o parâmetro de taxa e para ψ positivo e inteiro,

$\Gamma(\psi)$ pode ser aproximado por $(\psi - 1)!$.

Aplicando a expressão 2.9 é possível transformar a função de probabilidade da distribuição Gama apresentada na Equação 2.10, como uma função do valor esperado de ocorrências da distribuição Poisson μ_i da Equação 2.11.

$$p(\mu_i) = \frac{\left(\frac{\psi}{\mu_i}\right)^\psi \cdot \lambda_i^{\psi-1} \cdot e^{-\frac{\lambda_i}{\mu_i} \delta}}{\Gamma(\psi)} \quad (2.11)$$

Logo, combinando as expressões 2.3 e 2.11 pode ser deduzida a Equação 2.12 da função de probabilidade de uma distribuição BN para calcular a probabilidade de ocorrência de uma contagem m , dada determinada exposição.

$$p(Y_i = m) = \int_0^\infty \frac{e^{(-\lambda_i)} \cdot \lambda_i^m}{m!} \cdot \frac{\left(\frac{\psi}{\mu_i}\right)^\psi \cdot \lambda_i^{\psi-1} \cdot e^{-\frac{\lambda_i}{\mu_i} \delta}}{\Gamma(\psi)} d\lambda_i \quad (2.12)$$

Conforme discutido anteriormente, a estimação dos parâmetros do modelo BN ϕ, α, β_j pode ser efetuada pelo método de máxima verossimilhança, com base na Equação 2.13 do logaritmo da função de verossimilhança de um modelo de regressão BN, que deve ser iterado até atingir seu valor máximo.

$$LL = \sum_{i=1}^n \left[Y_i \cdot \ln \left(\frac{\phi \cdot \mu_i}{1 + \phi \cdot \mu_i} \right) - \frac{\ln(1 + \phi \cdot \mu_i)}{\phi} + \ln \Gamma(Y_i + \phi^{-1}) - \ln \Gamma(Y_i + 1) - \ln \Gamma(\phi^{-1}) \right] \quad (2.13)$$

2.3 Modelo de regressão Zero-Inflated (ZI)

Modelos de contagem comumente apresentam uma quantidade excessiva de zeros que pode enviesar os modelos de regressão tradicionais Poisson e BN. Nesses casos, são adequados os modelos de regressão Zero-Inflated (ZI). A definição da existência ou não de uma quantidade excessiva de zeros na variável resultado é determinada através da análise das saídas do teste de Vuong (VUONG, 1989). Desmarais e Harden (2013) propõem uma correção ao teste de Vuong, que se baseia nas estatísticas Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC), uma vez que os resultados do teste mencionado podem ser tendenciosos para apoiar modelos com inflação de zero.

De acordo com Lambert (1992), os modelos de regressão ZI adotam uma abordagem

que combina um modelo para dados binários com um modelo para dados de contagem. A distribuição binária indica se as contagens são zero ou não, chamados de zeros estruturais. Por outro lado, a distribuição tradicional Poisson ou BN determina a quantidade de ocorrências de um fenômeno, incluindo contagens zero que ocorreram coincidentemente, chamados de zeros amostrais. Portanto, um modelo ZI assume que a probabilidade geral de zero contagens é a probabilidade combinada de zeros de amostragem e estruturais.

Ngatchou-Wandji e Paris (2011) define os modelos ZI com a seguinte Equação 2.14.

$$p(Y = y|\omega) = \omega\delta_0(y) + (1 - \omega)f(y) \quad (2.14)$$

Onde, Y é a variável de contagem, ω é o proporção de excesso de zeros, $\delta_0(y) = 1$ se $y = 0$, $\delta_0(y) = 0$, para outros casos; e $f(y)$ é a densidade de uma distribuição de contagem, como Poisson ou BN. A escolha da distribuição dependerá da existência de superdispersão nos dados (PERUMEAN-CHANEY et al., 2013).

2.3.1 Modelo de regressão Zero-Inflated Poisson (ZIP)

Nos modelos de regressão Zero-Inflated Poisson (ZIP), a probabilidade de ocorrência de uma contagem zero $p(Y_i = 0)$ para dada observação $i (i = 1, 2, \dots, n)$ com tamanho de mostra n , pode ser calculada levando-se em consideração a soma de um componente dicotômico com um componente de contagem. O componente dicotômico pode ser definido com a probabilidade p_{logit_i} de ocorrência de uma contagem zero. Por outro lado, a probabilidade de ocorrência de uma contagem específica $m (m = 1, 2, \dots)$, ou seja, $p(Y_i = m)$, segue a expressão da distribuição de probabilidade de Poisson, multiplicada por $(1 - p_{logit_i})$ (FÁVERO; BELFIORE, 2017). Desta forma é deduzida a Expressão 2.15.

$$\begin{cases} p(Y_i = 0) = p_{logit_i} + (1 - p_{logit_i}) \cdot e^{-\lambda_i} \\ p(Y_i = m) = (1 - p_{logit_i}) \cdot \frac{e^{-\lambda_i} \cdot \lambda_i^m}{m!}, \quad m = 1, 2, \dots \end{cases} \quad (2.15)$$

Sendo $Y \sim ZIP(\lambda, p_{logit_i})$, onde λ é o número esperado de ocorrências da variável de resposta dada uma exposição. Se $p_{logit_i} = 0$ a distribuição de probabilidade da Equação 2.15 é simplificada à distribuição Poisson, inclusive para casos em que $Y_i = 0$. Por tanto, os modelos de regressão ZIP, consideram dois processos geradores de zeros, sendo um devido à distribuição binária (os chamados zeros estruturais) e outro devido à distribuição de Poisson (os chamados zeros amostrais).

Os parâmetros λ e p_{logit_i} de determinado modelo de regressão ZIP podem ser estimados por meio do método de máxima verossimilhança. De 2.15, a função de logaritmo de verossimilhança é dada pela Equação 2.16.

$$LL = \sum_{Y_i=0} \ln [p_{logit_i} + (1 - p_{logit_i}) \cdot e^{-\lambda}] + \sum_{Y_i>0} [\ln(1 - p_{logit_i}) - \lambda + (Y_i) \cdot \ln(\lambda) - \ln(Y_i!)] \quad (2.16)$$

2.3.2 Modelo de regressão Zero-Inflated Negative Binomial (ZINB)

Nos modelos de regressão ZINB, a probabilidade de ocorrência de uma contagem zero $p(Y_i = 0)$ é também calculada levando-se em consideração a soma de um componente dicotômico com um componente de contagem e a probabilidade de ocorrência de uma determinada contagem $m(m = 1, 2, \dots)$, ou seja, $p(Y_i = m)$, segue agora a expressão da distribuição de probabilidade de Poisson-Gamma, representada na Equação 2.17.

$$\begin{cases} p(Y_i = 0) = p_{logit_i} + (1 - p_{logit_i}) \cdot \left(\frac{1}{1 + \phi\mu_i}\right)^{\frac{1}{\phi}} \\ p(Y_i = m) = (1 - p_{logit_i}) \cdot \left[\binom{m + \phi^{-1} - 1}{\phi^{-1} - 1} \cdot \left(\frac{1}{1 + \phi\mu_i}\right) \cdot \left(\frac{\phi\mu_i}{\phi\mu_i + 1}\right)^m \right], \quad m = 1, 2, \dots \end{cases} \quad (2.17)$$

Sendo $Y \sim ZINB(\phi, \lambda, p_{logit_i})$, onde ϕ representa o inverso do parâmetro de forma de determinada distribuição Gamma. De forma análoga aos modelos de regressão ZIP, os modelos ZINB apresentam dois processos geradores de zeros, derivados da distribuição binária e da distribuição Poisson-Gamma. Portanto, um modelo ZINB assume que as observações zero se originam de zeros estruturais e de amostragem (CAMERON; TRIVEDI et al., 2010).

A Equação 2.18 define a função de probabilidade logarítmica para estimar os parâmetros de um modelo de regressão ZINB através de métodos de máxima verossimilhança.

$$\begin{aligned}
LL = & \sum_{Y_i=0} \ln \left[p_{\text{logit}_i} + (1 - p_{\text{logit}_i}) \cdot \left(\frac{1}{1 + \phi\lambda_i} \right)^{\frac{1}{\phi}} \right] \\
& + \sum_{Y_i>0} \left[\ln(1 - p_{\text{logit}_i}) + Y_i \ln \left(\frac{\phi\lambda_i}{\phi\lambda_i + 1} \right) - \frac{\ln(1 + \phi\lambda_i)}{\phi} + \ln\Gamma(Y_i + 1) - \ln\Gamma\phi^{-1} \right]
\end{aligned} \tag{2.18}$$

2.4 Generalized Linear Mixed Models (GLMM)

Os modelos de regressão para dados de contagem com efeitos mistos podem ser utilizados quando a variável de resposta é quantitativa, porém com valores discretos e não negativos, quando dispostos em determinada estrutura aninhada em níveis, podendo haver dados agrupados ou com medidas repetidas (FÁVERO; BELFIORE, 2017).

A modelagem dos zeros estruturais e amostrais proposto por Lambert (1992) foi ampliado por Hall (2000) para o campo de estimativas GLMM, adicionando à modelagem do componente de contagem os agrupamentos hierárquicos no conjunto de dados. A designação mista no termo GLMM vem do fato de que as variáveis preditoras podem ser consideradas nos componentes de efeitos fixos e aleatórios do modelo de regressão. Os parâmetros estimados de efeitos fixos indicam a relação entre as variáveis preditoras e a variável de resposta, da mesma forma de estimativas GLM; enquanto o componente de efeitos aleatórios pode ser representado pela combinação de termos de erro e variáveis preditoras. As estimativas GLMM levam em consideração a existência de dependência entre observações de um mesmo grupo e, conseqüentemente, geram erros padrão sem vieses.

Segundo Fávero (2017), existem muitas situações em que os dados são dispostos dentro de uma estrutura mista ou aninhada, e as hierarquias referem-se ao fato de que observações dos mesmos grupos, compartilham características que representam uma espécie de homogeneidade. Nesta perspectiva, pode ser especificado o modelo misto de dados de contagem ZI de dois níveis apresentado na Equação 2.19 e 2.20. Onde o primeiro nível refere-se a observações $i(i = 1, 2, \dots, n)$ aninhadas em unidades de dois níveis $j(j = 1, 2, \dots, J)$.

$$\ln \left(\frac{p_{\text{logit}_i}}{(1 - p_{\text{logit}_i})} \right) = \mathbf{Z}_{ij}^T \boldsymbol{\delta} \quad \text{sendo} \quad \boldsymbol{\delta} = \mathbf{G}_j^T \boldsymbol{\alpha} \tag{2.19}$$

$$\ln(\lambda_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\pi} + \mathbf{v}_j \quad \text{sendo} \quad \boldsymbol{\pi} = \mathbf{H}_j^T \boldsymbol{\beta} \tag{2.20}$$

Onde as matrizes das variáveis preditoras \mathbf{Z} e \mathbf{X} , de níveis i (nível 1) e j (nível 2), que aparecem nos respectivos componentes logísticos (Bernoulli) e Poisson ou Poisson-Gamma não são necessariamente as mesmas (condições iguais valem para matrizes das variáveis (\mathbf{G}_j e \mathbf{H}_j no nível 2). $\boldsymbol{\delta}$, $\boldsymbol{\pi}$, $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ são as respectivas matrizes dos parâmetros de regressão, ou seja, $\boldsymbol{\delta}$ pode ser interpretado em termos da proporção de inflação de zeros, $\boldsymbol{\pi}$ está relacionado com a resposta média na parte dos dados de contagem, e $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ correspondem às diferenças entre o contexto do nível dois em zeros estruturais e amostrais, respectivamente, devido ao comportamento das variáveis preditoras no nível 2 (\mathbf{G}_j e \mathbf{H}_j). $\boldsymbol{\delta}_j$ representa as variações aleatórias no segundo nível, o que significa que a heterogeneidade entre níveis superiores de análise (grupos, por exemplo) e entre indivíduos é permitida através dos efeitos aleatórios $\boldsymbol{\delta}$, com variância igual a $\boldsymbol{\sigma}^2\boldsymbol{v}_j$ (LEE et al., 2006; FÁVERO, 2017).

2.5 Métricas de avaliação

O ajuste dos modelos GLM pode ser avaliado através de AIC, BIC e de várias quantidades baseadas em desvio (nulo, residual e explicado).

Em estatística, o AIC é mais frequentemente usado para seleção de modelos. Ao calcular e comparar as pontuações do AIC de vários modelos possíveis, pode-se identificar aquele que se ajusta melhor aos dados. O AIC penaliza os modelos que usam mais parâmetros. Portanto, se dois modelos explicarem a mesma quantidade de variação, aquele que apresentar menos parâmetros terá uma pontuação de AIC menor e será o modelo de melhor ajuste (AKAIKE, 1974).

A Equação 2.21 determina o valor de AIC do modelo usando a estimativa de máxima verossimilhança e o número de parâmetros (variáveis independentes) no modelo.

$$AIC = 2k - 2\ln(L) \quad (2.21)$$

Onde k é o número de variáveis independentes do modelo e L corresponde à estimativa de probabilidade logarítmica.

Considerando os dados de demanda deste estudo, que assumem um limitado conjunto de valores positivos, uma perspectiva adequada apresentam os modelos para dados de contagem. Por outro lado, a existência excessiva de zeros e a variabilidade evidenciada entre municípios, sugerem uma abordagem ZI com estrutura de modelo multinível.

3 REVISÃO BIBLIOGRÁFICA

Duas revisões são apresentadas nesta seção. A primeira revisão exhibe o estado da arte da previsão de demanda no contexto humanitário. A segunda é relacionada ao método de previsão adotado no estudo.

3.1 Revisão da literatura de previsão de demanda em operações humanitárias

Nesta seção são apresentados trabalhos existentes sobre a previsão da demanda de suprimentos de socorro. As buscas foram realizadas nas bases de dados *Scopus* e *Web of Science*, as palavras chave “forecasting”, “prediction”, “demand”, “resource”, “humanitarian”, “relief” e “emergency”, que foram procuradas no título, resumo e palavras chaves do artigo para o período de publicação 1990 a 2022.

Os resultados da pesquisa constataram que estudos de previsão dentro do contexto humanitário são limitados. A literatura é dominada por pesquisas sobre a previsão do desastre ou do evento que causa a resposta. No entanto, uma peça importante que está faltando é prever os recursos necessários para mitigar ou se recuperar de um desastre.

Revisões de literatura de previsão em operações humanitárias realizadas por Altay e Narayanan (2020) e Rostami-Tabar et al. (2022), reafirmaram a escassez de contribuições acadêmicas no que concerne a previsão de recursos no contexto humanitário; apontaram como possíveis causas a falta de conscientização, habilidades e compreensão do valor da previsão por parte das organizações com missões sociais. Com o objetivo de fomentar pesquisas e facilitar a troca construtiva de ideias entre a comunidade acadêmica e os setores público e privado, a International Institute of Forecasters (IIF) recentemente introduziram a Forecasting for Social Good (FSG) como uma área autônoma de estudos acadêmicos (ROSTAMI-TABAR et al., 2022), com temas relacionados a previsão para operações humanitárias e de socorro a desastres, incluindo a uma série de temas importantes com impacto social.

A Tabela 2 representa uma síntese das pesquisas que usaram estimativas de modelos climáticos ou estudos climatológicos para tomada de decisões humanitárias que levaram a uma ação precoce. Dois dos onze artigos abordaram a fase de preparação na gestão de desastres, sendo que as previsões foram aplicadas na tomada de decisão de estoque e pré posicionamento de itens de socorro; os outros nove artigos focaram na fase de resposta, usando as previsões para, em caso de desastres potenciais, enviar alertas precoces que acionam planos de curto prazo, reduzindo, assim, os riscos do impacto. A maioria, 6 dos 11 documentos, estudaram desastres associados a inundações, o qual pode ser influenciado pela significativa quantidade de previsões pluviométricas disponibilizadas pela área meteorológica.

Tabela 2: Artigos de modelos de previsão de desastres com aplicações humanitárias

Referência	Tipo de Desastre	Aplicação de Previsão
Dore (2003)	-	Priorização de esforço segundo a fase e tipo do desastre
Taskin e Jr (2010)	Furacão	Políticas de controle de estoque para suprimentos de socorro
Tall et al. (2012)	Inundação	Tomada de decisão na gestão de desastres
Braman et al. (2013)	Inundação	Estratégias de ação e preparação antecipada a desastres
Perez et al. (2015)	-	Sistema de financiamento para execução de planos de emergência
Jongman et al. (2015)	Inundação	Apoio a operações de resposta rápida a desastres
Pacheco e Batta (2016)	Furacão	Pré-posicionamento de itens de socorro
Perez et al. (2016)	Inundação	Alertas para iniciar as atividades de preparação a desastres
Lopez et al. (2020)	-	Abordagem de avaliação de previsões de desastres e envio de alertas para início de operações humanitárias.
Emerton et al. (2020)	Inundação	Avaliar o processo de criação de boletins de risco e sua aplicação na preparação e resposta humanitária
MacLeod et al. (2021)	Inundação	Avaliar a viabilidade de usar previsões relacionadas a eventos climáticos extremos para desenvolver sistemas e respostas de alerta precoce

Fonte: Elaborado pelo autor

Os artigos levantados na revisão de literatura de previsão para operações humanitárias e de socorro a desastres são apresentados na Tabela 3.

Tabela 3: Resumo de artigos que desenvolveram modelos de previsão de demanda humanitária

Referência	Tipo de Desastre	Variável resposta	Variáveis predictoras	Tamanho amostra	Método de Previsão	Métrica de Avaliação
Mude et al. (2009)	-	Probabilidade de fome	Tamanho do rebanho, taxa de mortalidade e lactação, dados de nutrição infantil, dados climáticos, qualidade e disponibilidade de água	-	LSDV	RMSE
Xu, Qi e Hua (2010)	Avalanches	Demanda de produtos agrícolas	Consumo diário de vegetais	93	EMD-ARIMA	MAPE
Deqiang, Yun e Changbing (2011)	Terremoto	Demanda de material de socorro	Magnitude e profundidade do terremoto, densidade populacional, número de pessoas afetadas	150	CBR, BP-NN	-
Holgún-Veras e Jaller (2012)	Furacão	Demanda por tipo de material de socorro	Número de solicitações de produtos de socorro	689	ARIMA	-
Wang, Wu e Sun (2012)	Inundação	Demanda de alimentos	População afetada, número de mortos, desaparecidos e feridos, área agrícola afetada, estoque de alimentos	71	SVR	MAE
Florez et al. (2013)	Terremoto	Número de pessoas afetadas	População, IDH, renda familiar per capita, percentual da população com acesso a água e eletricidade, vivendas vulneráveis, taxa de natalidade, grau de instrução, antecedentes criminais	-	PCA, MRA	Desvio percentual
Mohammadi, Ghomi e Zeinali (2014)	Terremoto	Demanda de alimentos	Consumo diário de pão, água e comida enlatada	90	RBF-NN	MAPE, RMSE
Xing, Zhonglin e Shaoyu (2015)	Terremoto	Número de pessoas afetadas	Magnitude e intensidade do terremoto, densidade populacional, taxa de morte, lesões, nível de pré-aviso, probabilidade de uma pessoa estar na área afetada, localização, suporte de abastecimento, taxa de colapso do edifício	19	v SVM	RMSE

Davis et al. (2016)	-	Doações de alimentos por tipo de doador e local	Tipo e quantidade de alimento, tipo de doador, agência de doação, data de recebimento	86.211	MA, EWMA, ARIMA	MAPE, Coeficiente de variação
Basu, Roy e DasBit (2018)	Terremoto	Demanda de alimentos	Número de famílias, tamanho médio da família, percentual de homens, mulheres e crianças, percentual de famílias abaixo da linha de pobreza nacional, poder aquisitivo, incremento no preço dos alimentos, lojas locais em funcionamento, percentual da demanda total que pode ser suprida por lojas locais	-	CBR, PCR	PB, NSE, AF
Jiang, Liu e Zheng (2019)	Terremoto	Demanda de sangue	Consumo de sangue diário	-	SVR	MAE, RMSE
Shao et al. (2020)	Terremoto	Demanda de material de socorro	Magnitude e intensidade do terremoto, população afetada, área afetada, destruição de habitações	-	IF-CBR	Validação cruzada
Paul e Davis (2021)	-	Doações por grupos de doadores	Data de doação, local de recebimento, afiliação do doador, qualidade e tipo do produto, classificação de armazenamento, peso da doação	105.433	MA, ARIMA, SVR, ETS, EM	MAPE
Fuqua e Hespeler (2022)	Furacão, Tsunami	Demanda de combustível	Lead time, quantidade necessária e aprovada, preço listado e pago, custo de estoque, quantidade inicial e final em estoque	1.186	RPCA e LSTM	RMSE, MAPE

Fonte: Elaborado pelo autor

Através da revisão da literatura, foi evidenciada uma preferência dos autores por modelos de Inteligência Artificial (IA). Support Vector Machine (SVM), Neural Network (NN), Case-Based Reasoning (CBR), Principal Component Analysis (PCA) ou modelos híbridos foram apresentados em dez dos quatorze artigos de previsão de demandas de desastres. Quatro artigos abordaram métodos de previsão de séries temporais, sendo o modelo Autoregressive integrated moving average (ARIMA) comum em todos eles. As métricas de avaliação usadas com maior frequência são a Mean absolute Error (MAE) e Root Mean Squared Error (RMSE).

Os terremotos foram o tipo de desastre mais abordado na literatura de previsão de demanda de recursos humanitários; furações, inundações, avalanches e outros desastres não especificados apresentaram-se em menor proporção. Coletivamente, esses documentos abordam todas as quatro fases do gerenciamento de desastres, porém a maioria dos autores abordaram questões de recursos na fase de resposta.

Alguns dos autores incluíram técnicas de clusterização para reduzir a variabilidade dos dados, assim, melhorar a performance dos modelos preditivos. Holguín-Veras e Jaller (2012) analisam padrões estatísticos das necessidades de recursos após um grande desastre e identificam 8 grupos principais de materiais de socorro, logo, estimam modelos estatísticos para obter as previsões de demanda para cada um desses grupos. Davis et al. (2016) propõem uma técnica para identificar o comportamento dos doadores e agrupá-los com base nesses atributos. Kmedoids e Kmeans foram os métodos de clusterização usados. Em seguida, desenvolveram um modelo de previsão com uma abordagem *ensemble-based* que incorpora vários modelos de previsão de séries temporais e Aprendizagem de Máquinas (AM). Paul e Davis (2021) dividiram os dados em oito classes, a partir das características da doação (ou seja, fonte, local, tipo de alimento). Os modelos preditivos foram construídos para cada uma das classes, para entender a relação entre a informação da doação e a precisão da previsão.

Diferentemente do presente estudo, que procura estimar a demanda anual de suprimentos de socorro sem assumir a ocorrência previa do desastre, todos os artigos levantados na revisão, apresentaram estimativas de demanda após o acontecimento de um desastre de grande magnitude e em geral de início súbito. Sendo assim, as informações para treinamento desses modelos foram coletadas após o impacto do desastre e, por conseguinte, não apresentam excesso de zeros nos dados.

3.2 Análise bibliométrica de modelos de regressão de dados de contagem com estimações inflacionadas de zeros

Para complementar a revisão da literatura, realizou-se uma busca de artigos que desenvolveram modelos de regressão de dados de contagem com estimações inflacionadas de zeros.

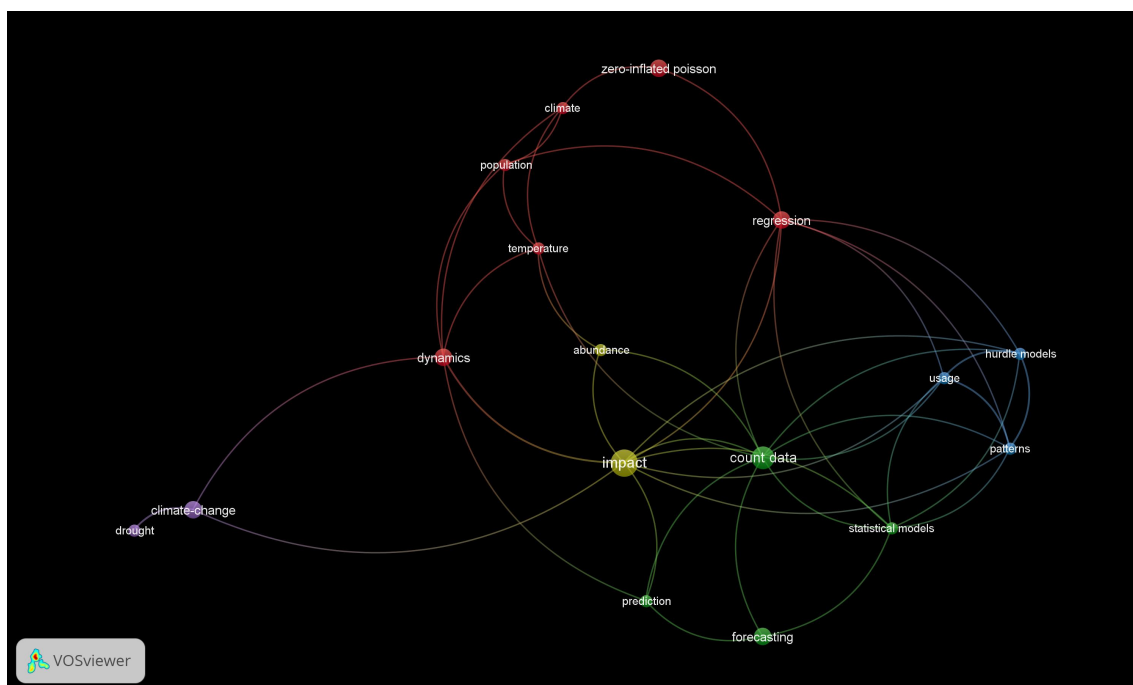
As palavras “count data”, “zero inflated” e “forecasting” foram usadas para filtrar os 31 documentos obtidos das bases de dados Scopus e Web of Science. Após remover artigos duplicados e algumas pesquisas com abordagens fora do foco do projeto, ficaram 24 documentos para análise.

Com o objetivo de identificar as aplicações dos modelos de regressão para dados de contagem com variável de resposta inflacionada de zeros e as áreas que estão explorando esses modelos preditivos, é apresentada uma análise bibliométrica dos artigos levantados na literatura.

As palavras *count data* e *zero inflated poisson* foram apresentadas pela primeira vez em 2016, o que mostra plenamente que são uma classe de modelos recentemente explorados na previsão. As fontes com maior destaque são *Analytic methods in accident research* e *climatic change* com 120 e 84 citações, respectivamente.

O software VOSviewer foi usado para gerar a visão de clusters de co-ocorrência das palavras-chaves. Um total de 17 palavras-chave com um frequência maior ou igual a 2 foram selecionadas de 192 palavras-chave, conforme mostrado na Figura 3.

Figura 3: Clusterização de co-ocorrência de palavras-chaves



Fonte: Elaborado pelo autor com base no VOSViewer

Na Figura 3 a área do nó e o tamanho da fonte dependem do valor do peso da palavra-chave. Quanto mais vezes a palavra-chave aparece e maior o nó; a linha entre os nós indica que uma palavra-chave ocorre simultaneamente com outra palavra-chave. As palavras chaves *impact* e *count data* são as mais frequentes. O termo *impacto*, ocorre simultaneamente com as expressões *câmbios climáticos* e *dinâmica*, sendo assim, pode ser relacionado a eventos de desastres. Por outro lado, o termo *dados de contagem* aparece junto a palavras relacionadas a *modelagem* e *previsão*. A ligação direta que existe entre as palavras chaves mais relevantes da rede indica a ocorrência simultânea de *impacto* e *dados de contagem* na maioria dos artigos da amostra.

Cinco clusters foram gerados a partir da análise de co-ocorrência de palavras chaves. O cluster roxo faz referência a desastres e mudanças climáticas, o vermelho agrupa modelos e regressões aplicados questões climáticas e populacionais. Os clusters amarelo e verde abarcam técnicas de modelagem e previsão. Finalmente, no cluster azul destaca-se o termo *hurdle models*, que representa modelos estatísticos onde uma variável aleatória é modelada por duas partes, a primeira que é a probabilidade de atingir o valor zero, e a segunda parte modela a probabilidade dos valores diferentes de zero.

Os modelos ZI e Hurdle Models (HM) diferem com base em sua conceituação dos zeros e interpretação dos parâmetros do modelo (FENG, 2021). Como mencionado na

Seção 2.2 o modelo ZI assume valores zero de uma mistura das duas distribuições, que são frequentemente chamados de zeros estruturais e zeros de amostragem (LAMBERT, 1992). Em contraste, o HM assume que todos os dados zero são de uma fonte estrutura, com uma parte do modelo sendo um modelo binário para modelar se a variável de resposta é zero ou positiva, e outra parte usando um modelo truncado, como um Poisson truncado ou uma distribuição BN truncada para os dados positivos (MULLAHY, 1986), (HEILBRON, 1994).

O modelo ZI assume que as observações zero se originam de zeros estruturais e de amostragem. Os dados de contagem deste estudo incluem excesso de zeros de amostragem e zeros estruturais. Por exemplo, zeros de amostragem ocorrem quando as contagens de demanda de suprimentos de socorro não são observadas em um município, no entanto, as contagens podem ser registradas para outros intervalos de tempo. Em contraste, municípios que nunca registraram contagens de demanda são considerados zeros estruturais. Portanto, o modelo ZINB é mais adequado.

Além do excesso de zeros, são evidenciados agrupamentos hierárquicos no conjunto de dados de demanda de materiais de socorro. As vulnerabilidades diferenciadas entre municípios, (localização em áreas propensas a riscos naturais, e fatores econômicos, políticos, sociais e culturais), são ignoradas pelo modelo ZINB. Nesse sentido, o modelo de regressão Zero-Inflated Negative Binomial Multilevel (ZINBM) que leva em consideração a existência de dependência entre observações de um mesmo grupo e, conseqüentemente, geram erros padrão não viesados, pode ser a melhor alternativa para este estudo.

4 METODOLOGIA

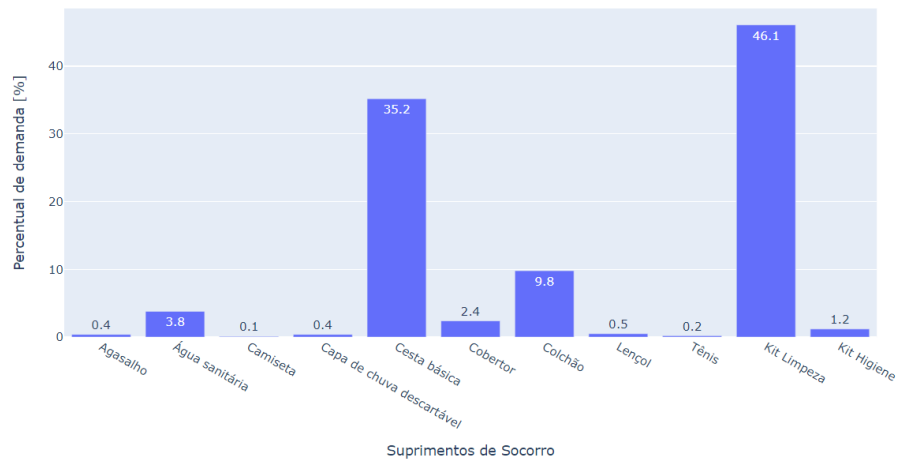
O presente trabalho busca desenvolver um modelo preditivo de demanda de suprimentos de socorro; de modo que as estimativas geradas serão usadas para decisões de localização de depósitos da CEPDEC-SP. Nesse sentido, esta seção apresenta os materiais e o método que serão usados para a modelagem.

4.1 Materiais

Dados de demanda de materiais de alívio foram fornecidos pela CEPDEC-SP. O histórico de 5 anos (2015-2020) contém 443 registros com informações anuais relacionadas ao município que gerou a demanda, além da data, da quantidade e da descrição do suprimento de socorro. Segundo o registro, 161 municípios solicitaram suprimentos de alívio pelo menos uma vez no espaço de tempo analisado; os outros 484 municípios não geraram nenhum registro, portanto sua demanda é considerada nula.

Mais do 80% da demanda é composta por cestas básicas de alimentos e kits de limpeza, enquanto os outros materiais foram solicitados em menor proporção, como mostra-se na Figura 4.

Figura 4: Composição da demanda por tipo de suprimento de socorro



Fonte: Elaborado pelo autor

A cesta básica de alimentos é repartida por família e contém quantidade suficiente para alimentar 4 pessoas por um período de 15 dias. O kit de limpeza também é enviado por família e é composto por itens básicos de higiene. A Figura 5 expõe esses materiais.

Figura 5: Cesta básica e parte do kit de limpeza estocado no armazém central da CEPDEC-SP

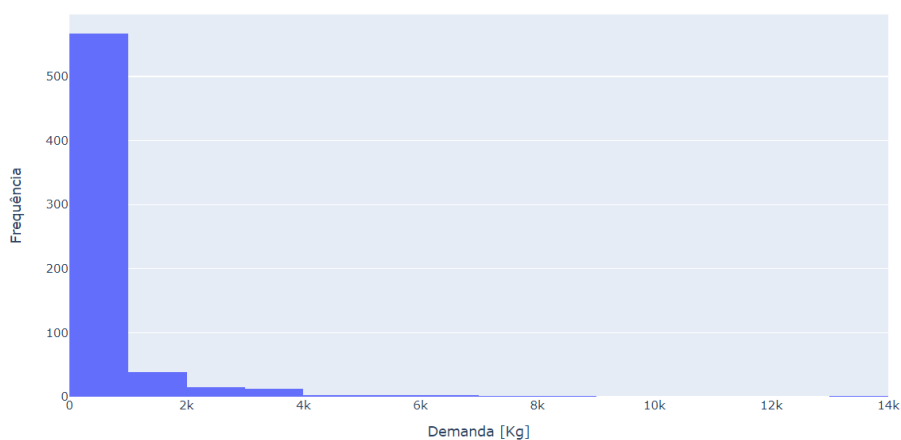


Fonte: Elaborado pelo autor

Para facilitar a modelagem, todos os itens de demanda foram transformados a unidades de peso (kg), com a informação levantada por Junior, Leiras e Yoshizaki (2020) no trabalho de campo realizado no depósito central da CEPDEC-SP, onde todos os materiais foram pesados e mensurados.

A Figura 6 apresenta a frequência na quantidade de solicitações realizadas por cada município, no período de 5 anos. Esta figura fornece uma representação visual dos valores zero contidos na variável de resultado, e evidencia uma provável superdispersão nos dados de demanda de suprimentos de socorro.

Figura 6: Histograma da demanda média de suprimentos de socorro



Fonte: Elaborado pelo autor

Os artigos levantados na revisão da literatura, selecionaram preditores associados a aspectos climáticos, demográficos, econômicos, características próprias do desastre e informações de vulnerabilidade social com o intuito de desenvolver os modelos preditivos no contexto humanitário. Consenza (2015) usou três variáveis explicativas na previsão do número de vítimas de desastres, a saber: quantidade de solicitações de ajuda humanitária, população e Índice de Desenvolvimento Humano (IDH). Considerando, que a maioria de desastres estão relacionados a inundações e deslizamentos de terra no Estado de São Paulo, Okane (2019) tentou adicionar o nível de chuvas às variáveis propostas por Consenza (2015), porém os resultados mostraram baixa correlação com a variável de resposta e nenhuma melhoria na performance do modelo.

Com finalidade de corroborar com os resultados apresentados por Okane (2019), uma análise das informações de precipitações disponíveis dos municípios do Estado de São Paulo, foi realizada. A análise das informações disponibilizadas pelo Departamento de Águas e Energia Elétrica (DAEE), expôs a ausência de medições pluviométricas para

certos municípios. Outra limitação é a exposição dos dados em períodos de tempo consideravelmente espaçados e a falta de informações relacionadas a magnitude e frequência das precipitações, que são essenciais para a avaliação de um potencial desastre.

Não obstante, a Figura 7 evidencia um incremento considerável de demanda de materiais de socorro no ano 2016 em relação aos outros anos do histórico. Esse pico de demanda coincide com a ocorrência de El Niño de 2015–2016 que foi considerado o evento mais forte em quase duas décadas (POTERIE et al., 2018).

Figura 7: Demanda de suprimentos de socorro por ano



Fonte: Elaborado pelo autor

Neste estudo, as estratégias para seleção das variáveis preditivas foram baseadas na compreensão das necessidades de suprimentos de socorro derivadas da vulnerabilidade, dos aspectos demográficos e das anomalias climáticas relacionadas ao fenômeno de El Niño, as quais são detalhadas na Tabela 4.

Tabela 4: Fatores preditores de previsão da demanda

Variáveis	Preditores	Descrição	Fonte
Índice Paulista de Vulnerabilidade Social (IPVS)	de So-	Indicador sintético de caracterização socioeconômica dos setores censitários no Estado de São Paulo	Governo Aberto SP
Anomalia Climática		Monitoramento de variações da temperatura, precipitações, pressão do ar e circulação atmosférica no Oceano Pacífico equatorial	El Niño Southern Oscillation (ENSO)
População estimada		Projeções da população baseadas no Censo Demográfico de 2010	Instituto Brasileiro de Geografia e Estatística (IBGE)

Fonte: Elaborado pelo autor

O IPVS contempla integralmente a questão de desigualdade dos municípios e a situação das suas áreas de concentração de pobreza. Informações de condições de vida da população, como renda, escolaridade, saúde, inserção no mercado de trabalho, acesso aos serviços prestados pelo Estado e as oportunidades de mobilidade social, foram consideradas pela Fundação Sede para avaliar e classificar os municípios do Estado de São Paulo em sete grupos segundo seu grau de vulnerabilidade (SEADE, 2010). A variável preditora de vulnerabilidade usada no modelo, foi resultado da agregação dos sete grupos de IPVS, em três grupos, a saber: vulnerabilidade “muito alta”, “alta” e “média”.

O preditor de anomalia climática é uma variável categórica relacionada diretamente com a presença (“sim”) ou ausência (“não”) do fenômeno de El Niño. Anualmente o sistema de previsão ENSO publica uma discussão diagnóstica e emite um status de alerta do fenômeno. ENSO é um padrão climático recorrente que envolve mudanças na temperatura das águas no Oceano Pacífico tropical, central e oriental. Esse padrão oscilante de aquecimento e resfriamento, afeta diretamente a distribuição das chuvas nos trópicos e pode ter uma forte influência no clima em diversas partes do mundo. El Niño e La Niña são as fases extremas do ciclo ENSO; entre essas duas fases possuiu uma terceira

fase neutra (NOAA, 2022).

O IBGE disponibiliza anualmente estimativas de população para os municípios brasileiros. A metodologia adotada para estimar os contingentes populacionais dos municípios baseia-se na relação da tendência de crescimento populacional do município, observada entre dois censos demográficos consecutivos IBGE. O preditor de população é constituído pelo logaritmo das estimativas da população.

Para determinar os níveis no modelo ZINBM, uma variável categórica foi gerada aplicando o método de agrupamento K-means (HARTIGAN; WONG, 1979). A suscetibilidade a desastres de cada município foi considerada como a única variável para calcular a distância entre as observações e gerar três clusters: “a”, “b” e “c”.

A Tabela 5 apresenta estatísticas descritivas e a frequência das variáveis quantitativas e qualitativas, respectivamente.

Tabela 5: Estatísticas descritivas ou frequência das variáveis envolvidas no modelo

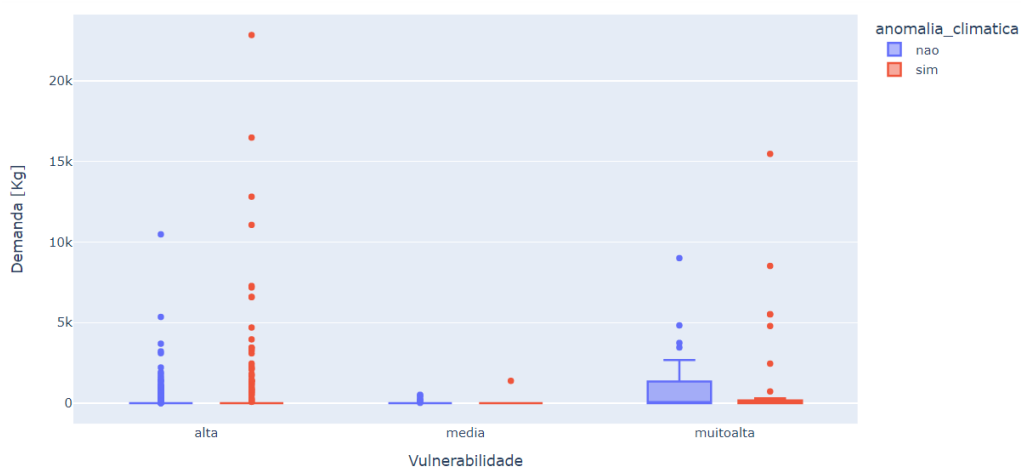
Demanda	Logaritmo população	Anomalia climática	Vulnerabilidade	Clusters
Min.: 0.0	Min.: 6.7			
1Q.: 0.0	1Q.: 8.7		freq. “média”: 270	freq. “a”: 15
Mediana: 0.0	Mediana: 9.6	Freq. “não”: 645	Freq. “alta”: 942	Freq. “b”: 1,124
3Q.: 0.0	3Q.: 10.7	freq. “sim”: 645	freq. “muitoalta”: 78	freq. “c”: 151
Max.: 22,847	Max.: 16.3			
Média: 241.9	Média: 9.7			

Fonte: Elaborado pelo autor

A variabilidade dos dados da demanda e sua relação com os preditores de vulnerabilidade e anomalia climática é apresentada na Figura 8. Constata-se uma relação direta entre a vulnerabilidade e a demanda, municípios mais vulneráveis geram maior demanda de suprimentos de socorro. Por outro lado, é evidenciada a relação direta entre a demanda e a ocorrência de anomalia climática. A demanda do ano 2016, correspondente ao fenômeno de El Niño (anomalia climática “sim”) são os que alcançam os valores máximos de demanda de suprimentos de socorro.

Pode-se observar para níveis de vulnerabilidade alta e média os valores de demanda do limite inferior, primeiro quartil (1Q), mediana (2Q), terceiro quartil (3Q) e limite superior, coincidem e estão concentrados no valor de zero. Por outro lado, municípios com vulnerabilidade alta apresentaram uma super-dispersão maior de demanda em relação a media e muito alta. Contudo, os três níveis de vulnerabilidade apresentam valores discrepantes de demanda que estão fora do intervalo definido, altos demais.

Figura 8: Boxplot da demanda de suprimentos de socorro



Fonte: Elaborado pelo autor

Para fines ilustrativos, extraiu-se parte dos dados das variáveis envolvidas no modelo no Apêndice A.

No processo foram exploradas, sem muito sucesso, outras variáveis socio-econômicas, a saber:

- Registros do CADÚnico, que identifica as famílias de baixa renda existentes no Brasil.
- Índice de Desenvolvimento Humano (IDH), que determina o grau de desenvolvimento de educação, saúde e renda.
- Índice Paulista de Responsabilidade Social (IPRS), que classifica o desenvolvimento e a qualidade de vida dos municípios paulistas.
- Mapas de áreas de risco de municípios, que avaliam as regiões com base na abordagem da paisagem.

Não encontrou-se relação entre a demanda de suprimentos de socorro e os dados levantados de CADÚnico, IDH, IPRS. Pode ser que as estratégias de ponderação e de

operacionalização das diferentes dimensões (renda, escolaridade, acesso a saúde e serviços básicos, etc.) distorceram a caracterização de vulnerabilidade nesses indicadores. No caso dos mapas de áreas de risco, o impedimento foi a carência de informações para todos os municípios.

4.2 Método

Os modelos de regressão para dados de contagem apresentam uma perspectiva adequada para o desenvolvimento do modelo de previsão de necessidades de suprimentos de socorro.

Posto que o estudo apresenta dados de demanda que assumem um limitado conjunto de valores estritamente positivos e não lineares em relação as variáveis preditoras, o modelo de previsão de necessidades de suprimentos de socorro é abordado com a técnica avançada de modelagem estatística GLM, especificamente com os modelos de regressão para dados de contagem.

Dada a existência de vulnerabilidades diferenciadas dos municípios, relacionadas a localização em áreas propensas a riscos naturais e onde questões econômicas, políticas, sociais e culturais reduzem as oportunidades de prevenção de desastres naturais; é proposta a estrutura de modelo multinível GLMM que leva em consideração esses agrupamentos hierárquicos evidenciados no conjunto de dados.

Para determinar os níveis do modelo ZINBM, inicialmente se considerou o agrupamento dos municípios nas onze regiões geográficas intermediárias, estabelecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Porém, com essa estrutura o modelo GLMM não apresentou melhoria em relação ao modelo GLM. Portanto, realizou-se uma análise de cluster para identificar e particionar os dados em grupos semelhantes. Assim, os 645 municípios do Estado de São Paulo foram particionados em três grupos, através do algoritmo *K-means*, considerando como única dimensão a quantidade de desastres registrados no histórico da CEPDEC-SP. O método *K-means* minimiza a distância euclidiana entre cada ponto e o centros de cluster (HARTIGAN; WONG, 1979).

Os modelos de regressão Poisson e BN são apresentados como ponto de partida. A existência excessiva de zeros na variável de resposta foi confirmada pelo Teste de Vong (VUONG, 1989) e o teste proposto por (CAMERON; TRIVEDI, 1990) evidenciou a característica de superdispersão nos dados da demanda de suprimentos de socorro. Assim, os modelos análogos que consideram inflação de zeros ZIP e ZINB, bem como o modelo

que considera a inflação de zeros sob a perspectiva multinível ZINBM, foram desenvolvidos no estudo.

Por fim, os valores de AIC e log-verossimilhança dos modelos estimados foram comparados para verificar a melhor adequação aos dados considerados, e apontar os vieses gerados por não considerar a inflação de zeros e a estrutura hierárquica dos dados. Neste trabalho, todas as estimativas são obtidas através do software R versão 4.0.4, usando o pacote MASS para os modelos Poisson e BN, o pacote pscl para o modelo ZINB e o pacote glmmTMB para o modelo ZINBM.

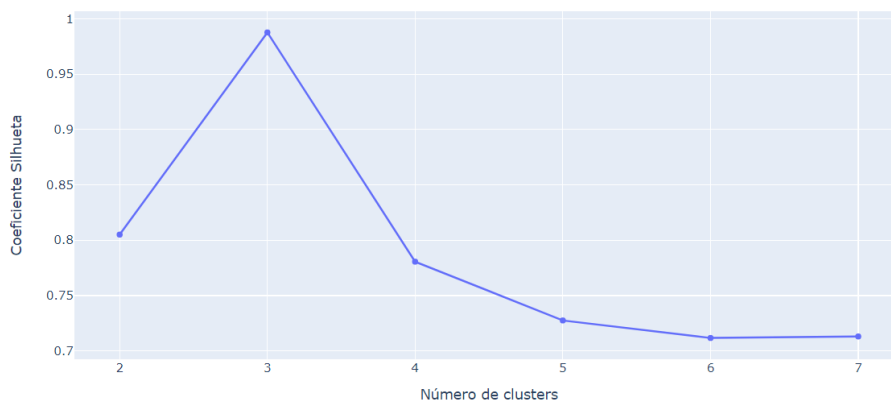
5 RESULTADOS

Nesta secção são apresentados os resultados e análises dos modelos desenvolvidos para previsão da demanda de suprimentos de socorro.

5.1 Resultados e Discussão

Os 645 municípios do Estado de São Paulo foram particionados em três grupos, através do algoritmo *K-means*, considerando como única dimensão a quantidade de desastres registrados no histórico da CEPDEC-SP. Para definir o número ótimo de clusters, foi testado diferentes valores de clusters e calculado o correspondente coeficiente silhueta para cada configuração (ROUSSEEUW, 1987); A Figura 9 mostra os valores do coeficiente silhueta, onde pode ser evidenciado que o valor ótimo de clusters é três.

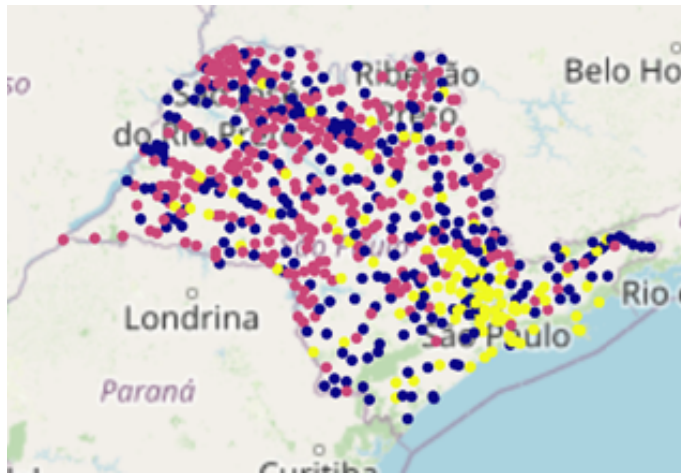
Figura 9: Coeficiente silhueta por número de clusters



Fonte: Elaborado pelo autor

A Figura 10 apresenta uma visão geográfica onde cada grupo é representado por uma cor. Esses clusters de municípios caracterizam os níveis hierárquicos do modelo ZINBM. Assim, o modelo proposto apresenta diferentes termos de erro relacionados a interceptos e a inclinações para cada grupo.

Figura 10: Agrupamento dos municípios do Estado de São Paulo.



Cores azul, amarelo e rosa correspondem aos clusters a, b e c, respectivamente.

Fonte: Elaborado pelo autor

O resultado do teste proposto por Cameron e Trivedi (1990) é apresentado na Figura 11. Como o valor *p-value* do teste *t* é menor ao nível de significância (5%) a existência de superdispersão é identificada nos dados da variável de resposta, fazendo com que o modelo de regressão BN seja mais adequado que o modelo Poisson.

Figura 11: Resultado do teste para verificação de existência de superdispersão

Overdispersion Test - Cameron & Trivedi (1990)

```
data: df
Lambda t test score: = 4.4755, p-value = 8.295e-06
alternative hypothesis: overdispersion if lambda p-value is less than
or equal to the stipulated significance level
```

Fonte: Elaborado pelo autor

Para confirmar o excesso de zeros na variável de resposta da demanda de suprimentos de socorro foi aplicado o teste de Vuong (VUONG, 1989), que compara as funções de verossimilhança entre o modelo inflacionado de zeros com o modelo análogo tradicional. A Tabela 6 mostra os resultados do teste de Vuong e as correções de Desmarais e Harden (2013) ao comparar o modelo de regressão BN com o modelo ZINB.

Tabela 6: Resultados do teste de Voung do modelo de regressão BN x ZINB

Estatística-z de Voung -16		
original	40.6094	$2.22e^{-16}$
AIC-corrigido	40.6088	$2.22e^{-16}$
BIC-corrigido	40.6073	$2.22e^{-16}$

Fonte: Elaborado pelo autor

O valor da estimativa do teste Voung é $z = 40.6094$, as estatísticas corrigidas AIC e BIC são $z = 40.6088$ e $z = 40.6073$, respectivamente. Todos os resultados apresentam valores de $p < 0.05$. Portanto, o modelo ZINB é mais apropriado em comparação com o modelo BN.

Os resultados da estimação dos modelos BN, ZINB, e ZINBM, são apresentados na Tabela 7. Verifica-se que todos os parâmetros apresentam significância estatística considerando um nível de significância de 1%.

Tabela 7: Estimativas do modelo BN, ZINB e ZINBM

	BN	ZINB	ZINBM
Efeitos Fixos			
Intercepto	-1.05***	2.98***	3.96***
anomalia_climatica sim	0.60***	1.26***	1.99***
pop_estimada	0.59***	0.35***	0.26***
Componente Logístico			
vulnerabilidade alta	-	1.74***	1.74***
vulnerabilidade media	-	1.62***	1.62***
vulnerabilidade muito alta	-	-1.43***	-1.43***
Efeitos Aleatórios			
grupo a	-	-	1.31
grupo b	-	-	-1.21
grupo c	-	-	0.19
Observações	1,290	1,290	1,290
Log-verossimilhança	-191,674	-1,964	-1,954

***Estatisticamente significativamente diferente de zero com 99% de confiança

Fonte: Elaborado pelo autor

As Equações 5.1, 5.2 e 5.3 correspondem aos modelos BN, ZINB e ZINBM, respectivamente.

$$d_i = -1.04968 + 0.60127anomalia_climatica_{sim_i} + 0.59088pop_estimada_i \quad (5.1)$$

$$d_i = \left\{ 1 - \frac{1}{1 + e[-(1.74095 + 1.62079vul_{media_i} - 1.43442vul_{muitoalta_i})]} \right\}^* e^{[2.98137 + 1.26171anomalia_climatica_{sim_i} - 0.34713pop_estimada_i]} \quad (5.2)$$

$$d_{ij} = \left\{ 1 - \frac{1}{1 + e[-(1.74001 + 1.6208vul_{media_{ij}} - 1.43433vul_{muitoalta_{ij}})]} \right\}^* e[3.95972 + 1.99652anomalia_{climatica_{sim_{ij}}} - 0.26204pop_{estimada_{ij}} + v_{0j}] \quad (5.3)$$

Ao considerar o modelo BN em relação aos modelos ZINB e ZINBM, percebe-se uma subestimação dos parâmetros de anomalia climática e uma superestimação na variável preditora do logaritmo de população estimada.

O modelo BN apresenta viés por não considerar o excesso de zeros na variável de resposta, além disso a estimativa do BN não identifica o efeito da vulnerabilidade. O componente logístico dos modelos ZINB e ZINBM, indica que o aumento de uma unidade na vulnerabilidade média incrementa a chance de zeros estruturais em 3% ($e^{1.74+1.62}/(1 + e^{(1.74+1.62)}) = 0.97$), e 15% e 42% para a vulnerabilidade alta e muito alta, respectivamente. Confirmando que a variável de vulnerabilidade tem uma preponderância na ausência de demanda de suprimentos de socorro nos municípios.

Mesmo que o modelo ZINB leve em consideração a inflação de zeros, ele não consegue capturar o aninhamento natural dos dados entre os municípios. Esse viés não é evidenciado no componente logístico, porém, observa-se diferentes termos de erro relacionados a interseptos e inclinações para cada grupo de municípios que caracteriza os níveis hierárquicos do modelo ZINBM, além de uma ligeira suavização na variável de logaritmo da população estimada.

O ajuste dos modelos GLM foram avaliados através de AIC e log-verossimilhança. Para fins de comparação, além do modelo ZINBM, este estudo estima os seguintes modelos: Poisson, ZIP, BN e ZINB. A Tabela 8 apresenta o valor de AIC e log-verossimilhança associado a cada modelo.

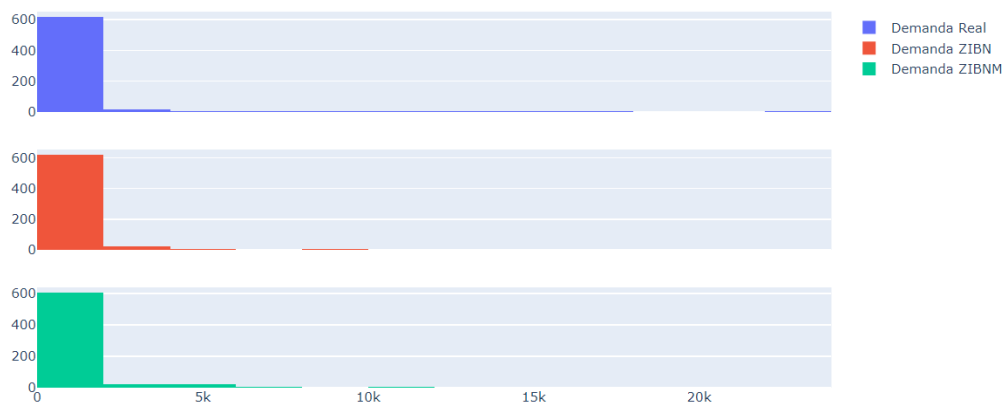
Tabela 8: Comparação de medidas de qualidade de ajuste dos modelos

Modelos	log-verossimilhança	AIC
Poisson	- 708,513	1,417,033
BN	- 708,058	1,416,125
ZIP	- 191,674	383,361
ZINB	- 1,964	3,942
ZINBM	- 1,954	3,925

Os resultados dos modelos são igualmente validados comparando os valores da demanda prevista com a demanda real.

A Figura 12 apresenta histogramas da demanda, e estimativas dos modelos ZINB e ZINBM. Ambos modelos conseguiram explicar o comportamento de excesso de zeros na demanda. Embora, os valores de demanda altos, que podem ser inferidos como resultado de ocorrências de desastres, são melhor estimados pelo modelo ZINBM, existem casos que as estimativas da demanda são sobrestimadas, o que pode induzir a um incremento nos custos de estoque. Contudo, no contexto de desastres e emergência é prioridade evitar a falta de suprimentos de socorro, que pode significar perdas humanas e sofrimento.

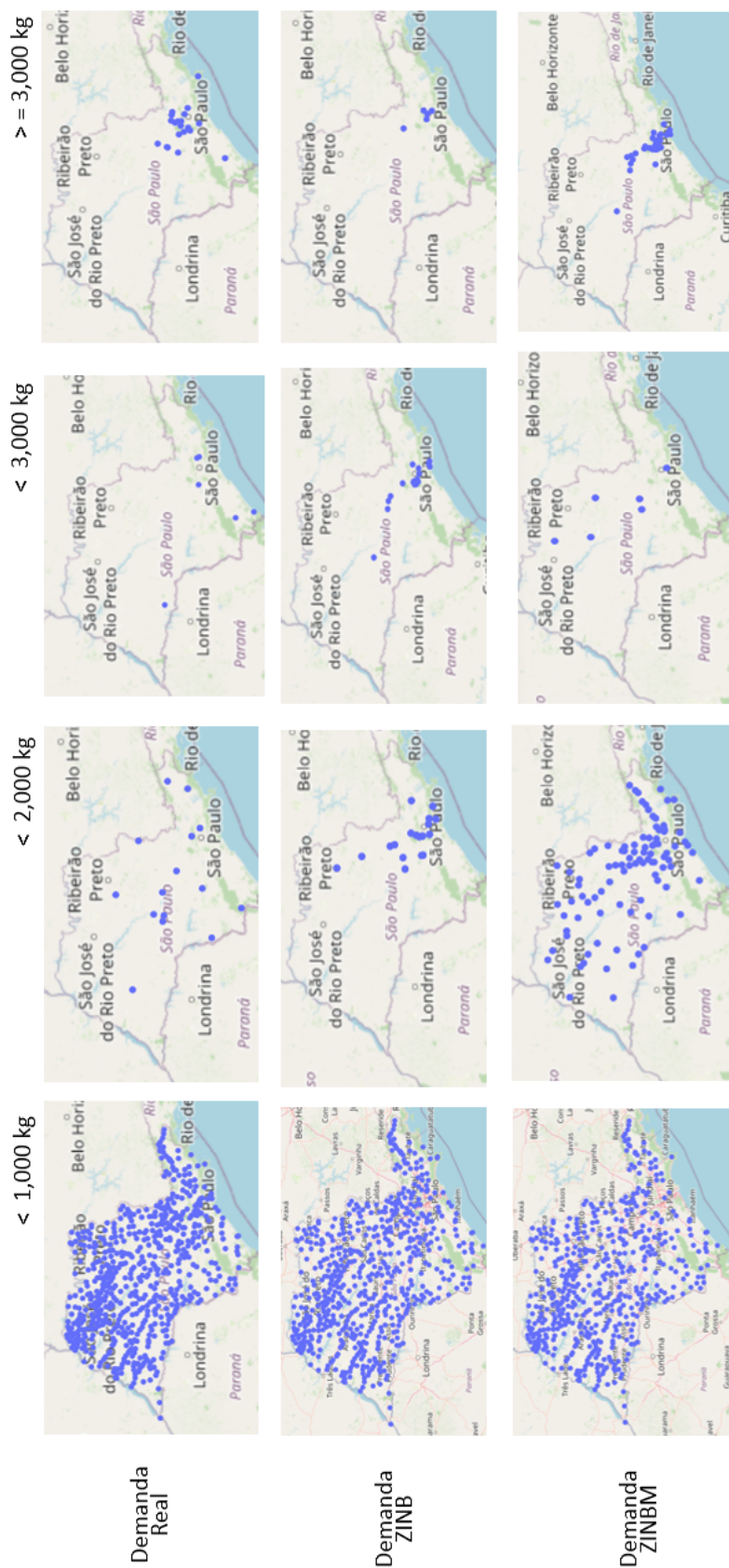
Figura 12: Histograma da demanda real x demanda estimada



Fonte: Elaborado pelo autor

A Figura 13 mostra a demanda geograficamente, onde cada coluna corresponde a um intervalo de demanda. Ambos modelos, ZINB e ZINBM, apresentaram uma concentração de estimativas no primeiro intervalo de 0 até 999 kg, que é provocada pelo excesso de zeros nos dados da demanda real. No intervalo de 1,000 a 1,999 kg o modelo ZINBM superestima a demanda para alguns municípios e pode-se considerar um melhor desempenho do modelo ZINB. Nos intervalos de 2,000 a 2,999 kg e maior a 3,000 kg de demanda o modelo ZINBM gera estimativas mais aderentes, em relação ao modelo ZINB.

Figura 13: Visão geográfica da demanda real x demanda estimada



Fonte: Elaborado pelo autor

É interessante notar que os municípios do cluster amarelo da Figura 10 correspondem a regiões com estimativas de demanda superiores a 2,000 kg na Figura 13, sendo que a maioria desses municípios constituem a região Metropolitana de São Paulo.

Os resultados da validação gráfica e as medidas de AIC e log-verossimilhança, sugerem que o desempenho do modelo ZINBM é satisfatório e superior em relação aos outros modelos apresentados, portanto, foi selecionado como o modelo final para estimar a demanda de suprimentos de demanda de suprimentos de socorro.

6 CONCLUSÕES

Nesta seção são expostas as conclusões, considerações finais e trabalhos futuros.

6.1 Conclusões

Para estimar a demanda anual de suprimentos de socorro dos 645 municípios que constituem o Estado de São Paulo, é proposto o modelo ZINBM, que considera a inflação de zeros sob a perspectiva multinível e mostrou-se superior a todos os outros modelos desenvolvidos.

As estratégias para seleção das variáveis preditivas foram baseadas na compreensão das necessidades de suprimentos de socorro. As variáveis preditoras são derivadas de indicadores de vulnerabilidade, fatores demográficos e ocorrências de anomalias climáticas relacionadas ao fenômeno de El Niño. Para estabelecer os níveis do modelo hierárquico considerou-se a quantidade de desastres como única dimensão.

Os modelos de regressão Poisson e BN foram desenvolvidos como ponto de partida. A existência excessiva de zeros na variável de resposta e a superdispersão nos dados da demanda de suprimentos de socorro foi confirmada pelo Teste de Vong (VUONG, 1989) e o teste proposto por Cameron e Trivedi (1990), respectivamente. Assim, os modelos análogos que consideram inflação de zeros ZIP e ZINB, bem como o modelo que considera a inflação de zeros sob a perspectiva multinível ZINBM, são desenvolvidos no estudo.

O ajuste dos modelos foi avaliado através das métricas de AIC e log-verossimilhança. Realizou-se também, um comparativo dos valores da demanda estimada com a demanda real para avaliar o desempenho dos modelos ZINB e ZINBM.

O modelo ZINBM alcançou os valores mais baixos de AIC e log-verossimilhança e gerou as estimativas mais aderentes dentre todos os modelos desenvolvidos. Portanto, foi selecionado como o modelo final para estimar a demanda de suprimentos de demanda de suprimentos de socorro.

O modelo proposto pode ser atualizado de forma fácil e regular com novas informações de população estimada e previsões ENSO. Tal modelo de previsão pode revelar-se uma ferramenta valiosa para a conscientização da demanda de suprimentos para resposta a emergências.

Reduzindo a incerteza na demanda, soluções adequadas para o problema de localização de depósitos podem ser atingidas pelo modelo de otimização. Finalmente, a escolha da solução será assistida pelo SAD que permite ao tomador de decisão personalizar os locais de instalação dos depósitos e rapidamente verificar os resultados em relação ao custo e tempo de entrega.

6.2 Limitações e Trabalhos Futuros

A quantidade e qualidade da informação disponível foi a principal limitação do estudo. O modelo foi desenvolvido com dados de demanda dos anos 2015 a 2019, as informações disponíveis do ano 2020 foram excluídas por ser um ano atípico de pandemia. Do mesmo modo, a ausência de medições pluviométricas para certos municípios, a exposição dos dados em períodos de tempo consideravelmente espaçados e a inexistência de informações relacionadas a magnitude e frequência, impediram o uso dessa potencial variável preditora. Limitações metodológicas para replicar os agrupamentos dos municípios, que foram empregados como níveis do modelo ZINBM, são originadas pelo uso do algoritmo não supervisionado *K-means* para o problema de clusterização.

Pesquisas futuras devem se concentrar na coleta de dados de demanda de períodos mais atuais para aprimorar o desempenho do modelo. Outra oportunidade consta na identificação de fatores explicativos da estrutura hierárquica para substituir os clusters dos municípios empregados no modelo ZINBM, a fim de obter resultados reproduzíveis baixo diversas condições.

Outras perspectivas de estudos são identificadas a partir da Figura 8; considerando a dispersão dos dados de demanda em relação a vulnerabilidade, aplicar modelos ZIP e Zero-Inflated Poisson Multilevel (ZIPM) para níveis de vulnerabilidade media e muito alta podem significar uma melhoria nas estimativas da demanda.

Por outro lado, no estudo só foram explorados modelos ZI que geram valores zeros da mistura de zeros estruturais e de amostragem; abordagens como os HM que assume que todos os dados zeros são de uma fonte estrutura, podem ser desenvolvidos para avaliar e/ou comparar seu desempenho em relação aos modelos ZI.

REFERÊNCIAS

- AGRESTI, A. *Foundations of linear and generalized linear models*. [S.l.]: John Wiley & Sons, 2015.
- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974.
- ALCÁNTARA-AYALA, I. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology*, Elsevier, v. 47, n. 2-4, p. 107–124, 2002.
- ALTAY, N.; NARAYANAN, A. Forecasting in humanitarian operations: Literature review and research needs. *International journal of forecasting*, Elsevier, 2020.
- BALCIK, B.; BEAMON, B. M. Facility location in humanitarian relief. *International Journal of logistics*, Taylor & Francis, v. 11, n. 2, p. 101–121, 2008.
- BALLOU, R. *Gerenciamento da Cadeia de Suprimentos: Logística Empresarial*. [S.l.]: Bookman Editora, 2009.
- BASU, S.; ROY, S.; DASBIT, S. A post-disaster demand forecasting system using principal component regression analysis and case-based reasoning over smartphone-based dtn. *IEEE Transactions on Engineering Management*, IEEE, v. 66, n. 2, p. 224–239, 2018.
- BELARDO, S. et al. A partial covering approach to siting response resources for major maritime oil spills. *Management Science*, INFORMS, v. 30, n. 10, p. 1184–1196, 1984.
- BRAMAN, L. M. et al. Climate forecasts in disaster management: Red cross flood operations in west africa, 2008. *Disasters*, Wiley Online Library, v. 37, n. 1, p. 144–164, 2013.
- BRITO, I. de J. *Location of relief supplies warehouses for São Paulo state Civil Defense*. Tese (Doutorado) — Universidade de São Paulo, 2015.
- BROLLO, M.; FERREIRA, C. Indicadores de desastres naturais no estado de são paulo. 2009.
- CAMERON, A. C.; TRIVEDI, P. K. Regression-based tests for overdispersion in the poisson model. *Journal of econometrics*, Elsevier, v. 46, n. 3, p. 347–364, 1990.
- CAMERON, A. C.; TRIVEDI, P. K. *Regression analysis of count data*. [S.l.]: Cambridge university press, 2013. v. 53.
- CAMERON, A. C.; TRIVEDI, P. K. et al. *Microeconometrics using stata*. [S.l.]: Stata press College Station, TX, 2010. v. 2.

CAUNHYE, A. M.; NIE, X.; POKHAREL, S. Optimization models in emergency logistics: A literature review. *Socio-economic planning sciences*, Elsevier, v. 46, n. 1, p. 4–13, 2012.

CEPDEC. *Histórico*. 1995. Disponível em: <http://www.defesacivil.sp.gov.br/historico>.

CHANG, M.-S.; TSENG, Y.-L.; CHEN, J.-W. A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, Elsevier, v. 43, n. 6, p. 737–754, 2007.

CONSENZA, A. *Location of relief supplies warehouses for São Paulo state Civil Defense*. Dissertação (Mestrado) — Universidade de São Paulo, 2015.

COTES, N.; CANTILLO, V. Including deprivation costs in facility location models for humanitarian relief logistics. *Socio-Economic Planning Sciences*, Elsevier, v. 65, p. 89–100, 2019.

DASKIN, M. *Network and Discrete Location*. [S.l.]: John Wiley & Sons, Inc, 1995.

DAVIS, L. B. et al. Analysis and prediction of food donation behavior for a domestic hunger relief organization. *International Journal of Production Economics*, Elsevier, v. 182, p. 26–37, 2016.

DEQIANG, F.; YUN, L.; CHANGBING, L. Forecasting the demand for emergency supplies: Based on the cbr theory and bp neural network. In: *Proceedings of international conference on innovation & management, Article*. [S.l.: s.n.], 2011. v. 45, n. 2011, p. 700–704.

DESMARAIS, B. A.; HARDEN, J. J. Testing for zero inflation in count models: Bias correction for the vuong test. *The Stata Journal*, SAGE Publications Sage CA: Los Angeles, CA, v. 13, n. 4, p. 810–835, 2013.

DESSOUKY, M. et al. Rapid distribution of medical supplies. In: *Patient Flow: Reducing Delay in Healthcare Delivery*. [S.l.]: Springer, 2006. p. 309–338.

DORE, M. H. Forecasting the conditional probabilities of natural disasters in canada as a guide for disaster preparedness. *Natural Hazards*, Springer, v. 28, n. 2, p. 249–269, 2003.

DURAN, S.; GUTIERREZ, M. A.; KESKINOCAK, P. Pre-positioning of emergency items for care international. *Interfaces*, Informs, v. 41, n. 3, p. 223–237, 2011.

EMERTON, R. et al. Emergency flood bulletins for cyclones idai and kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *International Journal of Disaster Risk Reduction*, Elsevier, v. 50, p. 101811, 2020.

FÁVERO, L. P.; BELFIORE, P. *Manual de análise de dados*. Cham: Elsevier Editora Ltda, 2017. 695–784 p. ISBN 978-85-352-7087-7.

FÁVERO, L. P. L. The zero-inflated negative binomial multilevel model: demonstrated by a brazilian dataset. *International Journal of Mathematics in Operational Research*, Inderscience Publishers (IEL), v. 11, n. 1, p. 90–106, 2017.

FEMA. *About disaster management*. 2019. Emergency Management Institute. Disponible em: [⟨https://training.fema.gov/emiweb⟩](https://training.fema.gov/emiweb).

FENG, C. X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of statistical distributions and applications*, Springer, v. 8, n. 1, p. 1–19, 2021.

FLOREZ, J. V. et al. Towards a demand forecast methodology for recurrent disasters. *WIT transactions on the built environment*, WIT Press, v. 133, p. 99–110, 2013.

FOX, J. *Applied regression analysis and generalized linear models*. [S.l.]: Sage Publications, 2015.

FRAZELLE, E. H.; SOJO, R. Q. *Logística de almacenamiento y manejo de materiales de clase mundial*. [S.l.]: Grupo Editorial Norma, 2006.

FUQUA, D.; HESPELER, S. Commodity demand forecasting using modulated rank reduction for humanitarian logistics planning. *Expert Systems with Applications*, Elsevier, p. 117753, 2022.

HALL, D. B. Zero-inflated poisson and binomial regression with random effects: a case study. *Biometrics*, Wiley Online Library, v. 56, n. 4, p. 1030–1039, 2000.

HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, JSTOR, v. 28, n. 1, p. 100–108, 1979.

HEILBRON, D. C. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, Wiley Online Library, v. 36, n. 5, p. 531–547, 1994.

HEZAM, I. M. et al. A systematic literature review on mathematical models of humanitarian logistics. *Symmetry*, Multidisciplinary Digital Publishing Institute, v. 13, n. 1, p. 11, 2021.

HOLGUÍN-VERAS, J.; JALLER, M. Immediate resource requirements after hurricane katrina. *Natural Hazards Review*, American Society of Civil Engineers, v. 13, n. 2, p. 117–131, 2012.

HORNER, M. W.; DOWNS, J. A. Optimizing hurricane disaster relief goods distribution: model development and application with respect to planning strategies. *Disasters*, Wiley Online Library, v. 34, n. 3, p. 821–844, 2010.

IFRC. *About disaster management*. 2021. International Federation of Red Cross and Red Crescent Societies. Disponible em: [⟨https://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management⟩](https://www.ifrc.org/en/what-we-do/disaster-management/about-disaster-management).

IFRC. *What is a disaster?* 2021. International Federation of Red Cross and Red Crescent Societies. Disponible em: [⟨https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster⟩](https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster).

JIA, H.; ORDÓÑEZ, F.; DESSOUKY, M. A modeling framework for facility location of medical services for large-scale emergencies. *IIE transactions*, Taylor & Francis, v. 39, n. 1, p. 41–55, 2007.

JIANG, P.; LIU, X.; ZHENG, M. Emergency blood demand forecasting after earthquakes. *IFAC-PapersOnLine*, Elsevier, v. 52, n. 13, p. 773–777, 2019.

JONGMAN, B. et al. Early flood detection for rapid humanitarian response: harnessing near real-time satellite and twitter signals. *ISPRS International Journal of Geo-Information*, MDPI, v. 4, n. 4, p. 2246–2266, 2015.

JUNIOR, I. d. B.; LEIRAS, A.; YOSHIKAZI, H. T. Y. A multi-criteria stochastic programming approach for pre-positioning disaster relief supplies in brazil. *Production, SciELO Brasil*, v. 30, 2020.

KINAY, Ö. B. et al. Modeling the shelter site location problem using chance constraints: A case study for istanbul. *European Journal of Operational Research*, Elsevier, v. 270, n. 1, p. 132–145, 2018.

KONGSOMSAKSAKUL, S.; YANG, C.; CHEN, A. Shelter location-allocation model for flood evacuation planning. *Journal of the Eastern Asia Society for Transportation Studies*, Eastern Asia Society for Transportation Studies, v. 6, p. 4237–4252, 2005.

LAAN, E. Van der et al. Demand forecasting and order planning for humanitarian logistics: An empirical assessment. *Journal of Operations Management*, Elsevier, v. 45, p. 114–122, 2016.

LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, Taylor & Francis, v. 34, n. 1, p. 1–14, 1992.

LEE, A. H. et al. Multi-level zero-inflated poisson regression modelling of correlated count data with excess zeros. *Statistical methods in medical research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 15, n. 1, p. 47–61, 2006.

LOPEZ, A. et al. Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings. *Weather and climate extremes*, Elsevier, v. 27, p. 100167, 2020.

MACLEOD, D. et al. Verification of forecasts for extreme rainfall, tropical cyclones, flood and storm surge over myanmar and the philippines. *Weather and climate extremes*, Elsevier, v. 33, p. 100325, 2021.

METE, H. O.; ZABINSKY, Z. B. Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics*, Elsevier, v. 126, n. 1, p. 76–84, 2010.

MIAOU, S.-P.; LUM, H. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, Elsevier, v. 25, n. 6, p. 689–709, 1993.

MOHAMMADI, R.; GHOMI, S. F.; ZEINALI, F. A new hybrid evolutionary based rbf networks method for forecasting time series: a case study of forecasting emergency supply demand time series. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 36, p. 204–214, 2014.

MONZÓN, J.; LIBERATORE, F.; VITORIANO, B. A mathematical pre-disaster model with uncertainty and multiple criteria for facility location and network fortification. *Mathematics*, Multidisciplinary Digital Publishing Institute, v. 8, n. 4, p. 529, 2020.

- MUDE, A. G. et al. Empirical forecasting of slow-onset disasters for improved emergency response: An application to kenya's arid north. *Food policy*, Elsevier, v. 34, n. 4, p. 329–339, 2009.
- MULLAHY, J. Specification and testing of some modified count data models. *Journal of econometrics*, Elsevier, v. 33, n. 3, p. 341–365, 1986.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.
- NGATCHOU-WANDJI, J.; PARIS, C. On the zero-inflated count models with application to modelling annual trends in incidences of some occupational allergic diseases in france. *Journal of Data Science*, School of Statistics, Renmin University of Chinaf, v. 9, n. 4, p. 639–659, 2011.
- NOAA. *El Niño Southern Oscillation (ENSO) Diagnostic Discussion*. 2022. Disponível em: https://www.cpc.ncep.noaa.gov/products/analysis/_monitoring/enso/_advisory.
- OKANE, D. T. *Criação de sistema de apoio à decisão interativo para a escolha da localização dos depósitos da CEPDEC-SP*. Trabalho de Formatura de Graduação — Universidade de São Paulo, 2019.
- ONU. *Los desastres son la nueva normalidad*. 2019. Disponível em: <https://news.un.org/es/interview/2019/11/1465021>.
- PACHECO, G. G.; BATTA, R. Forecast-driven model for prepositioning supplies in preparation for a foreseen hurricane. *Journal of the Operational Research Society*, Springer, v. 67, n. 1, p. 98–113, 2016.
- PAUL, S.; DAVIS, L. B. An ensemble forecasting model for predicting contribution of food donors based on supply behavior. *Annals of Operations Research*, Springer, p. 1–29, 2021.
- PEREZ, E. Coughlan de et al. Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, Copernicus GmbH, v. 20, n. 9, p. 3549–3560, 2016.
- PEREZ, E. Coughlan de et al. Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences*, Copernicus GmbH, v. 15, n. 4, p. 895–904, 2015.
- PÉREZ-GALARCE, F. et al. An optimization model for the location of disaster refuges. *Socio-Economic Planning Sciences*, Elsevier, v. 59, p. 56–66, 2017.
- PERUMEAN-CHANEY, S. E. et al. Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 83, n. 9, p. 1671–1683, 2013.
- POTERIE, A. S. T. de la et al. Understanding the use of 2015–2016 el niño forecasts in shaping early humanitarian action in eastern and southern africa. *International Journal of Disaster Risk Reduction*, Elsevier, v. 30, p. 81–94, 2018.

RAWLS, C. G.; TURNQUIST, M. A. Pre-positioning of emergency supplies for disaster response. *Transportation research part B: Methodological*, Elsevier, v. 44, n. 4, p. 521–534, 2010.

RENKLI, Ç.; DURAN, S. Pre-positioning disaster response facilities and relief items. *Human and Ecological Risk Assessment: An International Journal*, Taylor & Francis, v. 21, n. 5, p. 1169–1185, 2015.

ROSTAMI-TABAR, B. et al. Forecasting for social good. *International Journal of Forecasting*, Elsevier, v. 38, n. 3, p. 1245–1257, 2022.

ROUSSEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53–65, 1987. ISSN 0377-0427. Disponível em: <https://www.sciencedirect.com/science/article/pii/0377042787901257>.

RUI, S.; SHIWEI, H.; ZHANG, L. Optimum transit operations during the emergency evacuations. *Journal of Transportation Systems Engineering and Information Technology*, Elsevier, v. 9, n. 6, p. 154–160, 2009.

SEADE. *Índice Paulista de Vulnerabilidade Social*. 2010. Disponível em: <https://http://ipvs.seade.gov.br/>.

SEYEDAN, M.; MAFAKHERI, F. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*, SpringerOpen, v. 7, n. 1, p. 1–22, 2020.

SHANKAR, V.; MANNERING, F.; BARFIELD, W. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention*, Elsevier, v. 27, n. 3, p. 371–389, 1995.

SHAO, J. et al. Relief demand forecasting based on intuitionistic fuzzy case-based reasoning. *Socio-Economic Planning Sciences*, Elsevier, p. 100932, 2020.

SHERALI, H. D.; CARTER, T. B.; HOBEIKA, A. G. A location-allocation model and algorithm for evacuation planning under hurricane/flood conditions. *Transportation Research Part B: Methodological*, Elsevier, v. 25, n. 6, p. 439–452, 1991.

SOUZA, C. et al. *Desastres Naturais: conhecer para prevenir*. [S.l.]: Imprensa Oficial do Estado de São Paulo, 2009.

TALL, A. et al. Using seasonal climate forecasts to guide disaster management: the red cross experience during the 2008 west africa floods. *International Journal of Geophysics*, Hindawi, v. 2012, 2012.

TASKIN, S.; JR, E. J. L. Inventory decisions for emergency supplies based on hurricane count predictions. *International Journal of Production Economics*, Elsevier, v. 126, n. 1, p. 66–75, 2010.

VENABLES, W. N.; DICHMONT, C. M. Glms, gams and glmms: an overview of theory for applications in fisheries research. *Fisheries research*, Elsevier, v. 70, n. 2-3, p. 319–337, 2004.

- VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 307–333, 1989.
- WANG, G. et al. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, Elsevier, v. 176, p. 98–110, 2016.
- WANG, X.-L.; WU, X.-L.; SUN, B.-Y. Factor selection and regression for forecasting relief food demand. In: IEEE. *2012 8th International Conference on Natural Computation*. [S.l.], 2012. p. 226–228.
- WASSENHOVE, L. N. V. Humanitarian aid logistics: supply chain management in high gear. *Journal of the Operational research Society*, Taylor & Francis, v. 57, n. 5, p. 475–489, 2006.
- WASSENHOVE, L. N. V.; MARTINEZ, A. J. P. Using or to adapt supply chain management best practices to humanitarian logistics. *International Transactions in Operational Research*, Wiley Online Library, v. 19, n. 1-2, p. 307–322, 2012.
- WESTGATE, K. N. Some definitions of disaster. In: *Occasional Paper*. [S.l.]: Inglaterra. University of Bradford. Disaster Research Unit, 1976. v. 4.
- WILHELM, W.; SRINIVASA, A. V. A strategic, area-wide contingency planning model for oil spill cleanup operations with application demonstrated to the galveston bay area. *Decision Sciences*, Wiley Online Library, v. 27, n. 4, p. 767–799, 1996.
- WOOLDRIDGE, J. M. *Econometric analysis of cross section and panel data*. [S.l.]: MIT press, 2010.
- XING, H.; ZHONGLIN, Z.; SHAOYU, W. The prediction model of earthquake casualty based on robust wavelet v-svm. *Natural Hazards*, Springer, v. 77, n. 2, p. 717–732, 2015.
- XU, X.; QI, Y.; HUA, Z. Forecasting demand of commodities after natural disasters. *Expert Systems with Applications*, Elsevier, v. 37, n. 6, p. 4313–4317, 2010.
- YAACOB, W. F. W.; LAZIM, M. A.; WAH, Y. B. A practical approach in modelling count data. In: *Proceedings of the Regional Conference on Statistical Sciences. Malaysia*. [S.l.: s.n.], 2010. p. 176–183.
- ZHU, X.; ZHANG, G.; SUN, B. A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence. *Natural Hazards*, Springer, v. 97, n. 1, p. 65–82, 2019.

**APÊNDICE A – INFORMAÇÕES DA
DEMANDA E DOS
FATORES PREDITORES**

ID	clusters	demanda_kg	pop_estimada	vulnerabilidade	anomalia_climatica
ADAMANTINA	b	0	35153	alta	nao
ADOLFO	b	0	3545	media	nao
AGUAI	b	0	36981	alta	nao
AGUASDAPRATA	b	0	8262	media	nao
AGUASDELINDOIA	b	0	18908	alta	nao
AGUASDESANTABARBARA	b	0	6142	media	nao
AGUASDESAOPEDRO	b	0	3588	media	nao
AGUDOS	c	67,25	37582	alta	nao
ALAMBARI	b	0	6231	alta	nao
ALFREDOMARCONDES	b	0	4201	media	nao
ALTAIR	b	0	4211	alta	nao
ALTINOPOLIS	b	0	16221	alta	nao
ALTOALEGRE	b	0	4078	media	nao
ALUMINIO	b	0	18903	media	nao
ALVARESFLORENCE	b	0	3616	media	nao
ALVARESMACHADO	b	0	25078	alta	nao
ALVARODECARVALHO	b	0	5320	alta	nao
ALVINLANDIA	c	212	3251	media	nao
AMERICANA	b	9,25	244370	alta	nao
AMERICOBRASELENSE	b	0	41545	alta	nao
AMERICODECAMPOS	b	0	5993	media	nao
AMPARO	b	0	73145	alta	nao
ANALANDIA	b	0	5115	alta	nao
ANDRADINA	b	0	57245	alta	nao
ANGATUBA	b	0	25724	alta	nao
ANHEMBI	b	0	6911	alta	nao
ANHUMAS	b	0	4172	media	nao
APARECIDA	b	0	36211	alta	nao
APARECIDAOESTE	b	0	4122	alta	nao
APIAI	c	550,25	24081	alta	nao
ARACARIGUAMA	c	1092	23343	alta	nao
ARACATUBA	b	0	199210	alta	nao
ARACOIABADASERRA	c	317,5	35389	alta	nao
ARAMINA	b	0	5689	media	nao
ARANDU	b	0	6373	alta	nao
ARAPEI	b	0	2452	alta	nao
ARARAQUARA	c	226,25	240542	alta	nao
ARARAS	b	0	136739	alta	nao
ARCOIRIS	c	971,75	1755	alta	nao
AREALVA	b	0	8665	alta	nao
AREIAS	b	0	3906	alta	nao
AREIOPOLIS	b	0	11186	alta	nao
ARIRANHA	b	0	9851	media	nao
ARTURNOGUEIRA	b	0	56247	alta	nao
ARUJA	b	0	92453	alta	nao
ASPASIA	b	0	1815	media	nao
ASSIS	b	0	105768	alta	nao
ATIBAIA	b	0	145378	alta	nao
AURIFLAMA	b	0	15316	media	nao
AVAI	b	0	5467	alta	nao
AVANHANDAVA	b	0	14063	alta	nao
AVARE	b	0	91792	alta	nao
BADYBASSITT	b	0	18013	media	nao
BALBINOS	b	0	6127	alta	nao
BALSAMO	b	0	9209	alta	nao
BANANAL	b	0	11039	alta	nao
BARAODEANTONINA	b	0	3525	alta	nao
BARBOSA	b	0	7532	alta	nao

BARIRI	b	0	35844	alta	nao
BARRABONITA	b	0	36125	alta	nao
BARRADOCHAPEU	b	49,5	5794	alta	nao
BARRADOTURVO	c	738,5	7606	alta	nao
BARRETOS	b	0	123546	alta	nao
BARRINHA	b	0	33537	alta	nao
BARUERI	b	0	279704	muitoalta	nao
BASTOS	b	0	20952	alta	nao
BATATAIS	b	0	63438	alta	nao
BAURU	c	608,75	381706	muitoalta	nao
BEBEDOURO	b	0	77612	alta	nao
BENTODEABREU	b	0	3028	alta	nao
BERNARDINODECAMPOS	b	0	11168	media	nao
BERTIOGA	b	0	66154	muitoalta	nao
BILAC	b	0	8197	media	nao
BIRIGUI	b	0	126094	alta	nao
BIRITIBAMIRIM	b	0	33265	alta	nao
BOAESPERANCADOSUL	b	0	15111	alta	nao
BOCAINA	b	0	12571	media	nao
BOFETE	c	507,25	12107	alta	nao
BOITUVA	c	891	63310	alta	nao
BOMJESUSDOSPERDOES	c	688,5	26506	alta	nao
BOMSUCESSEDEITARARE	b	0	4013	alta	nao
BORA	b	0	839	media	nao
BORACEIA	b	0	4913	alta	nao
BORBOREMA	b	0	16278	alta	nao
BOREBI	b	0	2713	alta	nao
BOTUCATU	b	0	149718	alta	nao
BRAGANCAPAULISTA	b	0	172346	alta	nao
BRAUNA	b	0	5795	alta	nao
BREJOALEGRE	b	0	2911	alta	nao
BRODOWSKI	b	0	25605	alta	nao
BROTAS	b	0	24862	alta	nao
BURI	b	0	20050	alta	nao
BURITAMA	b	0	17414	alta	nao
BURITIZAL	b	0	4547	alta	nao
CABRALIAPAULISTA	b	0	4222	alta	nao
CABREUVA	c	531	51130	alta	nao
CACAPAVA	c	114,25	95752	alta	nao
CACHOEIRAPAULISTA	b	0	33827	alta	nao
CACONDE	c	210,5	19031	alta	nao
CAFELANDIA	b	0	17917	alta	nao
CAIABU	b	0	4195	alta	nao
CAIEIRAS	b	0	104044	alta	nao
CAIUA	b	0	6017	alta	nao
CAJAMAR	c	492,5	79034	muitoalta	nao
CAJATI	b	64,25	28441	alta	nao
CAJOBI	b	0	10649	alta	nao
CAJURU	b	0	26613	alta	nao
CAMPINADOMONTEALEGRE	b	0	6088	alta	nao
CAMPINAS	b	0	1223237	muitoalta	nao
CAMPOLIMPOPAULISTA	c	184	86407	alta	nao
CAMPOSDOJORDAO	b	8,25	52713	alta	nao
CAMPOSNOVOSPAULISTA	b	0	5028	alta	nao
CANANEIA	b	64,25	12542	alta	nao
CANAS	b	0	5268	alta	nao
CANDIDOMOTA	b	0	31410	alta	nao
CANDIDORODRIGUES	b	0	2805	media	nao
CANITAR	c	353,75	5365	alta	nao

ADAMANTINA	b	0	35153	alta	sim
ADOLFO	b	0	3545	media	sim
AGUAI	b	0	36981	alta	sim
AGUASDAPRATA	b	1385	8262	media	sim
AGUASDELINDOIA	b	0	18908	alta	sim
AGUASDESANTABARBARA	b	0	6142	media	sim
AGUASDESAOPEURO	b	0	3588	media	sim
AGUDOS	c	0	37582	alta	sim
ALAMBARI	b	0	6231	alta	sim
ALFREDOMARCONDES	b	0	4201	media	sim
ALTAIR	b	0	4211	alta	sim
ALTINOPOLIS	b	0	16221	alta	sim
ALTOALEGRE	b	0	4078	media	sim
ALUMINIO	b	0	18903	media	sim
ALVARESFLORENCE	b	0	3616	media	sim
ALVARESMACHADO	b	0	25078	alta	sim
ALVARODECARVALHO	b	0	5320	alta	sim
ALVINLANDIA	c	0	3251	media	sim
AMERICANA	b	812	244370	alta	sim
AMERICOBRASELENSE	b	0	41545	alta	sim
AMERICODECAMPOS	b	0	5993	media	sim
AMPARO	b	0	73145	alta	sim
ANALANDIA	b	0	5115	alta	sim
ANDRADINA	b	0	57245	alta	sim
ANGATUBA	b	0	25724	alta	sim
ANHEMBI	b	0	6911	alta	sim
ANHUMAS	b	0	4172	media	sim
APARECIDA	b	0	36211	alta	sim
APARECIDAOESTE	b	0	4122	alta	sim
APIAI	c	0	24081	alta	sim
ARACARIGUAMA	c	3092	23343	alta	sim
ARACATUBA	b	0	199210	alta	sim
ARACOIABADASERRA	c	0	35389	alta	sim
ARAMINA	b	0	5689	media	sim
ARANDU	b	0	6373	alta	sim
ARAPEI	b	0	2452	alta	sim
ARARAQUARA	c	0	240542	alta	sim
ARARAS	b	0	136739	alta	sim
ARCOIRIS	c	0	1755	alta	sim
AREALVA	b	0	8665	alta	sim
AREIAS	b	0	3906	alta	sim
AREIOPOLIS	b	0	11186	alta	sim
ARIRANHA	b	0	9851	media	sim
ARTURNOGUEIRA	b	0	56247	alta	sim
ARUJA	b	0	92453	alta	sim
ASPASIA	b	0	1815	media	sim
ASSIS	b	0	105768	alta	sim
ATIBAIA	b	22847	145378	alta	sim
AURIFLAMA	b	0	15316	media	sim
AVAI	b	0	5467	alta	sim
AVANHANDAVA	b	0	14063	alta	sim
AVARE	b	0	91792	alta	sim
BADYBASSITT	b	0	18013	media	sim
BALBINOS	b	0	6127	alta	sim
BALSAMO	b	0	9209	alta	sim
BANANAL	b	0	11039	alta	sim
BARAODEANTONINA	b	0	3525	alta	sim
BARBOSA	b	0	7532	alta	sim
BARIRI	b	0	35844	alta	sim

BARRABONITA	b	0	36125 alta	sim
BARRADOCHAPEU	b	0	5794 alta	sim
BARRADOTURVO	c	0	7606 alta	sim
BARRETOS	b	0	123546 alta	sim
BARRINHA	b	0	33537 alta	sim
BARUERI	b	0	279704 muitoalta	sim
BASTOS	b	0	20952 alta	sim
BATATAIS	b	0	63438 alta	sim
BAURU	c	0	381706 muitoalta	sim
BEBEDOURO	b	0	77612 alta	sim
BENTODEABREU	b	0	3028 alta	sim
BERNARDINODECAMPOS	b	0	11168 media	sim
BERTIOGA	b	0	66154 muitoalta	sim
BILAC	b	0	8197 media	sim
BIRIGUI	b	0	126094 alta	sim
BIRITIBAMIRIM	b	0	33265 alta	sim
BOAESPERANCADOSUL	b	322	15111 alta	sim
BOCAINA	b	0	12571 media	sim
BOFETE	c	0	12107 alta	sim
BOITUVA	c	3441	63310 alta	sim
BOMJESUSDOSPERDOES	c	6595	26506 alta	sim
BOMSUCESSEDEITARARE	b	0	4013 alta	sim
BORA	b	0	839 media	sim
BORACEIA	b	0	4913 alta	sim
BORBOREMA	b	0	16278 alta	sim
BOREBI	b	1266	2713 alta	sim
BOTUCATU	b	0	149718 alta	sim
BRAGANCAPAULISTA	b	0	172346 alta	sim
BRAUNA	b	0	5795 alta	sim
BREJOALEGRE	b	0	2911 alta	sim
BRODOWSKI	b	0	25605 alta	sim
BROTAS	b	0	24862 alta	sim
BURI	b	0	20050 alta	sim
BURITAMA	b	0	17414 alta	sim
BURITIZAL	b	0	4547 alta	sim
CABRALIAPAULISTA	b	0	4222 alta	sim
CABREUVA	c	11063	51130 alta	sim
CACAPAVA	c	0	95752 alta	sim
CACHOEIRAPAULISTA	b	0	33827 alta	sim
CACONDE	c	0	19031 alta	sim
CAFELANDIA	b	0	17917 alta	sim
CAIABU	b	0	4195 alta	sim
CAIEIRAS	b	3253	104044 alta	sim
CAIUA	b	0	6017 alta	sim
CAJAMAR	c	5521	79034 muitoalta	sim
CAJATI	b	156	28441 alta	sim
CAJOBI	b	0	10649 alta	sim
CAJURU	b	0	26613 alta	sim
CAMPINADOMONTEALEGRE	b	0	6088 alta	sim
CAMPINAS	b	0	1223237 muitoalta	sim
CAMPOLIMPOPAULISTA	c	6581	86407 alta	sim
CAMPOSDOJORDAO	b	0	52713 alta	sim
CAMPOSNOVOSPAULISTA	b	0	5028 alta	sim
CANANEIA	b	2459	12542 alta	sim
CANAS	b	1180	5268 alta	sim
CANDIDOMOTA	b	0	31410 alta	sim
CANDIDORODRIGUES	b	0	2805 media	sim
CANITAR	c	0	5365 alta	sim

ANEXO A – MODELO MATEMÁTICO DE OTIMIZAÇÃO DO PROBLEMA DE LOCALIZAÇÃO COM MÚLTIPLOS CRITÉRIOS

Decisões de localização de instalações de armazenagem de suprimentos de socorro influenciam o desempenho de operações humanitárias. O tempo de resposta e os custos incorridos são influenciados diretamente pelo número de localizações dos depósitos, além da quantidade de estoques de suprimentos de socorro mantidos.

Existem diversos modelos de otimização para problemas de localização. Ballou (2009) classifica problemas de localização conforme cinco critérios: Força direcionadora, número de instalações, descontinuidade das escolhas, agregação dos dados e horizonte de tempo. A força direcionadora está relacionada ao objetivo do modelo, em geral dominam os fatores econômicos. A classificação segundo o número de instalações diferencia problemas de localização de apenas uma instalação dos problemas que consideram múltiplas instalações; localizar uma única instalação simplifica no tratamento de forças competitivas, a divisão da demanda, e os custos de instalação. A descontinuidade das escolhas divide os problemas de acordo com a localização das instalações no espaço de modelagem, podendo ser contínuos ou discretos. O critério de classificação segundo o grau de agregação dos dados apresenta problemas de localização com alta agregação dos dados, isso resulta em métodos cuja precisão limita as localizações a amplas áreas geográficas. Por outro lado, os problemas com baixa agregação de informação podem diferenciar entre locais separados apenas em uma rua. Por fim, a classificação de acordo ao horizonte de tempo adotado pelo problema, que pode ser de período único (estático) ou múltiplo (dinâmico).

Daskin (1995) apresenta uma taxonomia de modelos de localização baseada na configuração espacial das demandas; classifica os modelos em analíticos, contínuos, em rede e

discretos. Modelos analíticos assumem que as demandas são distribuídas em alguma maneira sobre o espaço, geralmente uniformemente. Apesar, do nome os modelos contínuos consideram que as demandas são representadas em pontos discretos em qualquer lugar no plano. Modelos de localização de rede tratam demandas e instalações por uma rede composta de nós e links. Finalmente, os modelos discretos que recebem as localizações ou coordenadas dos nós de demanda e as localizações candidatas que são restringidas ao um número finito. Além dessas categorias, Daskin (1995) propõe classificação a problemas de localização baseadas na métrica de distância, na quantidade de instalações, na variedade de produtos, no número de objetivos, a capacidade da instalação, a natureza dos dados de entrada, bem como uma variedade de outros critérios de classificação.

Artigos que trabalharam com modelos de localização de instalações no planejamento pré-desastre foram divididos de acordo com o tipo dos dados de entrada, que podem ser determinísticos (certos) ou estocásticos (sujeitos a incertezas), e ao número de objetivos. A Tabela 9 apresenta os modelos de localização de instalações com base no tipo de dados e objetivos.

Tabela 9: Modelos de localização de instalações

	Determinístico	Estocástico
Único objetivo	Sherali, Carter e Hobeika (1991), Dessouky et al. (2006), Jia, Ordóñez e Dessouky (2007), Horner e Downs (2010), Pérez-Galarce et al. (2017), Cotes e Cantillo (2019)	Wilhelm e Srinivasa (1996), Chang, Tseng e Chen (2007), Balcik e Beamon (2008), Rui, Shiwei e Zhang (2009), Duran, Gutierrez e Keskinocak (2011), Renkli e Duran (2015), Kinay et al. (2018), Monzón, Liberatore e Vitoriano (2020)
Múltiplo objetivo	Kongsomsaksakul, Yang e Chen (2005)	Belardo et al. (1984), Mete e Zabinsky (2010), Rawls e Turnquist (2010)

Fonte: Adaptado de Caunhye, Nie e Pokharel (2012)

A maioria dos modelos de localização de instalações foram formulados como problemas de programação inteira mista com variáveis de localização binárias. Todos os trabalhos pesquisados que focam no planejamento pré-desastre, desenvolveram modelos estáticos ou uni periódicos. Além disso, são poucos os modelos que adotam padrões multi-objetivos,

a maioria dos trabalhos são baseados em um único objetivo (CAUNHYE; NIE; POKHAREL, 2012; HEZAM et al., 2021).

A.1 Formulação do modelo matemática

O objetivo do modelo é determinar as melhores opções de localização de instalações para armazenamento de suprimentos de socorro voltados ao atendimento à população atingida por desastres. Por conseguinte é proposto um modelo multiobjetivo, através do qual busca-se minimizar os custos totais e maximizar o nível de serviço. O segundo objetivo é capturado por meio da minimização do tempo máximo de atendimento. O modelo desenvolvido determina um conjunto de soluções ótimas, as quais serão avaliadas e selecionadas pelo tomador de decisão.

O modelo proposto é apresentado em detalhe a continuação.

Conjuntos:

i Municípios $I = \{1, 2, \dots, 645\}$

j Locais candidatos a receber depósitos $J = \{1, 2, \dots, 645\}$

Parâmetros:

k Peso atribuído ao objetivo de nível de serviço

a Número de instalações para ser localizadas

f_j Custo de instalação e operação do local candidato j [$R\$/ano$]

t_{ij} Custo de transporte entre o município i e local candidato j [$R\$/ano$]

h_i Demanda de suprimentos de socorro do município i [kg/ano]

d_{ij} Distância entre o município i e local candidato j [km]

\hat{d}_{ij} Índice de não cobertura

$$\hat{d}_{ij} = \begin{cases} 0, & \text{se } d_{ij} \leq D_{max} \\ 1, & \text{caso contrário} \end{cases}$$

Variáveis de decisão:

$$X_j = \begin{cases} 1, & \text{se abre depósito no local } j \\ 0, & \text{caso contrário} \end{cases}$$

$$Y_{ij} = \begin{cases} 1, & \text{se e atribuído o município } i \text{ ao depósito no local } j \\ 0, & \text{caso contrário} \end{cases}$$

Função objetivo:

$$\text{Min} \sum_{j \in J} f_j X_j + \sum_{i \in I} \sum_{j \in J} h_i (k \hat{d}_{ij} + t_{ij}) Y_{ij} \quad (\text{A.1})$$

Restrições:

$$\sum_{j \in J} Y_{ij} = 1 \quad \forall i \in I \quad (\text{A.2})$$

$$Y_{ij} - X_j \leq 0 \quad \forall i \in I \quad \forall j \in J \quad (\text{A.3})$$

$$\sum_{j \in J} X_j \leq a \quad (\text{A.4})$$

$$X_j \in 0, 1 \quad \forall j \in J \quad (\text{A.5})$$

$$Y_{ij} \in 0, 1 \quad \forall i \in I \quad \forall j \in J \quad (\text{A.6})$$

A função objetivo (A.1) está composta por dois objetivos. O primeiro objetivo é minimizar o custo ponderado pela demanda total entre o nó de demanda e o depósito mais próximo (objetivo de P mediana), o segundo objetivo é maximizar o número de demandas atendidas no tempo menor ao especificado (objetivo de cobertura máxima). A restrição A.2 garante que cada município $i \in I$ seja atribuído a exatamente um depósito $j \in J$. A restrição A.3 ligam as variáveis de localização X_j e as variáveis de alocação Y_{ij} . Assim asseguram que as demandas do nó $i \in I$ sejam atribuídas ao depósito no local $j \in J$ só se for aberto. A restrição A.4 afirma que exatamente a depósitos devem ser abertos. Por fim, as restrições A.5 e A.6 determinam os espaços de soluções das variáveis de decisão.

A.2 Determinação dos parâmetros

São apresentados os parâmetros considerados na modelagem matemática, com exceção da demanda.

Locais Candidatos

A lista de candidatos para o problema é composta por os 645 municípios do Estado de São Paulo. No caso do Sistema de Apoio à Decisão (SAD) o número de locais candidatos foi limitado a 15, os quais são selecionados pelo usuário.

Quantidade de depósitos a serem abertos

O modelo de localização de instalações humanitárias proposto, pode localizar uma quantidade de depósitos previamente estabelecida, oscilando entre um e o total de locais candidatos a depósitos.

Distância

Para a obtenção das distâncias entre os locais candidatos e os nós de demanda, utilizou-se a API (Interface de Programação de Aplicativos) “openrouteservice”. As consultas foram efetuadas com a geolocalização dos municípios e referentes à distância rodoviária.

Custo de instalação e operação do depósito

Os custos de instalação e operação de um depósito foram estimados com base na necessidade de mão de obra e na aquisição de equipamentos de armazenagem para as instalações, critérios propostos por Brito (2015) e avaliados pelo CEPDEC-SP. Os equipamentos de armazenagem e movimentação foram estimados considerando uma capacidade de 120 posições palete para cada depósito, além do equipamento de movimentação e pesagem. Os valores de custo de instalação equivale a 40 \$ por posição, com base em Frazelle e Sojo (2006).

Custo de Transporte

Os custos de transporte de suprimentos de socorro entre o depósito ao ponto de demanda, foram obtidos de tabelas de frete das entidades Associação Nacional do Transporte de Cargas e Logística (NTC&Logística) e Transporte de Cargas Rodoviárias (TRC), para carga fracionada e carga lotação, correspondentemente.

ANEXO B – SISTEMA DE APOIO À DECISÃO

B.1 Conceito

O Sistema de Apoio à Decisão (SAD), *LocDC*, foi desenvolvido para auxiliar a CEPDEC-SP do Estado de São Paulo na localização de depósitos de suprimentos de alívio através de Visualização Analítica e Modelagem Matemática Interativa. Sendo assim, buscou-se com que profissionais da CEPDEC-SP interajam de forma produtiva e intuitiva com modelos matemáticos de otimização.

O SAD combina técnicas de visualização interativas e análises automatizadas que auxiliam na compreensão de informações e no entendimento do problema, assim o tomador de decisão pode proporcionar soluções mais acertadas.

O caso abordado no sistema é um exemplo de problema de localização de instalações para armazenamento de suprimentos de alívio para resposta a desastres no Estado de São Paulo. O objetivo é encontrar o número e os locais dos depósitos que otimizem custos totais (fixos e variáveis) e nível de serviço (tempo de entrega) aos municípios vítimas de desastres. Os locais candidatos a ter um depósito são pré-selecionados pelo usuário dentre as 53 cidades principais das regiões geográficas do Estado de São Paulo; na mesma região geográfica pode-se substituir a cidade principal por uma outra. O nível de serviço é relacionado com o tempo máximo de atendimento, podendo ser personalizado num período de 3h a 10h, segundo critério do tomador de decisão.

B.2 Desenvolvimento

O SAD possui interface visual interativa que permite ao tomador de decisão personalizar os locais de instalação dos depósitos e rapidamente verificar os resultados.

Para a elaboração dos elementos de interface do sistema foram utilizadas as linguagens de programação Javascript e HTML. Além disso, foram utilizadas a linguagem de programação Python 3.7 e o framework web Flask. Já para a formulação do modelo de otimização foi usada a biblioteca Pyomo de Python, e para a resolução do problema foi selecionado o solver de código aberto CBC.

O sistema deve ser aberto com um navegador de internet. Na página inicial estão contidas informações sobre o projeto de pesquisa a partir do qual o software foi idealizado. Na aba Modelo o usuário é levado à página web interativa, onde é possível manipular alguns dados e as informações que são mostradas. Feitas as alterações desejadas nos dados, é possível otimizar.

Após o usuário processar o modelo, é gerada uma tabela e um gráfico, que apresentam as diversas soluções de Pareto, as quais facilitam a comparação, avaliação e seleção das alternativas de solução.

B.3 Estrutura

Ao abrir o sistema, na parte superior aparecem três abas: “CISLOG”, “Explicações” e “Modelo interativo”. A aba inicial “Explicações” possui informações sobre logística humanitária e contextualização do Estado de São Paulo, é apresentada a seguir na Figura 14. Após clicar na aba “CISLOG” o usuário é encaminhado ao site oficial do Centro de Inovação em Sistemas Logísticos.

Figura 14: Aba inicial “Explicações” do SAD



Fonte: Elaborado pelo autor

A aba de “Modelo interativo” é mostrada na Figura 15, que contém a entrada de dados ao modelo de localização de depósitos. Para explicar todas as configurações da

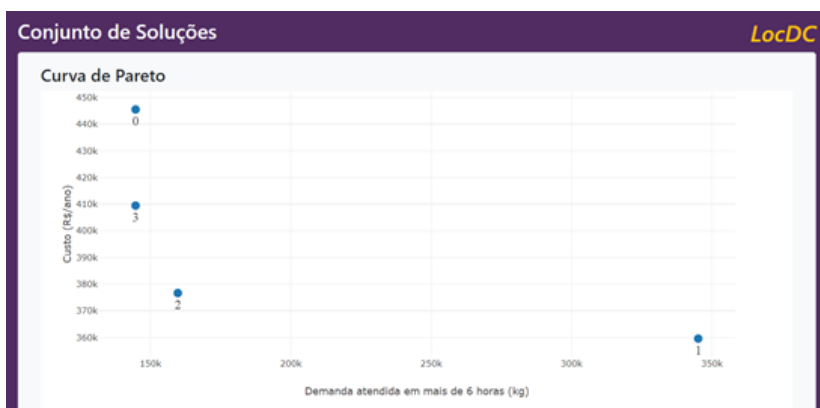
entrada de dados, as seções podem ser identificadas pela cor do quadro em volta de cada item. Foram destacadas quatro seções principais: locais candidatos (em verde), locais obrigatórios (em laranja), número de depósitos (em amarelo) e tempo de atendimento (em azul). O botão escrito “Otimizar”, na parte inferior da tela, inicia o processamento para resolver o modelo matemático após seu *setup*.

Figura 15: Aba “Modelo Iterativo” do SAD

Fonte: Elaborado pelo autor

Considerando que o custo e o tempo de atendimento são objetivos conflitantes, a saída de dados corresponde a um conjunto de soluções ótimas. O número de soluções ótimas varia de acordo aos dados de entrada. De forma a facilitar o entendimento dos resultados são apresentados, inicialmente, no formato de gráfico e tabela. O gráfico “Curva de Pareto” é apresentado na Figura 16.

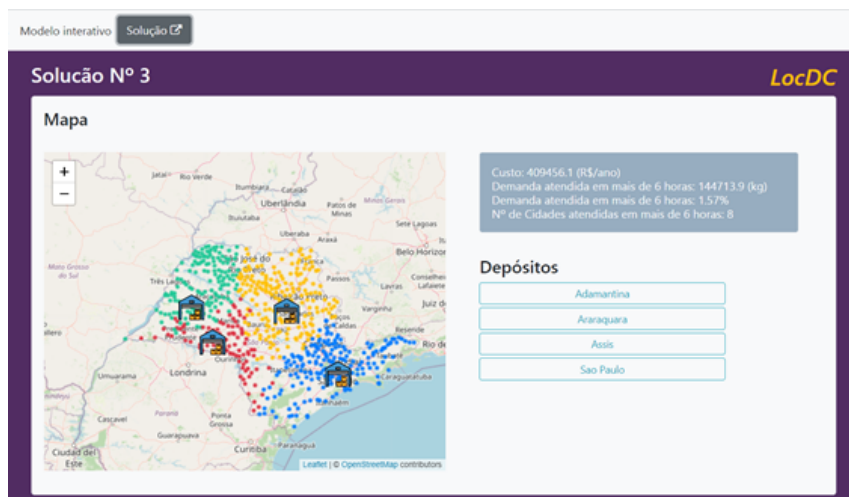
Figura 16: Gráfico de Curva de Pareto



Fonte: Elaborado pelo autor

As informações de uma solução específica são apresentadas na Figura 17, o mapa apresenta a localização dos depósitos abertos com os marcadores coloridos que referem se às cidades atribuídas a cada depósito, o box apresenta informação de custo e demanda atendida fora do tempo desejado.

Figura 17: Informação de uma solução específica



Fonte: Elaborado pelo autor

Um video manual com mais informações da ferramenta SAD, *LocDC*, pode-se encontrado no seguinte link <https://www.youtube.com/watch?v=fi0EHGB1DUA>.

A comunicação de criação foi aprovada pela Agência USP de inovação (AUSPIN) para proteção do sistema *LocDC*.