

RENATA AKEMI MARÇAL IMAI

**Análise do impacto do agrupamento da demanda na qualidade da
solução de problemas de localização de instalações**

SÃO PAULO

2022

RENATA AKEMI MARÇAL IMAI

**Análise do impacto do agrupamento da demanda na qualidade da
solução de problemas de localização de instalações**

Versão Corrigida

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

Área de concentração: Engenharia de Transportes

Orientador: Prof. Dr. Claudio Barbieri da Cunha

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, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Imai, Renata Akemi Marçal

Análise do impacto do agrupamento da demanda na qualidade da solução de problemas de localização de instalações / R. A. M. Imai -- versão corr. São Paulo, 2022.

138 p.

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

1.Agrupamento da demanda 2.Problema de localização 3.Modelagem
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II.t.

DEDICATÓRIA

*Gerson, finalmente posso te dedicar
mais que as 150 palavras de um bilhete.
Mamãe, que o coração se lembre um dia
de tudo que a memória esquecer.*

AGRADECIMENTOS

Aos meus pais, Arminda e Renato, por viabilizarem minha saída de casa para vir à São Paulo continuar meus estudos. A Laura e Bia, minhas irmãs, por todo carinho, risadas e companheirismo. E a Dona Thereza, minha avó, por todo amor e atenção.

Aos meus orientadores, Prof. Dr. Claudio Barbieri da Cunha e Dr. Cauê Sauter Guazzelli, por toda a paciência, confiança e orientação dos últimos anos. Muito obrigada por acreditarem em mim e viabilizarem esta pesquisa.

A Gerson, companheiro que (re)encontrei nessa cidade e decidi caminhar comigo. Obrigada por apoiar a conclusão dessa pesquisa e me motivar diariamente.

Aos amigos de mestrado, Gabriel Caldeira e Lucas Melo, pelas discussões, trabalhos em conjuntos e cafés ao longo desses anos. Aos amigos de longos anos Ana Clara, Camilla, Fernando, Lara, Letícia, Lucas Coutinho e Victor Vieira, pelo incentivo, carinho e apoio para a conclusão do mestrado. À Maria Josefina, minha terapeuta, por não me deixar desistir desse projeto. Aos amigos (se é que posso chamá-los assim) que São Paulo me presenteou: Ana Luiza dos Santos, Igor Florian e Leonardo Tegen.

Ao Conselho Nacional de Pesquisa e Tecnologia (CNPq), pelo apoio financeiro a esta pesquisa.

Nascer sabendo é uma limitação porque obriga a apenas repetir e, nunca, a criar, inovar, refazer, modificar. Quanto mais se nasce pronto, mais se é refém de que já se sabe e, portanto, do passado; aprender sempre é o que mais impede que nos tornemos prisioneiros de situações que, por serem inéditas, não saberíamos enfrentar. (CORTELLA, 2006).

RESUMO

Este trabalho trata da análise do impacto do agrupamento de pontos de demanda na qualidade da solução de problemas de localização. Foram desenvolvidos dois conjuntos de dados com localizações reais de farmácias no Brasil, a partir dos quais 18 instâncias do problema de localização de instalações com custo fixo foram criadas. São propostos dois métodos de agrupamentos de pontos de demanda: um que considera os limites geográficos do município e outro que emprega o algoritmo *K-Means*. Constatou-se que as soluções dos problemas com demanda agregada por município apresentam diferenças de 0,43%, enquanto os problemas com demanda agregada por *K-Means* apresentam diferenças de no máximo 0,03% com relação ao problema desagregado. Para identificar os municípios nos quais a alocação da demanda pudesse ser diferente entre os modelos agregados e o desagregado, utilizou-se o algoritmo DBSCAN – *Density Based Spatial Clustering of Applications with Noise*. Para analisar se o DBSCAN identifica adequadamente essas regiões, foram selecionadas quatro instâncias, sendo duas de cada conjunto. Em 3 das 4 instâncias analisadas, observou-se que o DBSCAN identificou as regiões do espaço nas quais a alocação de demanda é diferente entre modelos agregados e desagregados. Nestas mesmas instâncias, também se observou que as diferenças nos custos de transporte variam entre -84,99% e 662,90% no agrupamento por município e -23,25% e 110,31% no agrupamento por *K-Means*.

Palavras-chave: agrupamento da demanda; problema de localização; modelagem

ABSTRACT

This thesis analyses the impact in the quality of solution for location problems with aggregated demand points. Two datasets containing the location of drugstores in Brazil were developed, and 18 instances of the fixed cost facility location problem were created. Two aggregation methods are proposed: aggregating demand points regarding the municipality boundary or using the K-Means algorithm. The models' solutions with aggregated demand points by municipality showed a 0,43% difference in the objective function value at most, while the difference using K-Means algorithm was 0,03% at most. To identify the municipalities where differences in the allocation of unaggregated and aggregated models are different, DBSCAN (Density Based Spatial Clustering of Applications with Noise) algorithm was used. To identify whether DBSCAN properly identified these regions, four instances were selected, two from each dataset were selected. For 3 out of 4 instances, DBSCAN identified properly the location of demand points allocated to different facilities. For the same instances, the differences on transportation costs between aggregated and unaggregated problems range between -84,99% to 662,90% for demand points aggregated by municipality, and between -23,25% and 110,31% using K-Means clustering.

Keywords: demand aggregation; facility location; modelling

LISTA DE FIGURAS

| | |
|--|----|
| Figura 2.1 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo A. Adaptado de Current e Schilling (1987) | 11 |
| Figura 2.2 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo B. Adaptado de Current e Schilling (1987) | 11 |
| Figura 2.3 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo C. Adaptado de Current e Schilling (1987) | 12 |
| Figura 3.1 – Esquema metodológico | 15 |
| Figura 3.2 – Exemplo de agrupamento utilizando o método Ma1 | 19 |
| Figura 3.3 – Exemplo de aplicação do método Ma2 | 21 |
| Figura 3.4 – Alocação dos pontos de demanda com base na alocação da demanda do agrupamento..... | 23 |
| Figura 3.5 – Alocação dos pontos de demanda com base no menor custo definido pela etapa de desagregação do modelo | 23 |
| Figura 3.6 – Informações provenientes do <i>software</i> de otimização..... | 24 |
| Figura 3.7 – Exemplos de bases de dados. Fonte: Ester <i>et al.</i> (1996) | 26 |
| Figura 3.8 – Pontos ordenados conforme a distância para o terceiro vizinho mais próximo. Fonte: Rahmah e Sitanggang (2016)..... | 28 |
| Figura 3.9 – Localização das aglomerações urbanas com mais de 100.000 habitantes no estado de São Paulo (SP)..... | 29 |
| Figura 3.10 – Localização dos agrupamentos e ruídos considerando pontos de demanda no estado de São Paulo (SP) | 30 |
| Figura 4.1 – Polígono de Voronoi que delimita a região do espaço mais próxima a um ponto e sua respectiva área de influência..... | 33 |
| Figura 4.2 – Exemplo de alocação de demanda de uma grade | 33 |
| Figura 4.3 – Localização dos candidatos a instalação | 36 |
| Figura 4.4 – Pontos de demanda do Conjunto 1 | 37 |
| Figura 4.5 – Histograma da distância entre pontos de demanda e respectivos locais candidatos mais próximos para o Conjunto 1 | 39 |
| Figura 4.6 – Pontos de demanda do Conjunto 2 | 40 |
| Figura 4.7 – Histograma da distância entre pontos de demanda e locais candidatos mais próximos do Conjunto 2..... | 42 |

| | |
|--|----|
| Figura 5.1 – Localização dos baricentros dos municípios e suas respectivas demandas para os pontos de demanda do Conjunto 1..... | 46 |
| Figura 5.2 – Localização dos baricentros dos agrupamentos pelo método Ma2 e suas respectivas demandas para os pontos de demanda do Conjunto 1..... | 47 |
| Figura 5.3 – Localização do baricentro dos agrupamentos coincidentes entre os métodos Ma1 e Ma2 para o Conjunto 1..... | 49 |
| Figura 5.4 – Localização dos baricentros dos agrupamentos pelo método Ma1 e suas respectivas demandas para os pontos de demanda do Conjunto 2..... | 50 |
| Figura 5.5 – Localização dos baricentros dos agrupamentos pelo método Ma2 e suas respectivas demandas para os pontos de demanda do Conjunto 2..... | 51 |
| Figura 5.6 – Localização do baricentro dos agrupamentos coincidentes entre os métodos Ma1 e Ma2 para o Conjunto 2..... | 53 |
| Figura 5.7 – Cálculo do parâmetro Eps para o Conjunto 1..... | 62 |
| Figura 5.8 – Agrupamentos do DBSCAN para o Conjunto 1..... | 64 |
| Figura 5.9 – Exemplo de agrupamentos do DBSCAN | 65 |
| Figura 5.10 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 1-9 | 68 |
| Figura 5.11 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 1-9..... | 69 |
| Figura 5.12 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 1-9 | 70 |
| Figura 5.13 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 1-9..... | 71 |
| Figura 5.14 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 1-900..... | 73 |
| Figura 5.15 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 1-900..... | 74 |
| Figura 5.16 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 1-900..... | 75 |
| Figura 5.17 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 1-900..... | 76 |
| Figura 5.18 – Cálculo do parâmetro Eps para o Conjunto 2..... | 79 |
| Figura 5.19 – Agrupamentos do DBSCAN para o Conjunto 2..... | 81 |

| | |
|---|-----|
| Figura 5.20 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 2-9 | 84 |
| Figura 5.21 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 2-9..... | 85 |
| Figura 5.22 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 2-9 | 86 |
| Figura 5.23 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 2-9..... | 87 |
| Figura 5.24 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 2- 4500..... | 89 |
| Figura 5.25 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 2-4500..... | 90 |
| Figura 5.26 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 2- 4500..... | 91 |
| Figura 5.27 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 2-4500..... | 92 |
| Figura A.1 – Diagrama de Voronoi para oito pontos em um plano. Fonte: Aurenhammer (1991) | 105 |
| Figura A.2 – Diagramas de Voronoi obtidos utilizando (a) distância Manhattan, (b) distância euclidiana e (c) distância de Chebychev. Fonte: Burkey, Bhadury e Eiselt (2011)..... | 106 |
| Figura B.1 – Alocação da demanda para o modelo desagregado da instância 1-9..... | 108 |
| Figura B.2 – Alocação da demanda para o modelo Ma1 da instância 1-9..... | 109 |
| Figura B.3 – Alocação da demanda para o modelo Ma2 da instância 1-9..... | 110 |
| Figura B.4 – Alocação da demanda para o modelo desagregado da instância 1-90..... | 111 |
| Figura B.5 – Alocação da demanda para o modelo Ma1 da instância 1-90..... | 112 |
| Figura B.6 – Alocação da demanda para o modelo Ma2 da instância 1-90..... | 113 |
| Figura B.7 – Alocação da demanda para o modelo desagregado da instância 1-900..... | 114 |
| Figura B.8 – Alocação da demanda para o modelo Ma1 da instância 1-900..... | 115 |
| Figura B.9 – Alocação da demanda para o modelo Ma2 da instância 1-900..... | 116 |
| Figura B.10 – Alocação da demanda para os modelos desagregado e Ma2 da instância 1-1800 | 117 |
| Figura B.11 – Alocação da demanda para o modelo Ma1 da instância 1-1800..... | 118 |
| Figura B.12 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1- 4500..... | 119 |

| | |
|--|-----|
| Figura B.13 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-9000..... | 120 |
| Figura B.14 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-18000..... | 121 |
| Figura B.15 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-45000..... | 122 |
| Figura B.16 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-90000..... | 123 |
| Figura B.17 – Alocação da demanda para o modelo desagregado da instância 2-9..... | 124 |
| Figura B.18 – Alocação da demanda para o modelo Ma1 da instância 2-9..... | 125 |
| Figura B.19 – Alocação da demanda para o modelo Ma1 da instância 2-9..... | 126 |
| Figura B.20 – Alocação da demanda para os modelos desagregado e Ma2 da instância 2-90..... | 127 |
| Figura B.21 – Alocação da demanda para o modelo Ma1 da instância 2-90..... | 128 |
| Figura B.22 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-900..... | 129 |
| Figura B.23 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-1800..... | 130 |
| Figura B.24 – Alocação da demanda para o modelo desagregado da instância 2-4500..... | 131 |
| Figura B.25 – Alocação da demanda para o modelo Ma1 da instância 2-4500..... | 132 |
| Figura B.26 – Alocação da demanda para o modelo Ma2 da instância 2-4500..... | 133 |
| Figura B.27 – Alocação da demanda para os modelos desagregado e Ma2 da instância 2-9000..... | 134 |
| Figura B.28 – Alocação da demanda para o modelo Ma1 da instância 2-9000..... | 135 |
| Figura B.29 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-18000..... | 136 |
| Figura B.30 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-45000..... | 137 |
| Figura B.31 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-90000..... | 138 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 4.1 – Locais candidatos | 35 |
| Tabela 4.2 – Informações sobre os estados no Conjunto 1 | 38 |
| Tabela 4.3 – Informações sobre os estados no Conjunto 2..... | 41 |
| Tabela 4.4 – Características das instâncias solucionadas | 44 |
| Tabela 5.1 – Tamanho dos modelos do Conjunto 1..... | 48 |
| Tabela 5.2 – Tamanho dos modelos do Conjunto 2..... | 52 |
| Tabela 5.3 – Informações sobre as instâncias do Conjunto 1 | 55 |
| Tabela 5.4 – Informações sobre as instâncias do Conjunto 2 | 58 |
| Tabela 5.5 – Distâncias médias e maiores distâncias (em metros) entre os pontos de demanda e os candidatos abertos mais próximos para os modelos que representam as instâncias do Conjunto 1..... | 78 |
| Tabela 5.6 – Distâncias médias e maiores distâncias (em metros) entre os pontos de demanda e os candidatos abertos mais próximos para os modelos que representam as instâncias do Conjunto 2..... | 94 |
| Tabela B.1 – Figuras que mostram a alocação da demanda | 107 |

LISTA DE SIGLAS

| | |
|--------|---|
| FO | Função objetivo |
| FCFL | <i>Fixed Charge Facility Location Problem</i> (Problema de localização de instalações com custo fixo) |
| ONU | Organização das Nações Unidas |
| IBGE | Instituto Brasileiro de Geografia e Estatística |
| B2B | <i>Buisness-to-buisness</i> (empresas que negociam com outras empresas) |
| DBSCAN | <i>Density Based Spatial Clustering of Applications with Noise</i> (Agrupamento espacial de aplicações baseadas em densidade com ruído) |

SUMÁRIO

| | |
|---|----|
| 1. INTRODUÇÃO | 1 |
| 1.1. Justificativa e objetivo da pesquisa..... | 2 |
| 1.2. Organização do trabalho | 3 |
| 2. REVISÃO DA LITERATURA | 5 |
| 2.1. Problemas de Localização..... | 5 |
| 2.2. Métodos de agrupamento de pontos de demanda | 8 |
| 2.3. Erros decorrentes do agrupamento de pontos de demanda..... | 10 |
| 2.4. Considerações sobre a revisão de literatura | 13 |
| 3. ESCOPO E METODOLOGIA | 14 |
| 3.1. Escopo e formulação do modelo do problema de localização de instalações com custo fixo | 16 |
| 3.2. Métodos de formação de agrupamentos de pontos de demanda..... | 18 |
| 3.2.1. Método de agrupamento Ma1 – Pontos de demanda agregados por município..... | 18 |
| 3.2.2. Método de agrupamento Ma2 – Agrupamentos obtidos através do algoritmo K-Means | 19 |
| 3.3. Solução do modelo..... | 21 |
| 3.4. Cálculo da diferença e análise da qualidade da solução | 25 |
| 4. INSTÂNCIAS..... | 31 |
| 4.1. Pontos de demanda e locais candidatos | 31 |
| 4.2. Custos..... | 42 |
| 4.3. Síntese..... | 43 |
| 5. EXPERIMENTOS COMPUTACIONAIS | 45 |
| 5.1. Formação de agrupamentos para a modelagem..... | 45 |
| 5.1.1. Conjunto 1 | 45 |
| 5.1.2. Conjunto 2..... | 50 |

| | |
|--|-----|
| 5.2. Resultados dos modelos de otimização aplicados ao FCFL com e sem agregação da demanda | 54 |
| 5.2.1. Conjunto 1 | 54 |
| 5.2.2. Conjunto 2 | 57 |
| 5.3. Análise comparativa dos métodos de agregação propostos | 60 |
| 5.3.1. Conjunto 1 | 61 |
| Instância 1-9..... | 65 |
| Instância 1-900..... | 72 |
| 5.3.2. Conjunto 2 78 | |
| Instância 2-9..... | 81 |
| Instância 2-4500..... | 88 |
| 6. CONSIDERAÇÕES FINAIS..... | 95 |
| REFERÊNCIAS..... | 97 |
| APÊNDICE A – DIAGRAMAS DE VORONOI..... | 105 |
| APÊNDICE B – ALOCAÇÃO DE DEMANDA | 107 |

1. INTRODUÇÃO

Nas últimas décadas, diversas mudanças econômicas em um cenário cada vez mais globalizado, tem forçado as empresas a se tornarem ainda mais competitivas. Essa necessidade que as empresas têm de sobreviver à competição tem implicações diretas no “planejamento de sistemas logísticos e de transportes em seus aspectos estratégico, tático e operacional” (GUAZZELLI, 2018).

Um dos elementos do planejamento de sistemas logísticos é o projeto da rede logística que, segundo Ballou (2006), “se trata da especificação da estrutura ao longo da qual os produtos fluem desde os pontos de origem até os centros de demanda”, além de compreender uma série de decisões dentre as quais a quantidade e a localização das instalações, quer sejam de produção, armazenamento ou distribuição, por exemplo; a capacidade necessária em cada instalação para atender a demanda dos pontos, e qual e o fluxo dos produtos a partir de uma origem.

De acordo com Ballou (2006), para determinar a melhor configuração de rede logística, é possível utilizar métodos de simulação, heurísticos ou exatos. Dentre esses, os métodos exatos são bastante utilizados uma vez que possibilitam a obtenção de soluções ótimas e, uma forma de empregá-lo é através do uso de modelos matemáticos de otimização. Segundo Arenales *et al.* (2007), os modelos matemáticos são uma abstração (representação simplificada) de problemas reais, detalhada o suficiente para captar os elementos essenciais, mas suficientemente tratável para que seja resolvido.

Alguns dos modelos matemáticos que auxiliam na determinação da configuração da rede logística, são os modelos de localização, como, por exemplo, o problema de cobertura de conjunto e o de instalações com custo fixo. Em alguns casos, os problemas de localização não possuem algoritmos polinomiais para solucioná-los, de forma que os problemas de otimização associados a esses problemas não-polinomiais recebem o nome de problemas NP-*hard* (WOLSEY, 1998). A depender do tamanho do problema de localização a ser solucionado, é possível que suas soluções sejam rapidamente encontradas, especialmente se utilizado um *software* de otimização. Porém, à medida que tanto a quantidade de pontos de demanda quanto de locais candidatos a instalação aumenta, aumentam também a quantidade de variáveis de decisão do problema, o que eleva também o tempo de processamento. Uma possibilidade para reduzir a quantidade de variáveis de decisão é através do agrupamento de pontos de demanda, feito em etapa anterior à formulação do modelo matemático. O agrupamento de pontos de

demanda possibilita que um modelo de localização se torne tratável e necessite de uma quantidade menor de dados para a obtenção de soluções.

Ao se tratar de sistemas reais de distribuição, é possível que existam grandes quantidades de pontos de demanda, representando empresas, residências ou habitantes, por exemplo, concentrados em apenas algumas regiões do espaço. Para essas localidades, uma forma usual de agrupamento geográfico por município nem sempre pode ser a melhor forma de simplificar a modelagem do problema. Por exemplo: algumas regiões mais periféricas das cidades possuem maior quantidade de habitantes que a região central, tal que a demanda seja maior dessas localidades. Possivelmente representar toda essa demanda na região central do município seja uma maneira equivocada de simplificar a modelagem do problema, o que pode levar a obtenção de soluções não ótimas, caso o problema fosse modelado de outra maneira.

1.1. Justificativa e objetivo da pesquisa

Nas últimas décadas, houve melhorias consideráveis nos computadores e algoritmos utilizados para solucionar os problemas de localização. Ainda que resultados desses problemas com dezenas ou centenas de milhares de variáveis de decisão apresentem *gaps* cada vez menores o tempo de processamento ainda é considerável, podendo ser de mais de horas, ou até mesmo dias. O agrupamento de pontos de demanda surge como uma possibilidade de modelagem que visa a redução do total de variáveis de decisão do problema.

Assim, esta pesquisa busca apresentar um método confiável de agrupamento de pontos de demanda para reduzir o tempo de processamento em decorrência da diminuição do número de variáveis de decisão. Isso possibilita que um resultado com pequena diferença no valor da função objetivo (*gap*) seja obtido em menor tempo de processamento, o que permitiria analisar algumas mudanças de cenário no problema de localização de instalações mais rapidamente.

Esta pesquisa busca responder qual a melhor maneira de agregar pontos de demanda para simplificar a modelagem de um problema de localização de instalações em locais nos quais existem grandes regiões metropolitanas e megacidades, uma vez que essas localidades concentram tanto uma grande quantidade de pontos, quanto da demanda, se associando a milhares, ou dezenas de milhares de variáveis de decisão. O que diferencia a classificação das entre grandes regiões metropolitanas e megacidades é o número de habitantes, sendo que a primeira, possui entre 2 e 10 milhões de habitantes, e a segunda mais de 10 milhões de habitantes, incluindo os municípios das cidades vizinhas que compõem a região metropolitana

ou megacidade, em decorrência do fenômeno da conurbação (KIN; VERLINDE; MACHARIS, 2017).

Dentre os objetivos específicos encontram-se:

- i. Determinar qual o impacto na qualidade dos resultados dos problemas de localização decorrentes da utilização de diferentes métodos de formação de agrupamentos, sendo eles algoritmos ou regras, especialmente em megacidades, regiões metropolitanas, localidades ou partes de áreas de estudo com maior quantidade de pontos de demanda;
- ii. Identificar características das instâncias de problemas de localização que tornam algum algoritmo ou regra de formação de agrupamento melhor para modelar o conjunto de pontos de demanda;
- iii. Identificar regiões ou áreas no espaço geográfico do escopo do estudo de localização nas quais os pontos de demanda são atendidos por instalações diferentes nas soluções dos problemas agregado e desagregado e qual a diferença no valor dos custos de transporte isso representa localmente.

Para responder aos objetivos da pesquisa, utiliza-se um problema de distribuição B2B com entrega em última milha em cidades para empresas farmacêuticas, todavia, é possível aplicar o modelo proposto a outros produtos, como eletroeletrônicos, ou até mesmo para a definição de centros de distribuição, por exemplo.

1.2. Organização do trabalho

Esta dissertação se divide em mais cinco capítulos. O Capítulo 2 apresenta a revisão de literatura dos temas mais relevantes para a realização desta pesquisa e se subdivide em três subseções, abordando os temas de (i) problemas de localização, (ii) métodos de agrupamentos de pontos de demanda, e (iii) erros decorrentes do agrupamento de pontos de demanda.

Já o Capítulo 3 descreve a metodologia empregada para análise e comparação dos resultados. Este capítulo apresenta quatro subcapítulos, que discorrem sobre (i) o escopo e formulação do problema proposto, (ii) os métodos de formação de agrupamentos utilizados na dissertação, (iii) a solução do modelo proposto, (iv) o cálculo da diferença e análise da qualidade das soluções dos modelos com pontos de demanda agregados obtidos através dos métodos propostos.

O Capítulo 4 explica como as instâncias utilizadas neste trabalho foram criadas, além de apresentar suas características.

Por sua vez, o Capítulo 5 apresenta (i) as instâncias utilizadas no trabalho e suas características, (ii) os agrupamentos obtidos para os conjuntos de dados utilizados através dos métodos de formação de agrupamentos propostos, (iii) os resultados para cada um dos modelos de otimização e (iv) a análise da qualidade dos métodos propostos.

Por fim, o Capítulo 6 discorre sobre as considerações finais e limitações do trabalho, bem como de recomendações para maior aprofundamento do tema abordado nesta pesquisa.

2. REVISÃO DA LITERATURA

Este capítulo apresenta uma revisão de literatura que contempla os principais tópicos relacionados ao objeto da pesquisa, associados ao agrupamento de pontos de demanda para simplificar a modelagem de problemas de localização de instalações. As três primeiras seções do capítulo abordam os seguintes temas:

- i. Problemas de localização
- ii. Métodos de agrupamento de pontos de demanda
- iii. Erros decorrentes do agrupamento de pontos de demanda

Já a última seção apresenta as conclusões obtidas através dos pontos mais relevantes da revisão de literatura.

2.1. Problemas de Localização

O problema de localização de instalações é um tópico consolidado dentro do campo da pesquisa operacional e que, ao longo dos anos, foi incorporada como parte do gerenciamento da cadeia de suprimentos (MELO; NICKEL; SALDANHA-DA-GAMA, 2009). Para o setor privado, os problemas de localização podem indicar a melhor localização de lojas, armazéns e centros de distribuição, por exemplo, que buscam tanto minimizar os custos de transporte, quanto o custo de abertura e operação das instalações.

Segundo Daskin (2013), existem quatro classes de problemas de localização, dentre os quais os problemas de cobertura (CHURCH; REVELLE, 1974, TOREGAS *et al.*, 1971), de centro (HAKIMI, 1964, 1965), de medianas (REVELLE; SWAIN, 1970), e de custo fixo (KRARUP; PRUZAN, 1983). Os problemas de cobertura buscam minimizar o número de instalações abertas para atender toda a demanda, enquanto os problemas de centro buscam garantir a menor distância entre os pontos de demanda e as instalações. Por sua vez, os problemas de medianas buscam minimizar as parcelas de custos fixos, atendendo toda a demanda, e buscando a melhor localização para p instalações que devem ser abertas. Os problemas de custos fixos lembram os problemas de medianas, mas se diferenciam por incluírem na função objetivo a minimização da parcela de custos fixos, uma vez que o número de instalações abertas é definido endogenamente.

Tendo em vista a grande quantidade de trabalhos publicados sobre todas as classes de problemas de localização (REVELLE; EISELT; DASKIN, 2008), esta revisão de literatura se dedica ao estudo dos problemas com custo fixo. Maiores informações sobre modelagens e

aplicações para as demais classes de problemas de localização, podem ser encontradas nos livros de Drezner e Hamacher (2002), Farahani e Hekmatfar (2009) e Laporte, Nickel e Saldanha-da-Gama (2015).

De acordo com Hakli e Ortacay (2019), os problemas com custo fixo são utilizados como modelos de teste em estudos de otimização, uma vez que se assume que todas as instalações abertas possuem capacidades suficientemente grandes para atender toda a demanda e, diferentemente dos demais problemas, a quantidade de instalações abertas no problema é definida de maneira endógena.

Os problemas com custo fixo podem ser formulados como problemas de programação linear inteira mista (do inglês *Mixed Integer Linear Programming, MILP*) para os quais pode não existir algoritmos polinomiais para sua solução (problemas NP-completo). Alguns trabalhos, dentre os quais Cornuéjols, Nemhauser e Wolsey (1990) e Garey e Johnson (1979) demonstram que os problemas de otimização associados aos problemas de localização com custo fixo são classificados como NP-*hard*.

O fato de os problemas de custo fixo serem do tipo NP-completo implica que quanto maior o seu tamanho, maior a quantidade de operações matemáticas e de tempo de processamentos necessários para a obtenção de uma solução. Em alguns casos, é possível que não seja obtida a solução ótima do problema dado o seu tamanho.

Para lidar com o surgimento de problemas de aplicações reais cada vez maiores, as pesquisas buscaram tanto melhorar os métodos exatos quanto as heurísticas utilizadas para a obtenção de solução. Os métodos exatos são aqueles que, num tempo finito, encontram uma solução ótima para um problema, enquanto os métodos heurísticos são utilizados para obter soluções em problemas muito grandes para serem resolvidos por métodos exatos, e podem não encontrar uma solução ótima.

Alguns dos primeiros trabalhos sobre métodos para a solução de problemas de custo fixo utilizavam algoritmos com o *branch-and-bound* (KHUMAWALA, 1972), a relaxação lagrangeana (BARCELO *et al.*, 1990), o problema dual (ERLENKOTTER, 1978) e a programação linear (VAN ROY, 1986) para a obtenção de soluções ótimas. Estudos sobre as estruturas poliédricas da envoltória convexa dos problemas com custo fixo também foram realizados (AARDAL *et al.*, 1995), possibilitando melhorias nos algoritmos de *branch-and-cut* (AARDAL, 1998; MARCHAND *et al.*, 2002), por exemplo.

Em uma abordagem mais teórica e utilizando as estruturas poliédricas do problema de localização de instalações com custo fixo, Sankaran (2007) apresenta dois métodos para solucionar instâncias muito grandes do problema de custo fixo e suas implicações. O primeiro método baseia-se no agrupamento de pontos de demanda em um agrupamento grande que representa a maior parte da demanda, e outro menor, que representa o restante da demanda. Enquanto o segundo método baseia-se em remover algumas das restrições associadas ao limitante superior do problema, fazendo com que o *gap* do problema dual seja o menor possível.

Ainda assim, é possível que o tamanho dos problemas limite a capacidade desses algoritmos e relaxações de obter soluções, de maneira que mais pesquisas acerca do uso de métodos heurísticos têm acontecido. Dentre os trabalhos que utilizam métodos heurísticos, existem aqueles que empregam somente métodos heurísticos para a obtenção de soluções ótimas ou quase ótimas, tais como Kratica *et al.* (2001) e Hakli e Ortacay (2019) que utilizam algoritmos genéticos com diferentes mutações nos genes para obter soluções para o problema de localização de instalações com custo fixo. Além do algoritmo genético, outras heurísticas como busca tabu (SUN, 2006), *simulated annealing* (YIGIT; AYDIN; TURKBEY, 2006), *particle swarm optimization* (GUNNER; SEVKLI, 2008) e colônias de abelhas artificiais (KASHAN; NAHAVANDI; KASHAN, 2012) também são utilizados para a obtenção de soluções ótimas ou quase ótimas.

Além de serem empregadas na obtenção de soluções dos problemas de localização, também é possível empregar as heurísticas na inicialização de problemas para que algoritmos como o *branch-and-bound* e a relaxação lagrangeana sejam mais eficientes na obtenção de soluções ótimas ou quase ótimas de problemas maiores, em um tempo de processamento razoável. Galvão e Raggi (1989) apresentam um método em três etapas que aplica um algoritmo primal-dual, uma otimização do subgradiente para resolver uma relaxação lagrangeana e um algoritmo *branch-and-bound*, para a obtenção de soluções ótimas.

Por sua vez, Barahona e Chudak (2005) apresentam uma heurística para a obtenção de soluções utilizando o subgradiente para obter soluções do problema primal e aproximações aleatórias no problema dual, que permite melhorar as soluções obtidas. Já Hansen *et al.* (2007) utilizam *variable neighborhood search* para obter soluções com o problema primal, que posteriormente serão melhoradas utilizando o problema dual.

Existem diversas características da realidade que, muitas vezes, não são consideradas nos modelos. O problema de localização de instalações com custo fixo pode ser reformulado para considerar aspectos da realidade tais como múltiplos níveis (KRATICA; DUGOŠIJA; SAVIĆ, 2014), múltiplos horizontes de planejamento (MARQUES; DIAS, 2013) e múltiplos produtos (MELO; NICKEL; SALDANHA-DA-GAMA, 2006).

2.2. Métodos de agrupamento de pontos de demanda

Em situações nas quais o problema de localização possui uma grande quantidade de pontos de demanda, uma alternativa tanto para diminuir a quantidade de variáveis de decisão quanto para modelar problemas muito grandes que surgem na prática é agrupar os pontos de demanda. A formação de agrupamentos, ou *clusters*, é o processo de separar um conjunto de pontos de demanda em subconjuntos menores, de maneira que cada subconjunto (agrupamento), apresente pontos de demanda similares entre si, mas diferentes dos demais. Em alguns artigos e livros, os autores utilizam os termos observações ou objetos para se referir aos elementos que estão sendo agrupados (FÁVERO; BELFIORE, 2017; HAN; KAMBER; PEI, 2012; JAIN, 2010). Por este trabalho tratar do agrupamento de pontos de demanda, optou-se por manter a nomenclatura dos elementos a serem agrupados mais coerente ao decorrer do texto.

Na literatura, existem trabalhos que agrupam os pontos de demanda com base na sua localização geográfica. Um exemplo é o trabalho de Zhao e Batta (1999), que agrupa todas as residências (pontos de demanda) do condado de Ontário, nos Estados Unidos, com base no CEP de sua região. Para cada região geográfica associada a um CEP, todos os pontos de demanda ali existentes são agregados no centroide. Outra forma de agregar pontos de demanda com base na região geográfica é através das zonas censitárias, como realizado por Hodgson, Shmulevitz e Körkel (1997), que também agrupam os pontos de demanda no centroide da zona.

Já Sankaran (2007) cita o agrupamento de pontos de demanda considerando os limites do município. Ele sugere que, ao modelar um problema com uma quantidade muito grande de pontos de demanda em uma metrópole e suas cidades satélites, a demanda seja agregada em dois pontos: um que representa a metrópole e outro que representa os demais municípios, ou que, preferivelmente, a demanda de cada cidade seja individualmente representada no modelo. Galindo e Batta (2013), por sua vez, também utilizam a representação de metrópoles e cidades satélites proposta por Sankaran (2007) em um contexto de localização de instalações em caso de desastre humanitário.

Ghiani *et al.* (2012) formam agrupamentos considerando todos os pontos de demanda (habitantes) que se encontram em uma rua ou quadra, para a modelagem de um sistema de coleta de resíduos sólidos urbanos.

Por se basearem em localizações geográficas, esses agrupamentos podem não considerar adequadamente a forma como os pontos de demanda se distribuem no espaço. Uma forma de contornar possíveis equívocos ao agrupar de pontos de demanda é empregando algoritmos de formação de agrupamentos.

O'Kelly (1992) utiliza o algoritmo *K-Means* para agrupar os pontos de demanda e solucionar um problema de localização de terminais (*hub location problem*); enquanto Erkut e Bozkaya (1999) utilizam tanto o algoritmo *K-Means*, quanto um método hierárquico com três tipos de encadeamento (forma como as distâncias entre os pontos de demanda são calculadas) para agrupar pontos de demanda na cidade de Edmonton, Canadá. Nas soluções dos problemas de *p*-medianas, os autores observaram que as soluções dos modelos com pontos de demanda agregados pelo método *K-Means*, apresentavam as menores diferenças (valor do erro) com relação ao problema com pontos de demanda desagregado.

Também no contexto de localização de instalações para respostas de desastres humanitários, Dekle *et al.* (2005) agrupam os pontos de demanda utilizando um algoritmo baseado em um parâmetro de distância que considera os pontos mais distantes para formação de agrupamentos, enquanto Paul e MacDonald (2016) utilizam o algoritmo *K-Means* para formar agrupamentos de pontos de demanda e aplicar uma heurística que permita a obtenção de soluções. Já Levin e Ben-Israel (2004), Esnaf e Küçükdeniz (2009) e Gao (2021), utilizam heurísticas para agregarem o conjunto de pontos de demanda em *K* agrupamentos e solucionar tanto problemas de localização de instalações com custo fixo, quanto de localização de armazéns.

Outro método de formação de agrupamentos existente consiste na repartição do espaço em células (ou *grids*), a partir das quais as operações de agrupamento de pontos de demanda são realizadas utilizando algoritmos geradores de linhas e colunas. Nesse sentido, um dos principais trabalhos é o de Andersson *et al.* (1998), que agrega os nós da malha viária de uma cidade utilizando células para solucionar um problema de *p*-medianas em rede.

Em alguns trabalhos mais recentes como Irawan *et al.* (2017), Irawan e Salhi (2015b) e Cebecauer e Buzna (2017), os autores têm utilizado abordagens que integram a formação de

agrupamentos ao processo de solução dos problemas de localização, utilizando heurísticas para a obtenção de soluções ótimas ou aproximadas para problemas muito grandes.

Sabendo-se que os pontos de demanda podem ser classificados como dados espaciais, ainda é possível agrupá-los utilizando métodos de densidade, tais como o DBSCAN (ESTER *et al.*, 1996) e o OPTICS (ANKERST *et al.*, 1999). Diferentemente de alguns dos métodos de formação de agrupamentos, como o *K-Means* e os métodos de encadeamento, utilizados por Erkut e Bozkaya (1999), os métodos de densidade baseiam seu conceito de agrupamentos entre regiões de maior densidade de pontos de demanda no espaço, separadas por regiões com menor densidade de pontos (HAN; KAMBER; PEI, 2012). No que diz respeito à utilização dos métodos de densidade para a solução de problemas de localização, não foram encontrados trabalhos científicos que empregue algum destes métodos.

2.3. Erros decorrentes do agrupamento de pontos de demanda

Segundo Francis e Lowe (2014), a utilização de agrupamentos para representar os pontos de demanda é a causa dos erros de agregação. Existem diversas fontes através das quais os erros de agregação se originam. Um dos trabalhos pioneiros a abordar essas fontes de erro foi o de Hillsman e Rhoda (1978), que apresenta os erros de tipo A, B e C. O trabalho de Current e Schilling (1987) apresenta soluções para evitar os erros tipo A e B, enquanto Hogdson e Neuman (1993) propõem soluções para evitar os erros de tipo C. Posteriormente, Hogdson, Shmulevitz e Körkel (1997) definem os erros de tipo D e propõem formas de minimizá-los em sua pesquisa.

Os erros de tipo A ocorrem quando se utiliza a distância entre o ponto de demanda agregado (i.e., que representa um conjunto de pontos) e a instalação com a finalidade de representar a distância entre todos os pontos de demanda representados por ele e a mesma instalação. A Figura 2.1 ilustra este tipo de erro; ao utilizar a distância d_{ij} entre o ponto de demanda agregado i e a instalação j para representar a distância entre os pontos K e $K + 1$, é possível notar que essa distância pode não representar corretamente as distâncias desses pontos até a instalação j . Para remover este erro, Current e Schilling (1987) propõem o uso de uma distância (ou custo) ponderada pela demanda para melhorar a representação do problema.

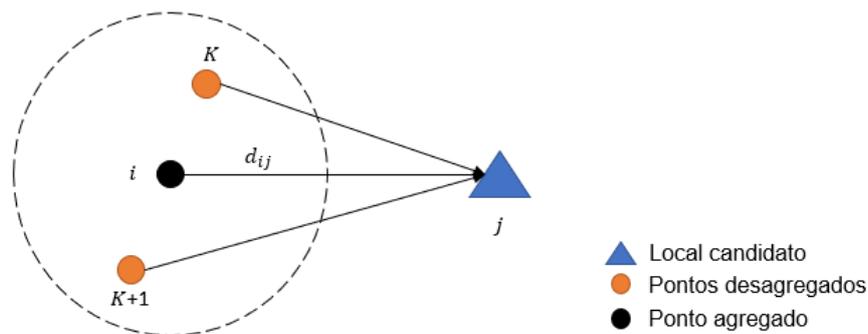


Figura 2.1 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo A. Adaptado de Current e Schilling (1987)

Os erros de tipo B são um caso especial de erro de tipo A, que deve ser separadamente analisado, uma vez que nem todos os problemas com pontos de demanda agregado apresentam este erro. Ele ocorre ao assumir que a instalação que atende à demanda de um ponto de demanda agregado e o próprio ponto de demanda agregado localizam-se no mesmo local no espaço. A Figura 2.2 mostra este erro. Observa-se que como o local candidato e o ponto de demanda agregado encontram-se no mesmo local no espaço, o que acarreta se erroneamente induzir que o custo para atender toda a semana seja igual a zero, quando na realidade não é. Com relação a este erro, uma forma de removê-lo, é substituindo a distância entre local candidato e ponto de demanda agregado (igual a zero), pela distância ponderada entre os pontos de demanda que compõem o agrupamento e o centroide do agrupamento.



Figura 2.2 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo B. Adaptado de Current e Schilling (1987)

Já os erros de tipo C ocorrem quando um ponto de demanda não é alocado para a instalação aberta mais próxima, uma vez que a decisão de alocação foi feita com base na instalação de menor custo com relação ao ponto de demanda agregado que o representa. Na Figura 2.3 observa-se que o ponto de demanda K é representado pelo ponto i , cuja demanda é alocada para a instalação j , diferentemente de K , cuja demanda seria alocada para $j + 1$, caso

não fosse representado pelo ponto agregado i . A remoção dos erros de tipo C pode ocorrer ao se desagregar os agrupamentos e buscar, para cada ponto de demanda, a instalação mais próxima capaz de atender a demanda.

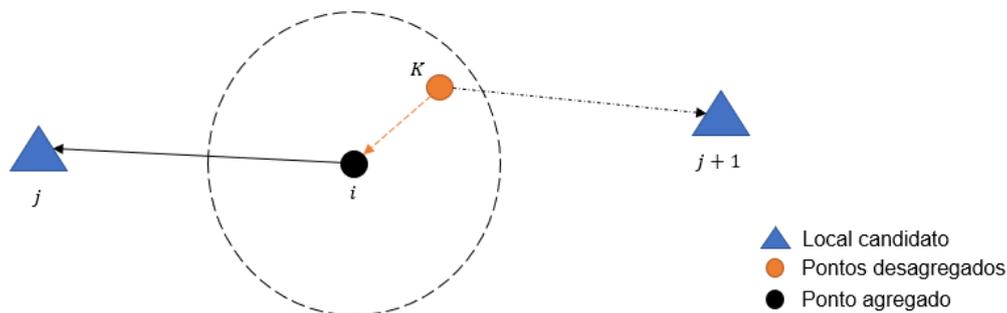


Figura 2.3 – Exemplo de agregação de pontos de demanda para demonstrar erros do tipo C. Adaptado de Current e Schilling (1987)

Os erros de tipo D acontecem ao definir locais candidatos a instalação a partir dos pontos de demanda agregado, de forma que o conjunto de locais candidatos não é ótimo. Uma maneira de remover este erro é desagregando o ponto de demanda agregado e utilizando o ponto mais próximo ao centroide como candidato a instalação.

Erkut e Bozkaya (1999) propõem seis tipos de erros que podem ser causados pelos analistas ao preparar os dados na etapa de modelagem. Os autores citam: a utilização de uma demanda uniforme (UA); formação de agrupamento de pontos de demanda sem a utilização de algoritmos que considerem agrupamentos populacionais dentro dos conjuntos de dados (RA); não considerar soluções alternativas na análise (FL); tomar decisões com base em uma única solução (EC); seleção de locais candidatos de maneira endógena ao problema (DF); e a não utilização de um nível correto de agregação (OA).

Em algumas classes de problemas de localização, é possível calcular a magnitude dos erros de agregação através de diferentes medidas (FRANCIS et al., 2009; IRAWAN; SALHI, 2015a). Em aplicações dos problemas de cobertura (CURRENT; SCHILLING, 1990; DASKIN et al., 1989), de centro (RAYCO; FRANCIS; LOWE, 1996) e de medianas (ANDERSSON et al. 1998; CEBECAUER; BUZNA, 2017; ERKUT; BOZKAYA, 1999; HODGSON; SHMULEVITZ; KÖRKEL, 1997; IRAWAN; SALHI, 2015b; ZHAO; BATTA, 1999), o que mais se observa é o cálculo da diferença no valor da função objetivo entre os problemas agregado e o desagregado.

2.4. Considerações sobre a revisão de literatura

Através da revisão da literatura, percebe-se que foram realizadas diversas melhorias nos métodos exatos e heurísticos utilizados para obter soluções de problemas de localização, principalmente decorrente da necessidade de solucionar problemas cada vez maiores. Todavia, o que se nota nessa seção é que houve uma mudança no foco dos trabalhos relacionados à obtenção de solução de problemas de localização com grande número de variáveis de decisão: foram da abordagem de aspectos do agrupamento de pontos de demanda, dentro do processo de modelagem do problema, para uma abordagem que buscava melhorar algoritmos e métodos de resolução dos problemas.

Dadas as limitações de tempo e de capacidade computacional, a busca por maneiras de diminuir adequadamente o tamanho do problema sem que haja perda na acurácia foi tema de diversas pesquisas. Observa-se que os pontos de demanda são, principalmente, agrupados com base em alguma subdivisão geográfica, mas que a aplicação de métodos de formação de agrupamentos e de heurísticas, capazes de obter uma quantidade pré-estabelecida de agrupamentos, também é empregada. Tantas maneiras de agrupar pontos de demanda corroboram o ponto de vista de Francis *et al.* (2009), que afirmam não existir consenso quanto ao melhor método de agrupamento de pontos de demanda. Nota-se também que a exceção do trabalho de Erkut e Bozkaya (1999), os demais não comparam os resultados dos problemas com pontos de demanda agregados a partir de diferentes métodos, analisando apenas se, nos contextos nos quais se encontram, os problemas agregados são simplificações adequadas dos problemas reais.

Por fim, foram apresentadas as principais fontes de erro decorrentes do uso de pontos de demanda agregados na modelagem dos problemas de localização. Muitas delas podem ser removidas se o tratamento dos dados de entrada do problema for realizado adequadamente. Nas situações nas quais as fontes de erro não podem ser removidas, é possível mensurar esse erro, mas, assim como no caso dos métodos de agrupamento de pontos, também não existe consenso na literatura quanto ao emprego das medidas de erro (qual é mais apropriada) sendo que, usualmente, calcula-se a diferença entre o valor da função objetivo do problema com pontos de demanda desagregados e agregados.

3. ESCOPO E METODOLOGIA

O objetivo deste capítulo é delimitar o escopo do problema abordado por esta pesquisa e apresentar a metodologia utilizada para agrupar os pontos de demanda e avaliar os efeitos nos resultados dos modelos de otimização de localização de instalações em decorrência da utilização de diferentes métodos para agrupar os pontos de demanda.

Na revisão de literatura foram apresentados tanto métodos de agrupamento com base na localização geográfica dos pontos de demanda dada a simplicidade com que podem ser definidos (ZHAO; BATTA, 1999; SANKARAN, 2007), quanto utilizando algoritmos de formação de agrupamentos (ERKUT; BOZKAYA, 1999). Com relação aos algoritmos de formação de agrupamento, o algoritmo *K-Means* foi utilizado em uma quantidade maior de trabalhos se comparado aos demais.

Em problemas reais nos quais se busca determinar a melhor configuração de uma rede logística para uma região ou área de estudo, tal como um país, parte dele, ou estado, a proposta desta pesquisa é investigar e avaliar a adoção de um método de agrupamento dos pontos de demanda visando simplificar o modelo e facilitar a sua resolução, como discutido anteriormente. Mais especificamente, objetiva-se avaliar alternativas à prática usual de agrupamento de pontos por município. Essa prática é usual; todavia, quando existe uma localidade que possua uma grande quantidade de pontos de demanda, e represente grande parte da demanda, como no caso de uma megacidade ou de uma grande região metropolitana, esta simplificação pode não ser a mais adequada, implicando em distorções que levam a diferenças nos resultados dos modelos de otimização quando comparados a uma modelagem sem agregação de pontos de demanda.

Sendo assim, esta metodologia busca responder como os pontos de demanda devem ser agrupados e considerados dentro da modelagem matemática de um problema de localização de instalações, ou seja, se é possível manter uma modelagem usual considerando os limites geográficos dos municípios, ou se é necessário utilizar outro método de formação de agrupamentos, além de avaliar qual o impacto na qualidade das soluções obtidas através de um o modelo matemático decorrente da utilização dos métodos de formação de agrupamentos que permitem simplificar a modelagem do problema. Adicionalmente, busca-se responder quais são as características presentes nas instâncias que eventualmente tornem os métodos de agrupamento mais adequados.

O esquema da Figura 3.1 resume a metodologia utilizada no trabalho. Todos os pontos de demanda existentes em um problema de localização de instalações, são agregados por dois métodos diferentes de formação de agrupamentos: por município e utilizando o algoritmo *K-Means* (MACQUEEN, 1967). Optou-se por utilizar esses métodos de formação de agrupamentos uma vez que o agrupamento por municípios proposto por Sankaran (2007) foi aplicado em um contexto de presença de municípios de diferentes tamanhos e demandas em regiões metropolitanas, como um sistema real B2B de bens de consumo nos quais existem diferenças na demanda entre localidades, enquanto o algoritmo *K-Means* foi aquele dentre os agrupamentos de pontos que apresentou valores da função objetivo mais próximos ao modelo desagregado no trabalho de Erkut e Bozkaya (1999). Além disso, utilizar um método de agrupamento geográfico e outro que utiliza um algoritmo permite comparar as soluções dos modelos com pontos agregados em locais do espaço nos quais a densidade de pontos de demanda não é uniforme, ainda que dentro de um mesmo município.

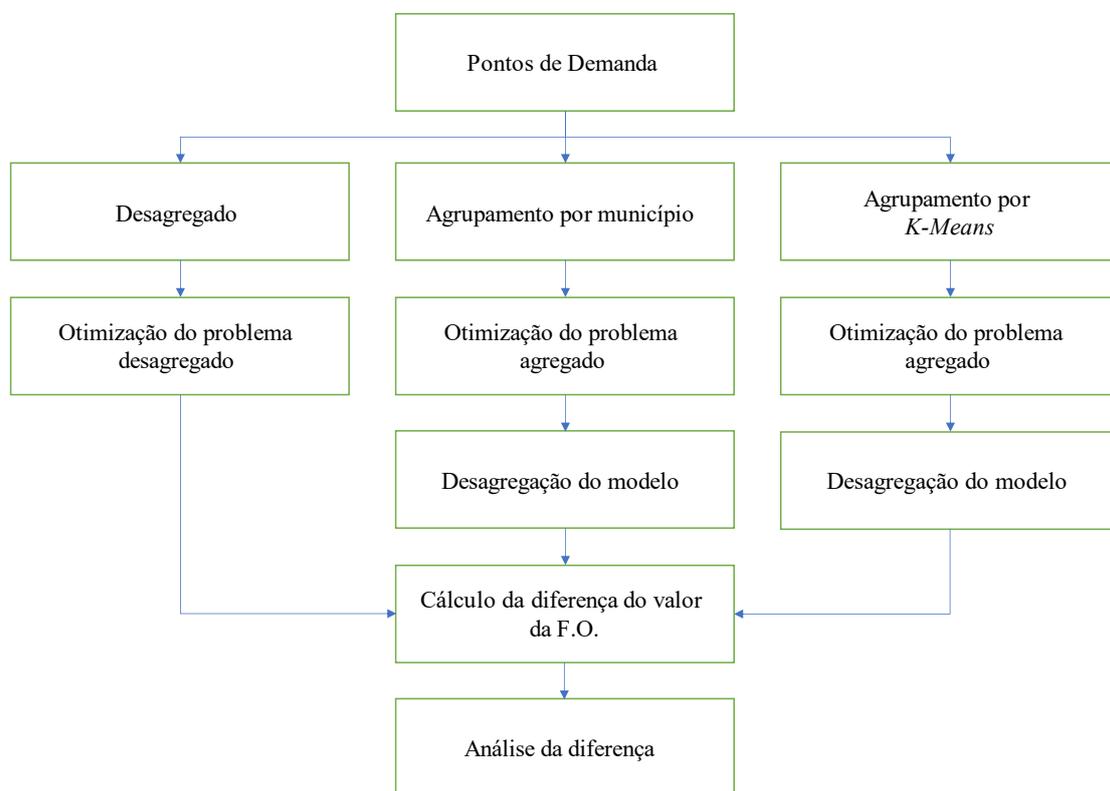


Figura 3.1 – Esquema metodológico

Os modelos que representam os problemas com pontos de demanda agregados são otimizados e posteriormente desagregados, para que os pontos de demanda sejam sempre alocados a instalação de menor custo aberta para atendê-los, eliminando os erros de tipo C da análise. O modelo com pontos de demanda desagregados também é solucionado. Posteriormente, com base no valor da função objetivo, calcula-se a diferença entre as soluções

dos problemas com pontos de demanda agregados e desagregado. Por fim, são analisados os casos nos quais existe diferença entre os resultados de modelos com pontos de demanda agregados e desagregados, buscando onde acontecem e quanto representam essas diferenças.

Nas próximas seções, são detalhados (i) o escopo e formulação do modelo de problema de localização de instalação que é utilizado no trabalho, (ii) os métodos de formação de agrupamentos utilizados para agrupar os pontos de demanda, (iii) solução do modelo; e (iv) cálculo da diferença e análise da qualidade da solução.

3.1. Escopo e formulação do modelo do problema de localização de instalações com custo fixo

Para Melo, Nickel e Saldanha-da-Gama (2009), um modelo qualquer de problema de localização de instalações compreende um conjunto de clientes¹ e outro de instalações para atender a demanda. Além disso, são necessárias outras informações, tais como distância, tempo de viagem e custos entre cada par origem-destino (instalações-clientes), cada qual com sua respectiva unidade de medida.

Um dos principais modelos de localização de instalações é denominado problema de localização de instalações com custo fixo (*Fixed Charge Facility Location Problem – FCFL*), no qual a função objetivo busca tanto a minimização dos custos variáveis de transportes entre os pares origem-destino para atender a demanda, quanto a minimização da parcela de custos fixos de operação de cada instalação, sendo essa quantidade definida de maneira endógena ao modelo.

O problema de localização é modelado como o FCFL, podendo os pontos de demanda serem desagregados ou agregados; em ambos os casos os conjuntos de locais candidatos a instalação são os mesmos. Apresenta-se a seguir a formulação matemática do FCFL (KRARUP; PRUZAN, 1983). As notações utilizadas na formulação foram padronizadas por conjunto de forma a facilitar a compreensão, como se segue:

I : Conjunto de locais candidatos à instalação, tal que $i \in I$

J : Conjunto dos pontos de demanda desagregados, tal que $j \in J$

K : Conjunto dos agrupamentos para o respectivo modelo, tal que $k \in K$

¹ Neste trabalho não se utiliza o termo clientes para representar os pontos de demanda, entretanto, neste contexto, ambos são tratados como equivalentes. Segundo Guazzelli (2018), os pontos de demanda representados por modelos de localização não são necessariamente clientes.

São conhecidos: a demanda q_j de cada ponto de demanda $j \in J$; o custo c_{ij} para transportar uma unidade de produto desde um local candidato $i \in I$ para um ponto de demanda $j \in J$. Tal representação é similar no caso de os pontos de demanda serem agrupados e representados por pontos/locais agregados. Fazemos apenas a diferenciação do índice k no lugar de j ; em outras palavras, a demanda em unidades por agrupamento é representada pelo parâmetro q_k , enquanto o custo para transportar uma unidade de produto desde um local candidato i para um agrupamento k , é representado pelo parâmetro, c_{ik} . Os custos de transporte c_{ij} e c_{ik} são compostos por uma parcela fixa, que representa o custo de mão-de-obra, e uma parcela variável proporcional à distância percorrida, de forma que são calculados como sendo uma função linear, como $c_{ij} = b + ax$. Tanto nos modelos com pontos de demanda desagregados ou agregados, o custo fixo² de operação das instalações é representado por f_i , para cada um dos locais candidatos $i \in I$.

Sejam X_{ij} as variáveis de decisão que representam o número de unidades de produto enviadas desde um local candidato i para atender a demanda de um ponto j ; em outras palavras, essas variáveis estão relacionadas ao fluxo de produtos entre os locais candidatos e pontos de demanda, o que permite que os fluxos sejam tanto representados pela quantidade de produto transportada de um ponto a outro, quanto pelo percentual de volume de demanda de um ponto de demanda quando atendido por uma instalação.

Por último, sejam Y_i as variáveis de decisão binárias, indicam que se um local candidato $i \in I$ é aberto ou não, de forma que a demanda dos pontos $j \in J$ somente seja atendida por instalações que se encontram abertas.

Desta forma, o problema de localização de instalações com custo fixo para um modelo com pontos de demanda desagregados pode ser formulado como se segue pelas Restrições (3.1) a (3.5) .

A função objetivo (3.1) busca minimizar as parcelas de custos fixos e variáveis da operação. As restrições (3.2) garantem que a demanda de todos os pontos seja atendida, enquanto as restrições (3.3) garantem que apenas locais candidatos abertos possam atender esta demanda. As restrições (3.4) e (3.5) garantem o domínio das variáveis.

² Ressalta-se que a longo prazo, nenhum custo é verdadeiramente fixo. Neste trabalho, o custo fixo de operação das instalações deve ser visto como um custo sujeito a variações ao longo do tempo, conforme o tamanho das instalações a que se associam.

Para diferenciar a modelagem do FCFL dos problemas com demanda agregada dos de demanda desagregada, a função objetivo, representada pela Equação (3.1) e as Restrições (3.2) a (3.5) devem ter todos os subscritos j substituídos pelos subscritos k .

$$\text{minimizar } \sum_{i \in I} \sum_{j \in J} X_{ij} * c_{ij} + \sum_{i \in I} Y_i * f_i \quad (3.1)$$

Sujeito a:

$$\sum_{i \in I} X_{ij} \geq q_j, \forall j \in J \quad (3.2)$$

$$X_{ij} \leq q_j * Y_i, \forall i \in I, \forall j \in J \quad (3.3)$$

$$X_{ij} \geq 0, \forall i \in I, \forall j \in J \quad (3.4)$$

$$Y_i \in \{0,1\}, \forall i \in I \quad (3.5)$$

3.2. Métodos de formação de agrupamentos de pontos de demanda

São propostos dois métodos de agrupamentos (Ma) de pontos de demanda, sendo um deles um agrupamento geográfico que considera o limite geográfico do município, e outro que aplica o algoritmo *K-Means* (MACQUEEN, 1967). Considerando uma megacidade como São Paulo, na qual 1/3 dos habitantes encontra-se na zona leste da cidade, representá-los por um ponto localizado na região central da cidade (como no agrupamento por municípios) pode ser uma representação incorreta da localização desses pontos na modelagem do problema de localização. Assim, ao utilizar dois métodos de agrupamentos diferentes, é possível comparar os resultados obtidos através de cada um deles, analisando também os efeitos dessa modelagem em regiões de maior demanda e quantidade de pontos de demanda, como nas megacidades de São Paulo e do Rio de Janeiro.

3.2.1. Método de agrupamento *Ma1* – Pontos de demanda agregados por município

Este método considera o limite geográfico do município para agrupar os pontos de demanda. Por esta simplificação, todos os pontos de demanda dentro do perímetro do município são representados por apenas um ponto, localizado ou no centroide ou no baricentro do agrupamento de pontos.

A Figura 3.2 ilustra um exemplo da aplicação deste método de agrupamento. Nela, estão presentes os municípios de Diadema, Santo André, São Bernardo do Campo e São Caetano do Sul, localizados na Região Metropolitana de São Paulo. Os pontos de demanda em cada

município são mostrados em cores diferentes, cujo diâmetro é proporcional a sua respectiva demanda. Ao agrupar os pontos de demanda pelo método Ma1, estes tornam-se representados pelos círculos opacos (mais claros), localizados no baricentro³ dos pontos de demanda do município, que agora também compreendem sua demanda. Para o município de São Caetano do Sul, não é possível observar o círculo opaco, e isso se deve ao fato de que existe apenas um ponto de demanda no município, que representa a si mesmo no modelo com pontos de demanda agregados por município.

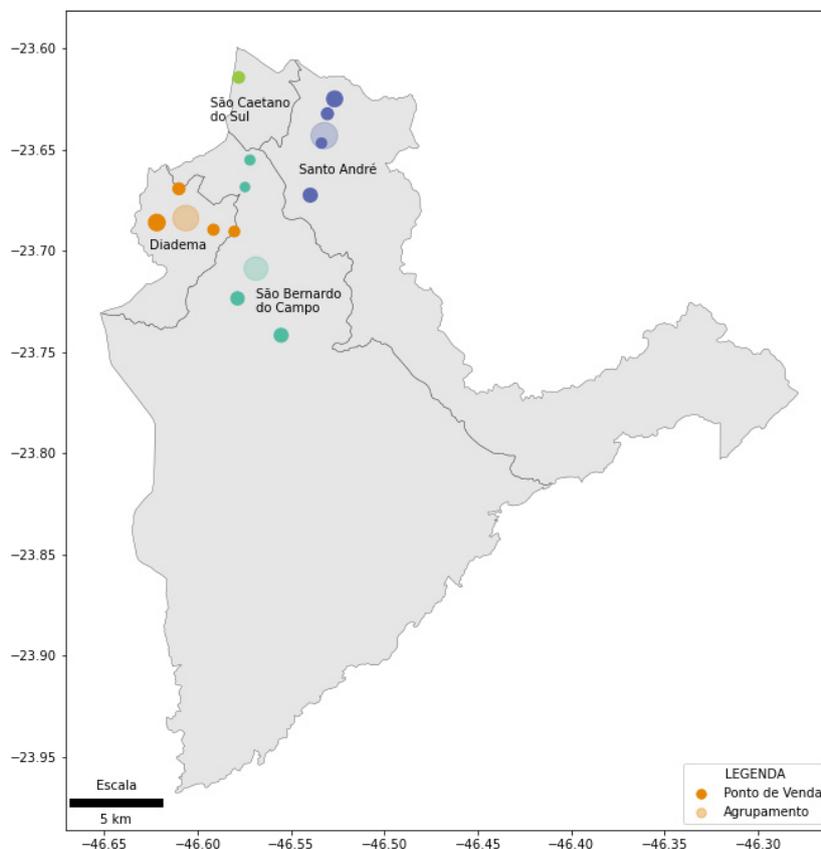


Figura 3.2 – Exemplo de agrupamento utilizando o método Ma1

3.2.2. Método de agrupamento Ma2 – Agrupamentos obtidos através do algoritmo K-Means

Neste método, utiliza-se o algoritmo *K-Means* (MACQUEEN, 1967) para agrupar os pontos de demanda de uma dada instância. O algoritmo consiste em selecionar aleatoriamente K pontos de demanda do conjunto, que inicialmente representam a média ou centro do agrupamento, e ir alocando os demais pontos ao agrupamento mais similar, com base na menor

³ Embora muitos trabalhos utilizem o centroide do agrupamento dos pontos de demanda como base para solucionar os modelos com pontos de demanda agregados, optou-se por utilizar o baricentro uma vez que a demanda dos pontos não é uniforme nas instâncias solucionadas, e por considerar que pontos cuja demanda é maior, deveriam ser mais próximos ao centro que outros que possuem menor demanda. Para maior aprofundamento sobre a utilização do centroide na solução de problemas de localização, sugere-se a leitura de Zhao e Batta (1999).

distância entre o ponto e a média do agrupamento (por isso o nome do algoritmo). A cada nova iteração, as médias dos agrupamentos são recalculadas e, uma vez que todos os pontos são alocados a um agrupamento, o algoritmo é encerrado (HAN; KAMBER; PEI, 2012).

Uma vez que não são consideradas restrições de capacidade, é possível que existam agrupamentos heterogêneos tanto no que diz respeito à quantidade de pontos de demanda, quanto de demanda em um mesmo conjunto de dados. Outro ponto com relação ao algoritmo *K-Means* diz respeito a quantidade K de agrupamentos que o algoritmo deve encontrar, que é parâmetro de entrada, ou seja, definida *a priori*. Para comparar os efeitos da formação de agrupamentos no caso em que dois métodos diferentes possuem a mesma quantidade K de agrupamentos, optou-se por considerar o valor K igual à quantidade de municípios. Em experimentos anteriores utilizando o método do pulo, do inglês *jump method* (SUGAR; JAMES, 2003), e o número de municípios, observou-se que o método do pulo indicava quantidades de agrupamento próximas ou iguais ao número de municípios, o que possibilitaria esta simplificação.

A Figura 3.3 exemplifica a modelagem pelo método Ma2. Nela, encontram-se os pontos de demanda nos municípios de Diadema, Santo André, São Bernardo do Campo e São Caetano do Sul apresentados anteriormente. Agrupando os pontos de demanda utilizando o algoritmo *K-Means*, com K igual a quantidade de municípios, observa-se que os agrupamentos (representados por cores diferentes) não respeitam os limites do município. Um mesmo baricentro pode representar pontos de demanda em mais de um município, como no caso do indicado em azul, ou um agrupamento de pontos dentro do próprio município, como o verde, no município de Santo André.

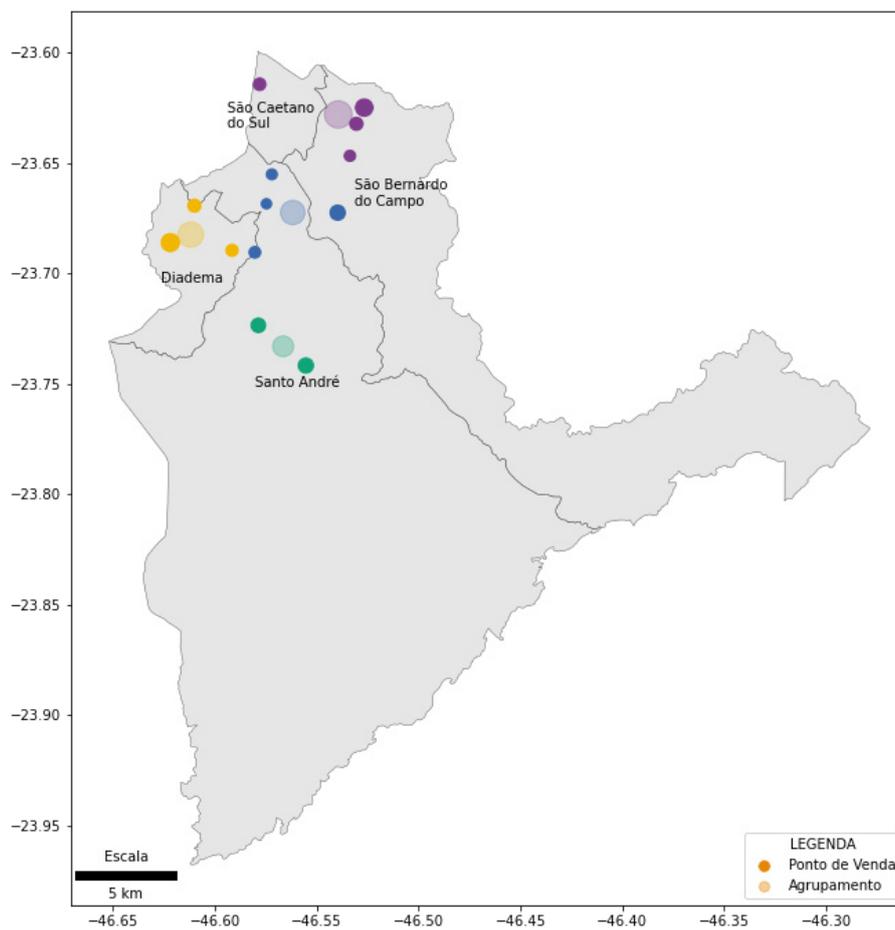


Figura 3.3 – Exemplo de aplicação do método Ma2

3.3. Solução do modelo

Conforme apresentado no fluxograma da Figura 3.1, após a definição dos agrupamentos, os modelos matemáticos resultantes são otimizados. Para efeitos de comparação entre os métodos de formação de agrupamento, cada problema de localização é solucionado três vezes, uma considerando seus pontos de demanda desagregados e outras duas considerando os pontos de demanda agregados pelos métodos anteriormente apresentados (Ma1 e Ma2).

No modelo desagregado, cada ponto de demanda é considerado individualmente, ou seja, a quantidade de agrupamentos (K) é igual a quantidade total de pontos de demanda. O modelo é o apresentado nas equações (3.1) a (3.5), com todos os subscritos j . Ainda que não represente um método de agrupamento, deve ser solucionado juntamente aos demais modelos pois é a base de comparação da qualidade dos resultados a partir dos dois modelos com pontos de demanda agregados. Como resultado deste processo de otimização, são obtidas a quantidade (número) e a localização dos candidatos abertos (i.e, aqueles para os quais $Y_i = 1$) e, a partir

de cada candidato aberto, qual deles é responsável por atender a demanda de cada ponto; além de retornar o valor da função objetivo do problema.

Por sua vez, os modelos simplificados (i.e., com pontos de demanda agregados) são otimizados e requerem uma etapa posterior de decomposição do modelo. Para que as diferenças entre as soluções sejam calculadas posteriormente, é necessário que os custos dos modelos estejam na mesma base, do contrário, essa comparação acarreta erro. Dessa forma, inicialmente soluciona-se um modelo matemático de problema de localização considerando os pontos de demanda agregados (utilizando subscritos k), que resulta na quantidade e na localização dos candidatos abertos para atender a demanda. Posteriormente, os agrupamentos são desfeitos e, para cada ponto de demanda desagregado são recalculados os custos considerando os candidatos abertos ($Y_i = 1$).

A Figura 3.4 mostra os pontos de demanda (círculos), locais candidatos abertos (triângulos) e como a demanda foi alocada (linhas). A cor de cada ponto de demanda está associada ao agrupamento ao qual pertencem. Por exemplo, todos os pontos verdes localizados na zona sul do município de São Paulo são parte de um mesmo agrupamento, bem como todos os pontos de cor laranja em Mogi das Cruzes são parte de outro agrupamento. Neste caso, nota-se que todos os pontos de demanda pertencentes a um mesmo agrupamento estão alocados à instalação de menor custo para atender o agrupamento. Todavia, esta pode não ser a instalação aberta de menor custo que atenda a demanda de cada ponto presente no agrupamento.

Já na Figura 3.5, mostra-se a alocação da demanda depois da etapa de desagregação do modelo. Observa-se que alguns pontos de demanda foram realocados para outras instalações, senão aquelas que foram inicialmente designadas para atender ao agrupamento, como no caso do ponto em lilás, indicado pela seta preta na figura. Desta forma, a realocação dos pontos de demanda às instalações abertas de menor custo ajusta o valor da função objetivo (FO) e remove os erros de tipo C decorrentes do processo de formação de agrupamentos.

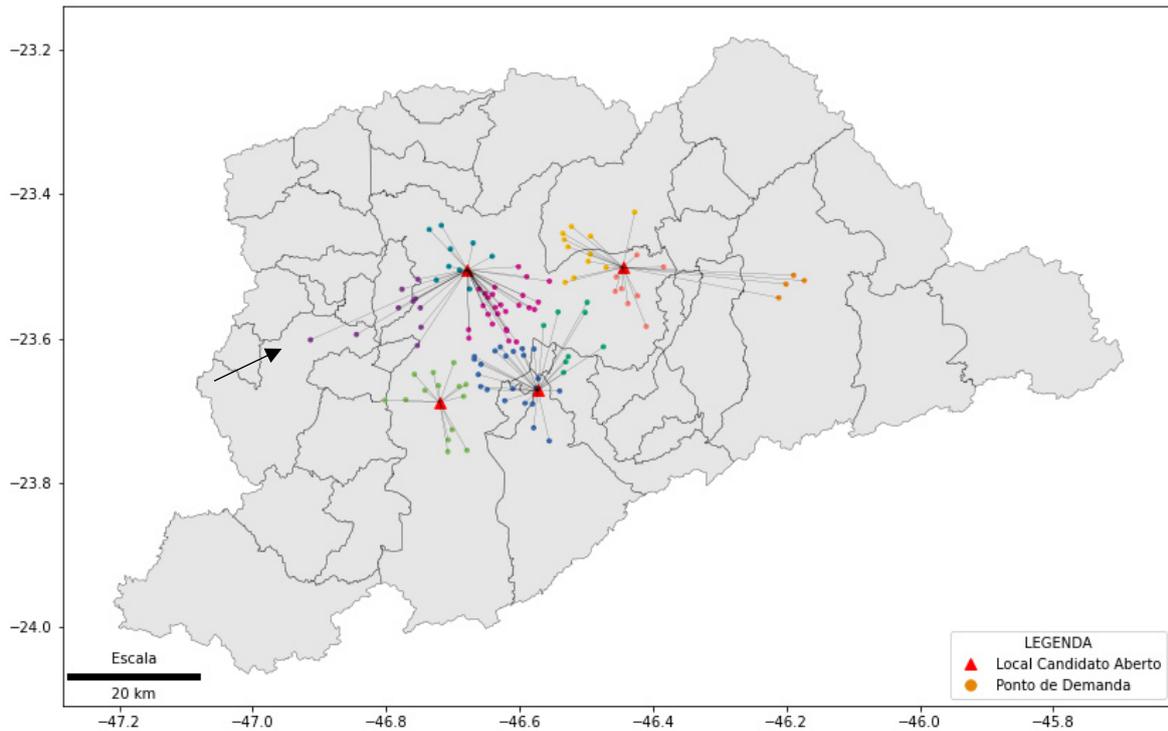


Figura 3.4 – Alocação dos pontos de demanda com base na alocação da demanda do agrupamento

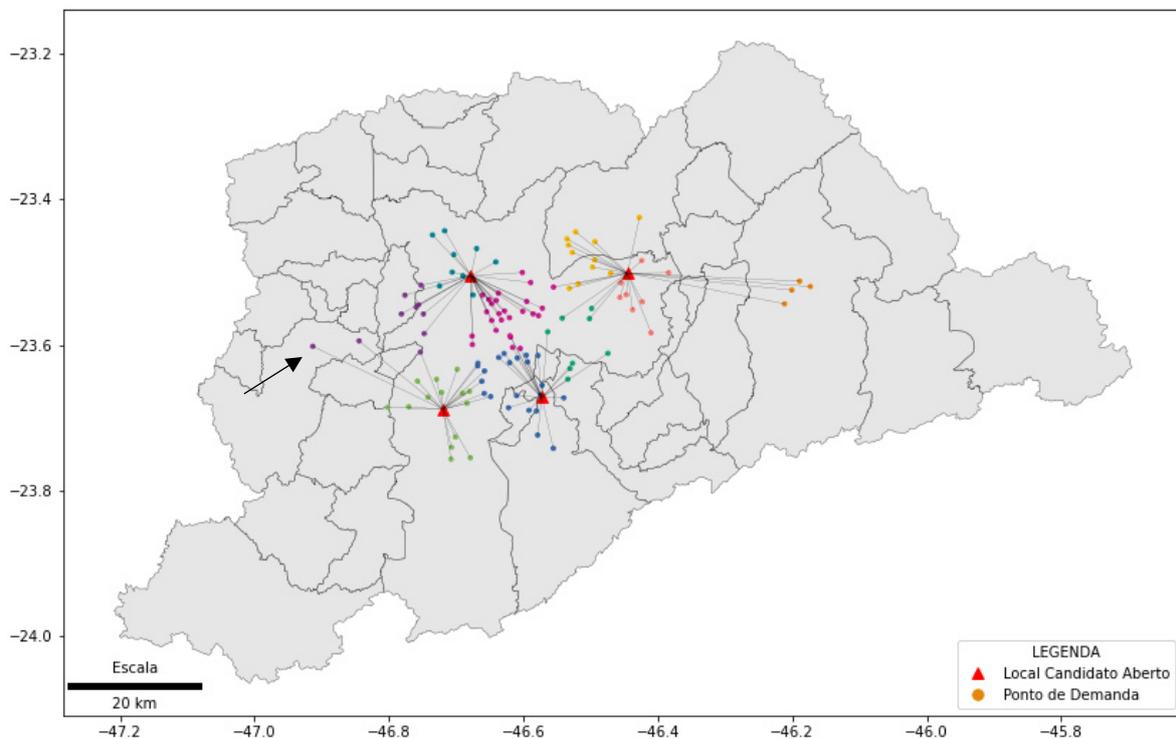


Figura 3.5 – Alocação dos pontos de demanda com base no menor custo definido pela etapa de desagregação do modelo

As informações fornecidas pelo *software* de otimização entre as quais o número de linhas e colunas e de variáveis diferentes de zero em cada modelo, bem como o tempo de

processamento, o número de planos de corte do algoritmo *branch-and-cut* e de iterações do *simplex* necessárias para a obtenção das soluções ótimas dos modelos são coletadas (ver Arenales *et al.* (2007) e Wolsey (1998)). A partir dessas informações, é possível verificar como os modelos agregados diminuem o tamanho dos modelos, bem como seu tempo de processamento, e se são modelos mais fáceis de se resolver que o modelo desagregado. Modelos mais complexos e com maior número de restrições, requerem um maior tempo de processamento, além de maior número de planos explorados do *branch-and-cut* e de iterações do *simplex*.

A Figura 3.6 exemplifica o resultado fornecido pelo *software* de otimização após solucionar um problema com 442 linhas, 432 colunas e 1.248 valores diferentes de zero. Em amarelo, além das informações sobre o tamanho do problema solucionado, são destacados o número de nós explorados e de iterações do *simplex* necessárias e o tempo de processamento. É possível observar que outras informações, tais como o tipo de variável presente no problema também estão disponíveis.

```
Academic license - for non-commercial use only - expires 2021-03-21
Using license file C:\Users\renat\gurobi.lic
Parameter method unchanged
  Value: -1 Min: -1 Max: 5 Default: -1
Gurobi Optimizer version 9.1.1 build v9.1.1rc0 (win64)
Thread count: 4 physical cores, 8 logical processors, using up to 8 threads
Optimize a model with 442 rows, 432 columns and 1248 nonzeros
Model fingerprint: 0xe24cc896
Variable types: 416 continuous, 16 integer (16 binary)
Coefficient statistics:
  Matrix range      [1e+00, 1e+04]
  Objective range   [4e-02, 1e+02]
  Bounds range      [1e+00, 1e+00]
  RHS range         [7e+01, 9e+02]
Found heuristic solution: objective 4021.4152500
Presolve time: 0.00s
Presolved: 442 rows, 432 columns, 1248 nonzeros
Variable types: 416 continuous, 16 integer (16 binary)

Root relaxation: objective 1.669296e+03, 52 iterations, 0.00 seconds

  Nodes |   Current Node |   Objective Bounds |   Work
  Expl Unexpl | Obj Depth IntInf | Incumbent BestBd Gap | It/Node Time
*    0     0 |             0 | 1669.2955500 1669.29555 0.00% | - 0s

Explored 0 nodes (52 simplex iterations) in 0.02 seconds
Thread count was 8 (of 8 available processors)

Solution count 2: 1669.3 4021.42

Optimal solution found (tolerance 1.00e-04)
Best objective 1.669295550000e+03, best bound 1.669295550000e+03, gap 0.0000%
```

Figura 3.6 – Informações provenientes do *software* de otimização

3.4. Cálculo da diferença e análise da qualidade da solução

Após a etapa de desagregação do modelo com pontos de demanda agregados e do recálculo dos custos, é calculada a diferença entre as soluções obtidas nos modelos com ponto de demanda agregados e desagregados, através da fórmula representada na Equação (3.6), para a qual:

m_A : conjunto de instalações abertas no problema agregado

m_D : conjunto de instalações abertas no problema desagregado

$f(m_A, c_{ij})$: função objetivo calculada utilizando m_A e o custo c_{ij}

$f(m_D, c_{ij})$: função objetivo calculada utilizando m_D e o custo c_{ij}

$$gap = \left[\frac{f(m_A, c_{ij}) - f(m_D, c_{ij})}{f(m_D, c_{ij})} \right] * 100 \quad (3.6)$$

Para cada problema de localização, analisa-se a diferença entre os resultados dos modelos em que os dados de entrada (pontos de demanda) estão desagregados e agregados, cujos valores são calculados com base na Equação (3.6). Caso a diferença seja nula, o valor da função objetivo tanto do modelo com pontos de demanda desagregado quanto do modelo com pontos de demanda agregados é igual, ou seja, não há diferença entre utilizar um método ou outro para solucionar o problema. Caso a diferença seja maior que zero (sinal positivo), o valor da função objetivo do modelo com pontos de demanda agregados é maior que a do modelo com pontos desagregados, indicando uma superestimativa da função objetivo ao agrupar os pontos demanda. Se o erro apresentar sinal negativo (menor que zero), ocorre uma subestimativa do valor da função objetivo do modelo com pontos de demanda desagregados, pelo modelo com pontos de demanda agregados. Para o modelo com pontos de demanda desagregados, não é calculada a diferença entre as soluções obtidas, uma vez que é redundante compará-lo consigo.

Além da comparação entre o valor da função objetivo da solução ótima encontrada para os problemas com pontos de demanda agregados e desagregados, analisa-se também no que difere a alocação de demanda e os custos de transporte a nível de município, buscando identificar as sobrestimativas e subestimativas em maior detalhe. A sobrestimativa ocorre quando os custos de transporte do modelo agregado são maiores que os do modelo desagregado em decorrência da ausência ou da maior distância entre instalações abertas e pontos de demanda. Por outro lado, a subestimativa ocorre quando os custos de transporte do modelo

agregado são menores que no modelo desagregado, em decorrência da abertura ou menor distância entre instalações abertas e pontos de demanda.

Com relação à análise a nível de município, busca-se identificar também características das regiões do espaço nas quais essas diferenças de alocação da demanda acontecem, com o propósito de responder qual seria a diferença na alocação da demanda em regiões com maior quantidade de pontos de demanda, como uma região metropolitana ou uma megacidade. Uma forma de identificar as regiões metropolitanas e megacidades de maneira mais automatizada e sem a necessidade de tantos conhecimentos prévios sobre a região é através do emprego do algoritmo DBSCAN (ESTER *et al.*, 1996).

O DBSCAN é um método de formação de agrupamentos, baseado na noção de densidade, ou seja, o quão próximos os pontos estão uns dos outros no espaço. A Figura 3.7 apresenta três bases de dados. Segundo Ester *et al.* (1996), o principal motivo pelo qual os agrupamentos nessa figura são identificados, é associado ao fato de que cada um possui uma densidade de pontos diferente dos demais, e que pontos que não pertencem a nenhum agrupamento (como observado em *database 3*) estão associados a regiões de menor densidade de pontos no espaço.

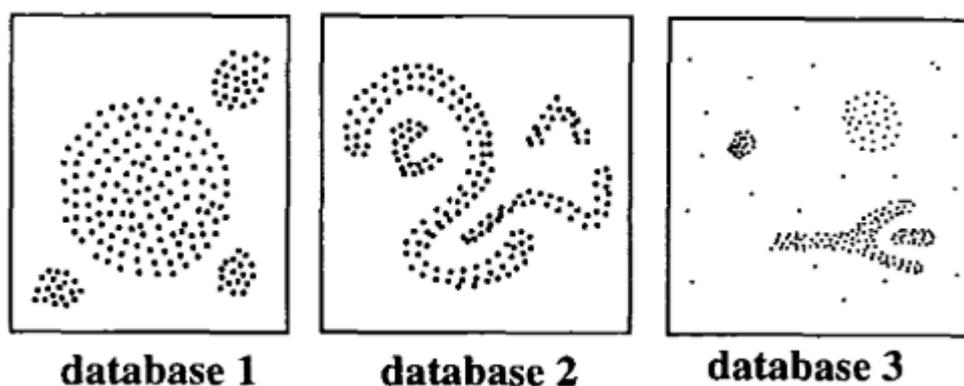


Figura 3.7 – Exemplos de bases de dados. Fonte: Ester *et al.* (1996)

Han, Kamber e Pei (2012) descrevem o algoritmo do DBSCAN da seguinte maneira: seja um local no espaço, no qual vários pontos se encontram e são classificados como não-visitados. O algoritmo DBSCAN aleatoriamente seleciona um ponto não-visitado p , alterando sua condição para visitado, e verifica se, numa distância máxima Eps existe uma quantidade mínima de pontos $MinPts$ na vizinhança do ponto p . O parâmetro $MinPts$ do DBSCAN representa a menor quantidade de pontos que juntos formem um agrupamento e deve ser definido de acordo com o problema a ser solucionado, assim como o parâmetro de distância Eps . Caso exista essa quantidade mínima de pontos, o DBSCAN cria um agrupamento; para

cada ponto na vizinhança do ponto p , sua condição é alterada para visitado e verifica-se a existência de outros pontos numa distância máxima Eps . Se esses pontos ainda não sejam parte do agrupamento, eles são adicionados ao agrupamento e repete-se o procedimento de visitá-los e buscar outros pontos a uma distância Eps a partir deles. Uma vez que não exista mais pontos a uma distância Eps dos pontos do agrupamento, o algoritmo escolhe, aleatoriamente, outro ponto não-visitado e repete o procedimento de busca. Se a não existir a quantidade $MinPts$ de pontos dentro de um raio de tamanho Eps a partir de um ponto, ele é classificado como ruído (no original *noise*).

A fim de se definir o parâmetro Eps , utiliza-se o método proposto por Rahmah e Sitanggang (2016), que consiste em calcular uma matriz de distância entre cada par de pontos do conjunto e selecionando, para cada um, as três menores distâncias. Em seguida, essas distâncias são ordenadas de maneira crescente com relação ao terceiro vizinho mais próximo e o ponto de inflexão desta curva é então utilizado como valor de Eps . A Figura 3.8 exemplifica a obtenção do valor do parâmetro conforme ponto de inflexão da curva.

A Figura 3.9 mostra as áreas urbanas com mais de cem mil habitantes para o estado de São Paulo, Brasil. Em azul, estão indicadas aquelas que possuem entre 100 mil e 300 mil habitantes, enquanto as destacadas em vermelho possuem mais de 300 mil habitantes. Por sua vez, a Figura 3.10 mostra a localização dos pontos de demanda no estado de São Paulo (SP), classificados em agrupamentos (pontos em vermelho) e ruídos (pontos em preto), após a classificação dos pontos de demanda com o algoritmo DBSCAN. Percebe-se que os pontos classificados como ruído estão associados a municípios tanto com menor quantidade de pontos de demanda quanto com menor população. Observa-se também que os agrupamentos identificados em vermelho estão associados às regiões urbanas apresentadas na Figura 3.10, indicando que o DBSCAN possibilita a identificação de grandes municípios e regiões metropolitanas sem a necessidade de conhecimento prévio da geografia local ou da definição de agrupamentos contíguos tais como regiões metropolitanas.

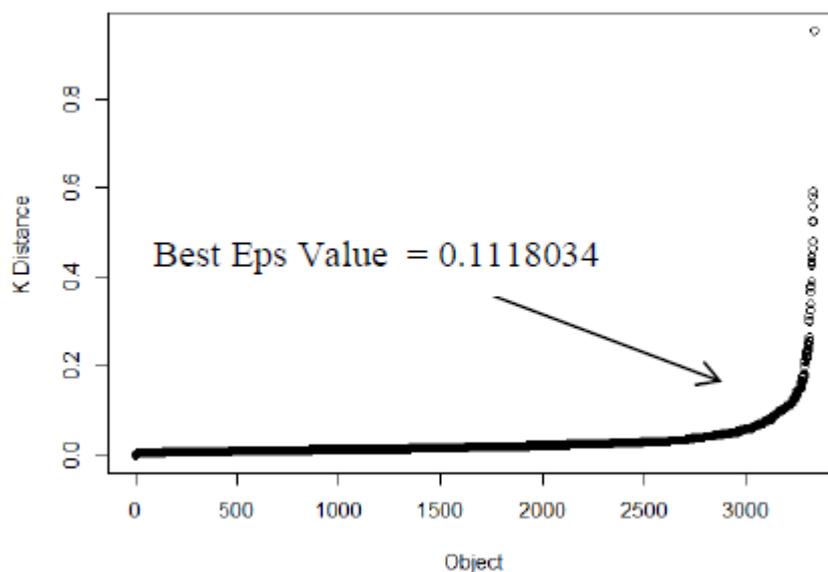


Figura 3.8 – Pontos ordenados conforme a distância para o terceiro vizinho mais próximo.
 Fonte: Rahmah e Sitanggang (2016)

Ao se empregar o DBSCAN, é possível que ocorra uma das seguintes situações (casos), para os quais busca-se entender se existem diferenças na alocação da demanda para nessas localidades:

- i. Todos os pontos de demanda existentes no município encontram-se em um mesmo agrupamento;
- ii. Existe mais de um agrupamento em um mesmo município, ou seja, dada a forma como os pontos de demanda se distribuem no espaço, os pontos de demanda não estão espacialmente próximos uns aos outros o suficiente para formarem um único agrupamento;
- iii. Existe mais de um município em um mesmo agrupamento, ou seja, existem regiões no espaço nas quais os pontos de demanda de dois municípios diferentes encontram-se próximos, de forma que componham um mesmo agrupamento juntos (este é um exemplo do que acontece em regiões metropolitanas e megacidades, uma vez que o efeito da conurbação dificulta a existência de localidades com menor densidade de pontos no espaço);
- iv. Em um dado município ocorre pelo menos um agrupamento cujos pontos de demanda estão todos localizados dentro da extensão territorial do município, como também pelo menos um agrupamento em que os pontos de demanda que o compõe se localizam em dois ou mais municípios, ou seja, equivalente a uma combinação das situações (ii) e (iii);

- v. Todos os pontos de demanda em um município são classificados como ruído, por não possuir a quantidade *MinPts* necessária para formar um agrupamento

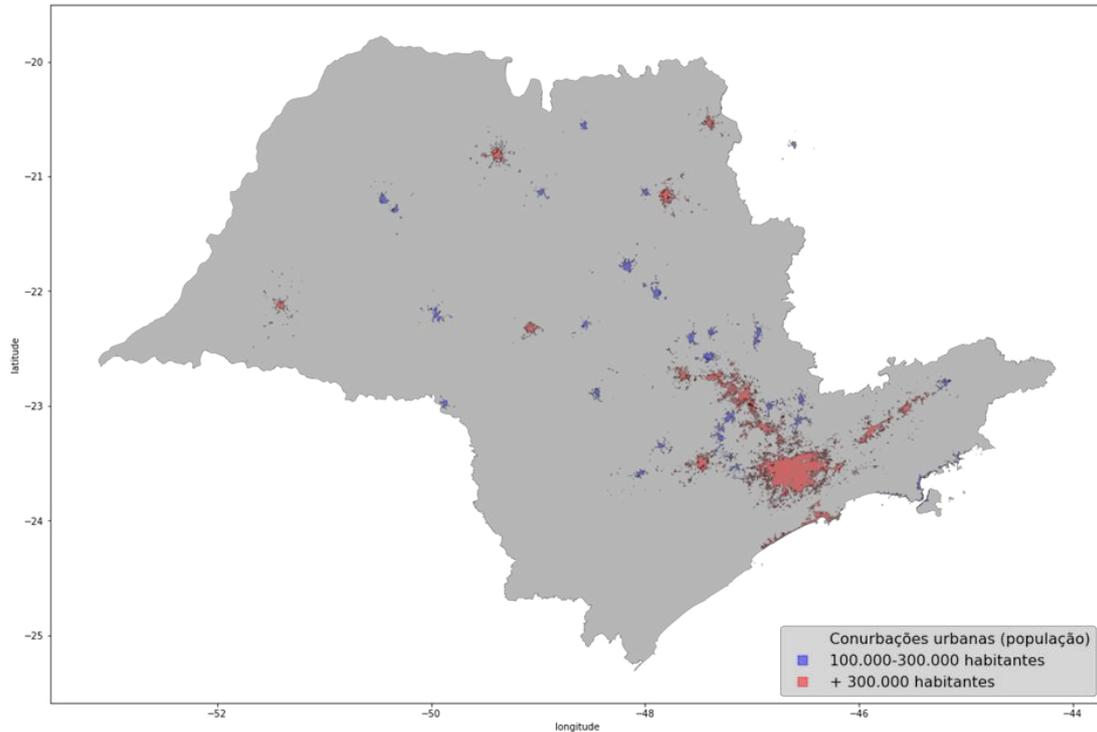


Figura 3.9 – Localização das aglomerações urbanas com mais de 100.000 habitantes no estado de São Paulo (SP)

Por último, calcula-se a distância média (D_{avg}) e a maior distância (D_{max}) entre os pontos de demanda e o local candidato mais próximo para atender a demanda. Para os modelos agregados, as distâncias são calculadas após a etapa de desagregação. Segundo Peeters e Thomas (2000), a distância média é uma medida de eficiência da solução ótima, enquanto a maior distância é uma medida de equidade, ou seja, do quão distantes da instalação aberta mais próxima encontram-se os pontos de demanda.

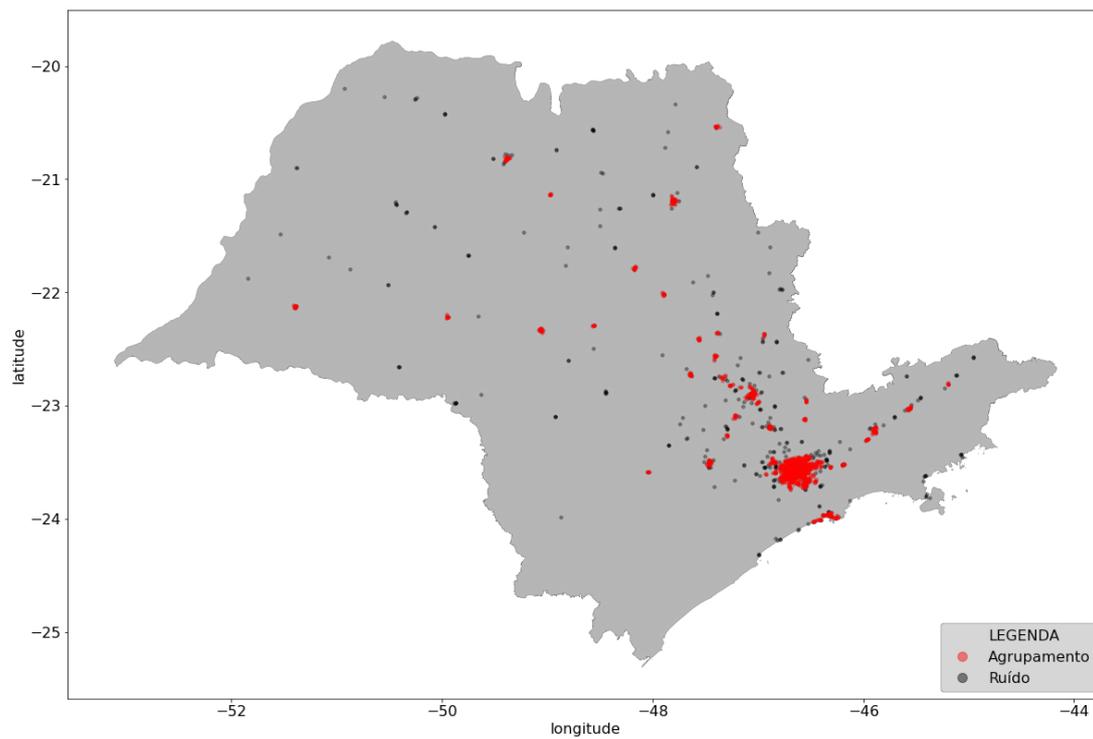


Figura 3.10 – Localização dos agrupamentos e ruídos considerando pontos de demanda no estado de São Paulo (SP)

4. INSTÂNCIAS

Neste capítulo são apresentadas as instâncias criadas para esta pesquisa. Foram geradas 18 instâncias do FCFL, a partir de dois conjuntos de dados que correspondem a uma operação real de distribuição B2B nos estados das regiões sul e sudeste, e parte dos estados da região nordeste do país.

4.1. Pontos de demanda e locais candidatos

Os dois conjuntos de dados resultantes são denominados Conjunto 1 e Conjunto 2. O Conjunto 1 possui 5.574 pontos de entrega, em 650 municípios, e uma demanda total de 633.265 unidades. Já o Conjunto 2, é uma variação do Conjunto 1, criada com a finalidade de gerar mais diversidade e compreende 585 municípios, 4.460 pontos e demanda total de 498.872 unidades de produto. A criação do Conjunto 2 possibilita que possíveis vieses na obtenção de localização e de estimativa de demanda fossem removidos, respaldando os resultados obtidos para o Conjunto 1.

É importante destacar que embora existam instâncias para problemas de localização comumente utilizadas na literatura científica, tais como, por exemplo, a *OR-library* (BEASLEY, 1990) e o conjunto M^* (KRATICA *et al.*, 2001), optou-se por desenvolver as instâncias utilizadas na pesquisa para retratar uma operação real, em uma grande área com densidades de pontos de demanda variáveis no espaço, visto que são baseadas em localizações reais para os pontos de demanda, o que possibilita a aplicação do método de agrupamento de pontos de demanda por município, essencial para o trabalho.

Para a construção do Conjunto 1, foram consideradas as localizações dos pontos de venda de algumas das maiores empresas do ramo farmacêutico do Brasil. Essas localizações foram selecionadas com base nas unidades da federação nas quais se localizavam. Neste caso, a população de cada estado, de acordo com os dados do IBGE (IBGE, 2011), foi ordenada de maneira decrescente e, aqueles pontos de venda localizados em estados cuja população representasse ao menos 80% da população do país e que fossem contíguos foram selecionados. Desta forma, ficaram excluídos estados das regiões Centro-Oeste e Norte, e os estados do Maranhão e do Piauí no Nordeste.

Os endereços das farmácias dos grupos Raia-Drogasil, Drogaria São Paulo e Pacheco, Extrafarma, Pannel, Farmácias Pague Menos, Drogaria Araújo e Farmácias São João foram obtidos em seus respectivos sites, e, a partir do endereço obtido, utilizou-se a *Places API* do *Google Maps* (GOOGLE, 2022) a fim de se obter o par de coordenadas (latitude e longitude)

mais próximas ao endereço em questão. Nesta etapa, também foram considerados pontos de vendas dessas redes de farmácias em outros estados que não constam nas instâncias.

Para cada empresa, os pontos de demanda são considerados do mesmo tipo, sem diferenciação exceto pela demanda a ser abastecida; suas demandas, representadas pelo número de unidades de produto, foram estimadas como sendo proporcionais ao número de habitantes que se encontram mais próximos a cada ponto. Para determinar a região de cada cidade mais próxima a cada ponto de demanda, foram utilizados polígonos de Voronoi⁴, obtidos pelo módulo *Grass*, do *software* QGIS (GEOSPATIAL FOUNDATION PROJECT, 2020). Também foi considerado que cada ponto de venda atenda a uma área com raio máximo de 3 km, uma distância duas vezes maior que a citada por Hofer *et al.* (2020), considerando que uma pessoa estaria disposta a andar mais para ter acesso a produtos para sua saúde. A fim de determinar o número de habitantes próximos a cada ponto, utilizou-se as informações da Grade Estatística do Instituto Brasileiro de Geografia e Estatística (IBGE, 2016), que contém dados populacionais do censo realizado no país em 2010.

A Figura 4.1 exemplifica uma sobreposição de camadas que determinam o número de habitantes mais próximos de um ponto de demanda. Os polígonos de Voronoi são delimitados pelas linhas vermelhas; o raio de caminhada é indicado pelo círculo laranja, enquanto as grades estatísticas do IBGE são representadas pelos quadrados azuis. São mostrados dois tamanhos de quadrados azuis, sendo que o menor representa uma área de 200m x 200m, enquanto o maior, uma área de 1km x 1km, indicando regiões com diferentes densidades populacionais dentro de um mesmo polígono. Os espaços em branco na figura representam regiões no espaço nas quais as grades estatísticas do IBGE não identificaram nenhum habitante.

Ao sobrepor essas camadas, observa-se que algumas quadrículas da grade podem ser alocadas a dois ou mais pontos de venda. Para evitar a divisão da demanda entre dois pontos de venda, optou-se por alocar toda a população representada pela célula da grade ao ponto de demanda de cuja área é mais significativa na grade. A Figura 4.2 ilustra um exemplo dessa situação. Nesse caso, a população da quadrícula vermelha, em destaque, foi alocada ao ponto 1.558, ao invés do ponto 863, uma vez que maior parte da área da grade está associada ao ponto 1.558.

⁴ O leitor é referido ao APÊNDICE A para uma breve contextualização e exemplos de aplicação dos polígonos de Voronoi.

Para cada rede de farmácias, foram divididas 100.000 unidades de produto entre seus respectivos pontos de demanda, de maneira proporcional à população atendida por eles. Ou seja, pontos de venda que atendiam a uma quantidade maior de habitantes receberiam uma quantidade maior de produtos por isso. Também se definiu que cada ponto de venda receberia, ao menos, dez unidades de produto.

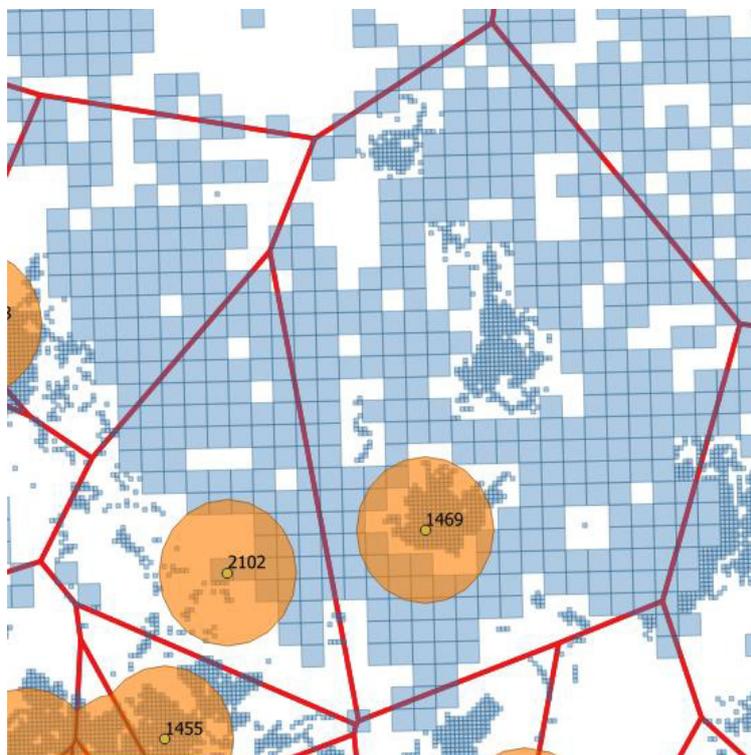


Figura 4.1 – Polígono de Voronoi que delimita a região do espaço mais próxima a um ponto e sua respectiva área de influência

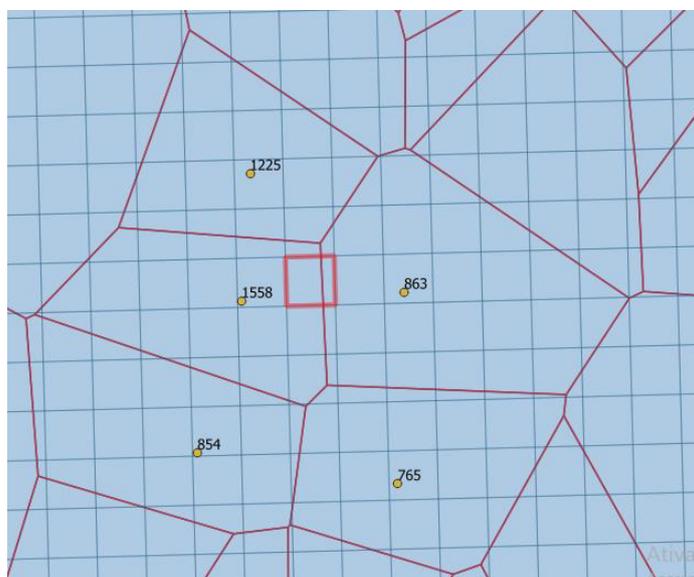


Figura 4.2 – Exemplo de alocação de demanda de uma grade

Sabendo-se que algumas redes de farmácias, dentre as quais, Raia-Drogasil, Drogaria São Paulo e Pacheco, e Extrafarma, também possuem filiais (pontos de demanda) em cidades de outros estados distintos desses presentes no Conjunto 1, optou-se por também incluir essas filiais na divisão das unidades de produto, mas não as considerar na formação do conjunto. Por esse motivo, a quantidade de unidades de demanda do Conjunto 1 mencionada anteriormente é de 633.265 unidades, e não 700.000, como seria esperado, ou seja, mais de 90% do total de unidades está representado pelos pontos existentes no Conjunto 1.

Os 150 locais candidatos a instalação não são parte do conjunto original de dados e foram definidos endogenamente com base nos pontos de demanda existentes em cada município. Inicialmente, os municípios são separados em pertencentes ou não a cidades-sede que representem regiões metropolitanas ou megacidades. As cidades-sede são os municípios mais importantes, geralmente capitais, que compõem a região metropolitana. Como regiões metropolitanas e megacidades foram considerados os municípios de Belo Horizonte, Porto Alegre, Salvador, São Paulo e Rio de Janeiro. Para os municípios de São Paulo e Rio de Janeiro, que apresentam maior número de pontos de venda, buscou-se definir 12 locais candidatos, uma proporção de aproximadamente 1 candidato para cada 20 pontos de demanda. Por sua vez, nos municípios de Belo Horizonte, Porto Alegre e Salvador, buscou-se definir 6 locais candidatos, respeitando uma proporção de 1 local candidato para cada 10 pontos de demanda. Assim, para cada região metropolitana, foi resolvido um problema de p -medianas diferente, no qual o número de medianas era igual ao número de instalações que se desejava localizar.

Para cada município não que não é uma cidade-sede, foi calculado o centroide com base na localização de todos os pontos de demanda do município. Neste caso, a demanda de cada ponto não é levada em consideração. Após obter todos os centroides, os municípios foram novamente divididos entre pertencentes ou não a uma das regiões metropolitanas. Para os municípios nas regiões metropolitanas de São Paulo e Rio de Janeiro, foram resolvidos outros problemas de p -medianas com p igual a 8, completando 20 locais candidatos; enquanto em Belo Horizonte, Porto Alegre e Salvador, os problemas de p -medianas foram resolvidos considerando p igual a 7, totalizando 15 locais candidatos. Ressalta-se que por não existir demanda em ao menos 9 municípios da região metropolitana de Salvador, esta possui 11 locais candidatos ao invés de 15.

Sabendo-se que o total de candidatos nas regiões metropolitanas é igual a 81, foi solucionado um problema de p -medianas, com p igual a 79, de maneira que os 79 centroides mais centrais foram selecionados para compor o restante do conjunto de locais candidatos.

A Tabela 4.1 apresenta o número de locais candidatos por estado e região metropolitana, bem como a quantidade de municípios às quais esses candidatos estão associados. É possível observar que somente nas regiões metropolitanas o número de locais candidatos é maior que a quantidade de municípios. Destaca-se a RM de Salvador, que possui 11 locais candidatos, ao invés de 15. Para totalizar 15 locais candidatos, seria necessário haver pontos de demanda em ao menos 9 municípios da região metropolitana de Salvador; todavia, apenas em apenas 5 municípios se localizavam pontos de demanda para o cálculo do centro de gravidade, de forma que para manter a quantidade de candidatos na cidade de Salvador igual a Belo Horizonte e Porto Alegre, não foram sorteados mais pontos de demanda, e manteve-se uma quantidade inferior de candidatos na RM.

Tabela 4.1 – Locais candidatos

| UF/ Reg. Metropolitana | No. Municípios | No. Candidatos | UF/ Reg. Metropolitana | No. Municípios | No. Candidatos |
|------------------------|----------------|----------------|------------------------|----------------|----------------|
| RM São Paulo | 9 | 20 | CE | 5 | 5 |
| RM Rio de Janeiro | 9 | 20 | PE | 5 | 5 |
| SP | 15 | 15 | ES | 3 | 3 |
| RM Belo Horizonte | 10 | 15 | RJ | 2 | 2 |
| RM Porto Alegre | 10 | 15 | SE | 2 | 2 |
| RS | 14 | 14 | AL | 1 | 1 |
| RM Salvador | 6 | 11 | PB | 1 | 1 |
| MG | 7 | 7 | RN | 1 | 1 |
| PR | 7 | 7 | SC | 1 | 1 |
| BA | 5 | 5 | | | |

A Figura 4.3 mostra a localização dos candidatos a instalação que são utilizados tanto no Conjunto 1 quanto no 2. Cada local candidato é representado por um triângulo de cor laranja, e existe ao menos um candidato por unidade da federação, como no caso do Rio Grande do Norte. Nota-se que algumas regiões, como o Triângulo Mineiro, norte de Minas Gerais, interior de Santa Catarina e da Bahia não possuem locais candidatos, em decorrência dos municípios ali localizados não serem centrais em relação aos demais municípios do Conjunto 1.

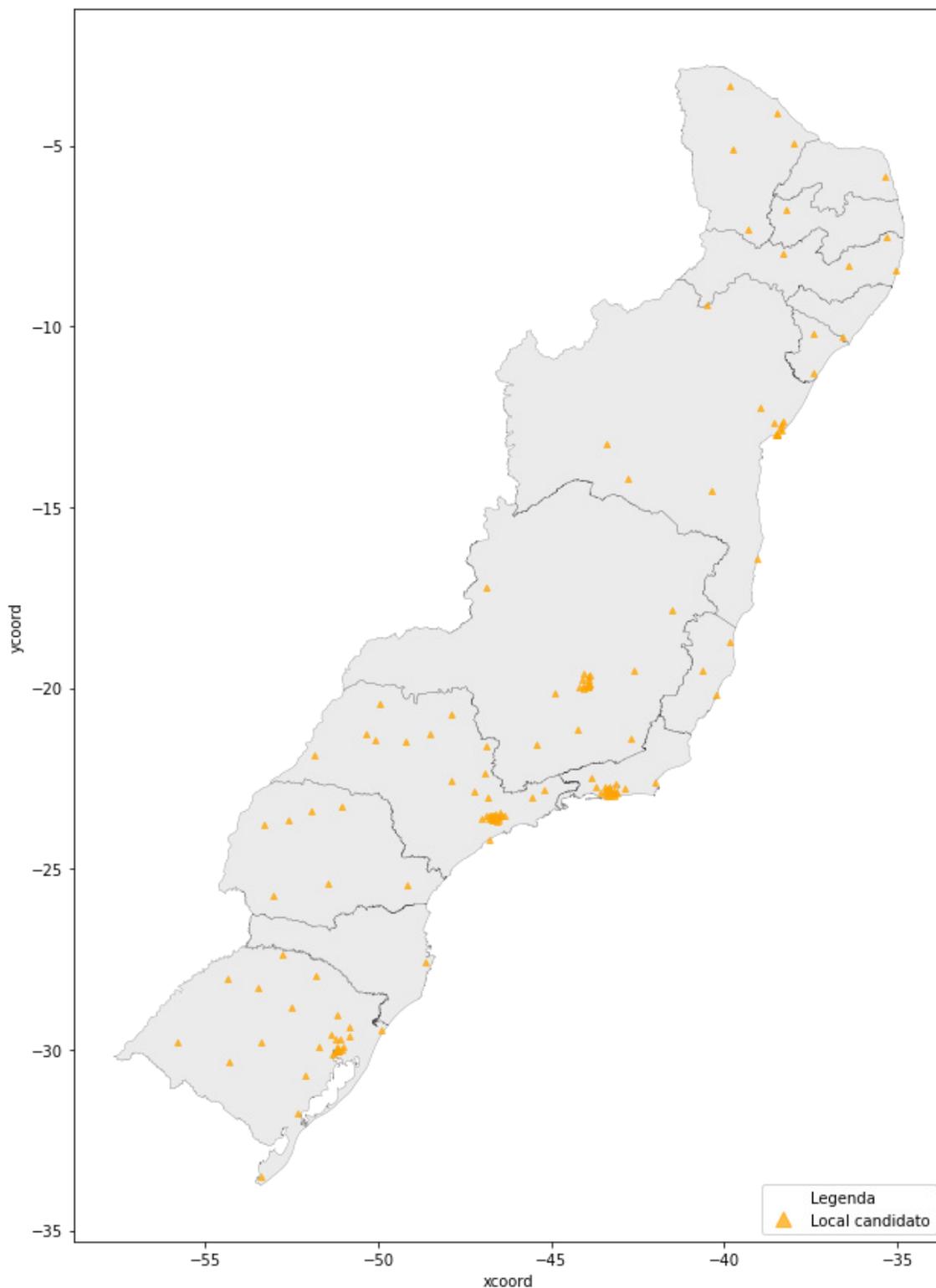


Figura 4.3 – Localização dos candidatos a instalação

Os pontos de demanda do Conjunto 1 são mostrados na Figura 4.4. A maioria deles encontra-se em regiões dos estados pertencentes às regiões sudeste e sul, como esperado dado o número de pontos de demanda e municípios presentes nessas regiões. Observa-se que em regiões como o norte de Minas Gerais, noroeste da Bahia, interior do Paraná e de Santa

Catarina, existem vazios, indicando a ausência de pontos de demanda dessas redes, possivelmente por serem regiões mais afastadas se comparadas às demais.

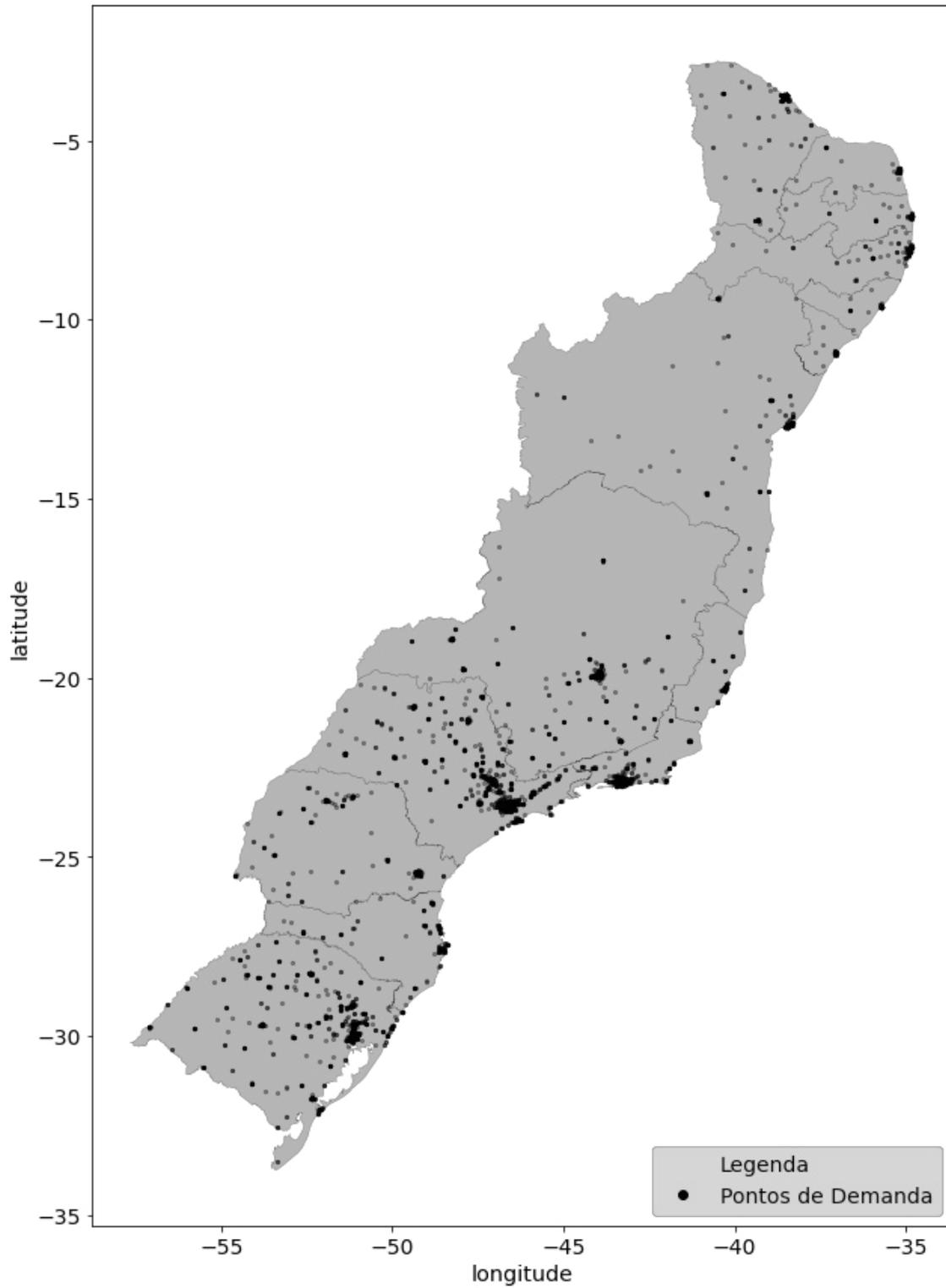


Figura 4.4 – Pontos de demanda do Conjunto 1

Na Tabela 4.2 são mostradas a demanda, a porcentagem do total que ela representa, o número de pontos de demanda e de municípios por unidade da federação no Conjunto 1. Como esperado, São Paulo é o estado com maior demanda e número de pontos de demanda, por se tratar do estado mais populoso do país. Todavia, é possível ver que o Rio Grande do Sul possui a segunda maior demanda e quantidade de pontos de demanda, o que não é esperado, uma vez que existem outros estados, como Rio de Janeiro e Minas Gerais, que são mais populosos. Isso acontece pelo fato da rede de Farmácia São João ser presente nos estados da região sul, em especial o Rio Grande do Sul, que pode ter afetado a abertura de candidatos mais próximos.

Tabela 4.2 – Informações sobre os estados no Conjunto 1

| Estado (sigla) | Pontos de Demanda | Municípios | Demanda | % da Demanda |
|----------------|-------------------|------------|---------|--------------|
| SP | 1.926 | 158 | 147.585 | 23,31% |
| RS | 1.005 | 148 | 130.148 | 20,55% |
| MG | 577 | 68 | 129.458 | 20,44% |
| RJ | 493 | 36 | 37.115 | 5,86% |
| CE | 308 | 38 | 29.924 | 4,73% |
| BA | 264 | 42 | 30.251 | 4,78% |
| PR | 259 | 46 | 45.186 | 7,14% |
| PE | 219 | 33 | 25.482 | 4,02% |
| SC | 175 | 34 | 30.526 | 4,82% |
| ES | 106 | 10 | 7.226 | 1,14% |
| RN | 70 | 11 | 8.644 | 1,36% |
| PB | 70 | 13 | 5.935 | 0,94% |
| SE | 54 | 7 | 2.538 | 0,40% |
| AL | 48 | 6 | 3.247 | 0,51% |
| Total | 5.574 | 650 | 633.265 | |

O histograma da distância entre pontos de demanda e os respectivos locais candidatos mais próximos para o Conjunto 1 é mostrado na Figura 4.5. Observa-se que há mais de 5.000 pontos de demanda que se encontram até 100 km de distância dos candidatos mais próximos; por outro lado, nota-se também que existem pontos de demanda estão a uma distância de aproximadamente 280 km do candidato mais próximo, sugerindo que são pontos de demanda em regiões com menos número de pontos de demanda e conseqüentemente mais afastadas dos locais candidatos mais próximos para atender a demanda.

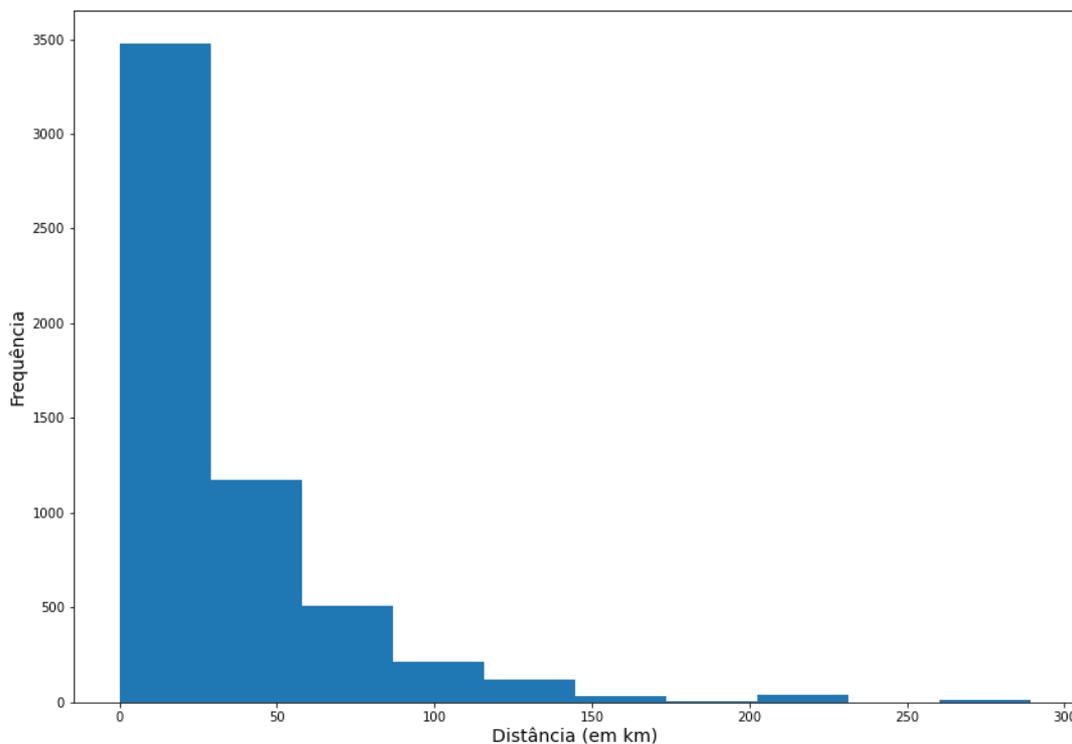


Figura 4.5 – Histograma da distância entre pontos de demanda e respectivos locais candidatos mais próximos para o Conjunto 1

Como mencionado anteriormente, o Conjunto 2 corresponde a uma variação do Conjunto 1, criado para gerar maior diversidade. Para sua construção, foram selecionados aleatoriamente cerca de 80% do número total de pontos de demanda do Conjunto 1, e a respectiva demanda foi multiplicada por um fator aleatório que varia entre 0,60 e 1,40. A localização dos pontos de demanda sorteados e que compõem o Conjunto 2 é mostrada na Figura 4.6. A princípio, observa-se que a maioria dos pontos de demanda se concentra nos estados das regiões sudeste e sul, em especial nas regiões metropolitanas de suas respectivas capitais. Também é possível ver que regiões, como o interior de Santa Catarina e do Paraná e o noroeste da Bahia permanecem com poucos pontos de demanda e mais afastados dos demais. Se comparada à Figura 4.4, observa-se na Figura 4.6 que alguns municípios nessas regiões não foram sorteados para compor o Conjunto 2.

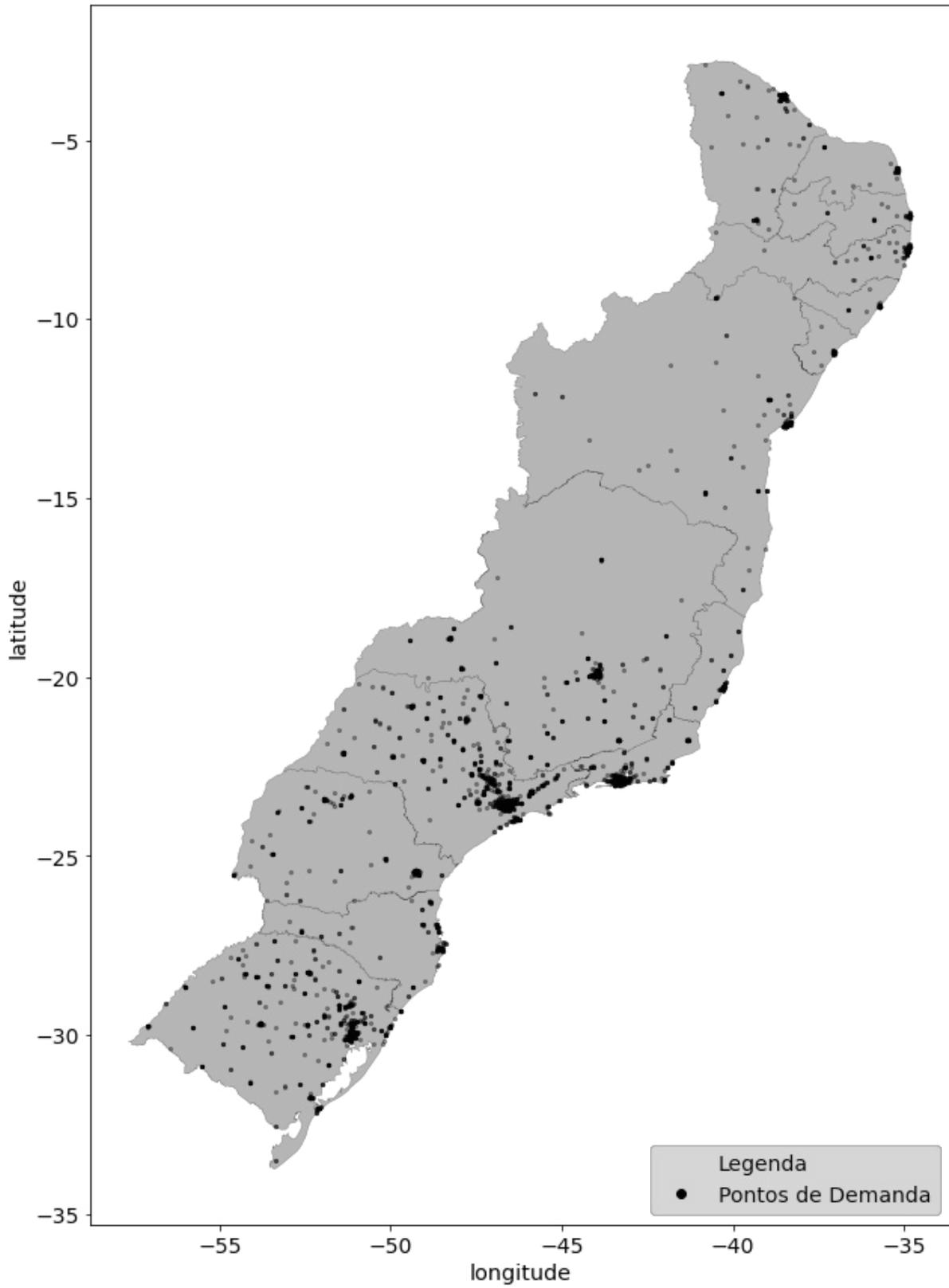


Figura 4.6 – Pontos de demanda do Conjunto 2

A Tabela 4.3 mostra o número de pontos de demanda, municípios, e de unidades de demanda e qual a porcentagem que essa demanda representa por estado para o Conjunto 2. A demanda total do Conjunto 2 é equivalente a 78,78% da demanda total do Conjunto 1. Mesmo multiplicando a demanda dos pontos por um fator aleatório, observou-se que os fatores que multiplicam a demanda por um valor menor que 1,0 foram mais frequentes, o que retorna um valor diferente da média (80% para o conjunto), uma vez que existe a probabilidade de que um valor diferente da média seja retornado, mesmo em sorteios honestos.

Tabela 4.3 – Informações sobre os estados no Conjunto 2

| Estado (sigla) | Pontos de Demanda | Municípios | Demanda | % da demanda |
|----------------|-------------------|------------|---------|--------------|
| SP | 1.555 | 152 | 112.412 | 22,53% |
| RS | 801 | 141 | 104.484 | 20,94% |
| MG | 461 | 58 | 101.142 | 20,27% |
| RJ | 405 | 34 | 30.581 | 6,13% |
| CE | 247 | 30 | 25.298 | 5,07% |
| BA | 218 | 38 | 24.984 | 5,01% |
| PR | 208 | 40 | 36.193 | 7,25% |
| PE | 165 | 26 | 20.752 | 4,16% |
| SC | 130 | 27 | 21.034 | 4,22% |
| ES | 90 | 10 | 6.525 | 1,31% |
| PB | 52 | 10 | 4.433 | 0,89% |
| RN | 50 | 9 | 6.683 | 1,34% |
| SE | 43 | 6 | 2.021 | 0,41% |
| AL | 35 | 4 | 2.330 | 0,47% |
| Total | 4.460 | 585 | 498.872 | |

A Figura 4.7 mostra o histograma da distância com relação ao candidato mais próximo para os pontos de demanda do Conjunto 2. Nota-se que aproximadamente 4.000 dos 4.460 pontos de demanda encontram-se a uma distância de até 100 km do local candidato mais próximo, e que existem pontos de demanda com uma distância aproximada de 280 km do local candidato aberto mais próximo, um indício de que alguns pontos de demanda em regiões mais distantes dos locais candidatos também foram sorteadas como parte do Conjunto 2.

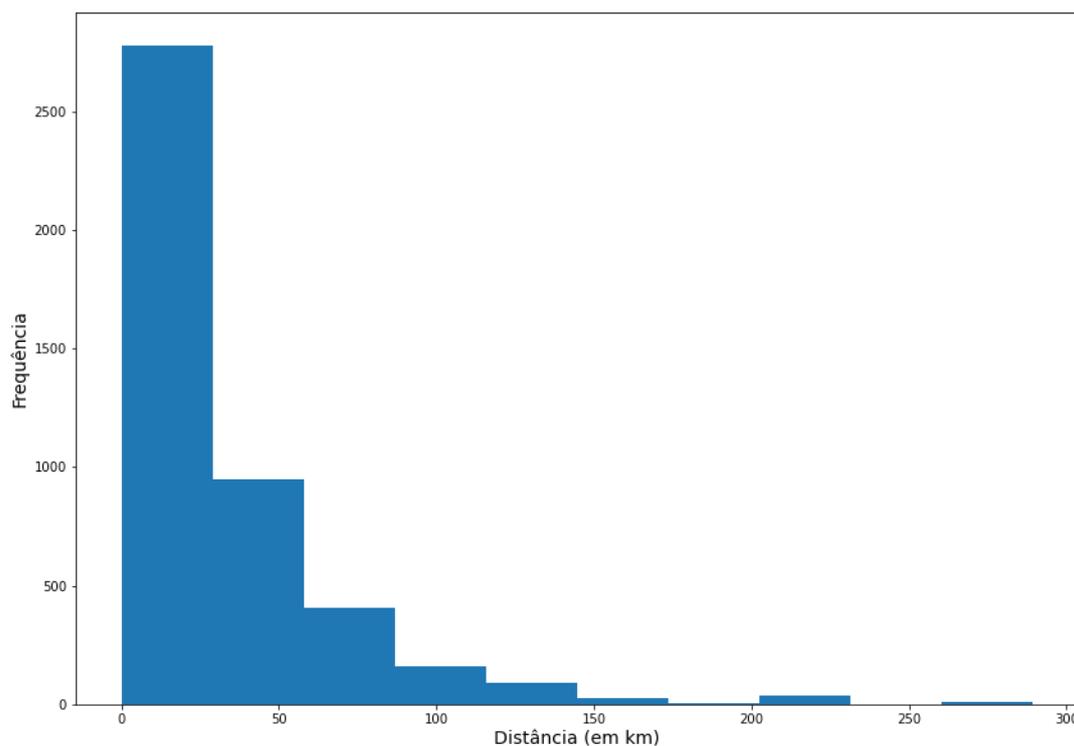


Figura 4.7 – Histograma da distância entre pontos de demanda e locais candidatos mais próximos do Conjunto 2

4.2. Custos

Para estimar o custo fixo de operação mensal das instalações, considerou-se um valor de R\$45,00/m² para imóveis comerciais, que é proposto pelo índice Fipe/ZAP (BUENO, 2019), para um imóvel de 200 m², totalizando um valor base de R\$ 9.000, aproximadamente. A área do imóvel foi escolhida como 200m² uma vez que é comparável à área mínima de 100m² citada por Rodero (2018) para instalações típicas de farmácias. A partir deste valor, outros foram escolhidos para que as variações de custo fixo na instância fossem analisadas, mantendo o custo de transporte constante. Foram consideradas as seguintes variações: multiplicação em 2, 5 e 10 vezes do valor base, e divisão por 2, 5, 10, 100 e 1.000 vezes do valor base. Optou-se por dividir o valor base mais vezes uma vez que seriam obtidas soluções com maior número de instalações abertas, o que não aconteceria caso o valor base fosse multiplicado por 100 e 1.000 vezes, soluções nas quais o número de candidatos abertos possivelmente seria 1.

O custo variável unitário de transporte foi definido com base na tabela de fretes da ANTT (BRASIL, 2021) e, para diferenciar o custo de transporte em regiões metropolitanas e/ou cidades muito próximas umas às outras, foram considerados dois tipos de custos: um para distâncias de até 40 km e outro para distâncias maiores. Sabendo-se que a tabela de fretes da ANTT também considera um custo de mão-de-obra para efetuar as atividades de carga e

descarga de mercadorias, optou-se por considerar este valor como parte do custo unitário de transporte, de maneira que este seja calculado na forma $c_{ij} = b + ax$. O custo do deslocamento R\$/km por unidade (associado ao termo dependente ax) foi definido como R\$0,0020412, se a distância entre o local candidato e ponto de venda for superior a 40 km, e metade deste valor, caso esteja a menos de 40 km, levando em consideração que existem locais candidatos em regiões metropolitanas ou em municípios nos quais existam pontos de demanda e que o custo de transporte, dada a proximidade, deve ser menor. Adicionalmente, o custo de carga e descarga (termo independente b) é dividido entre todas as unidades que vão para o mesmo ponto de venda, sendo somada ao custo do deslocamento calculado anteriormente, no valor de R\$0,094 por unidade para distâncias superiores a 40 km, e metade desse valor para distâncias inferiores a 40 km. Para este cálculo, foi considerada uma frota homogênea, composta por veículos genéricos de dois eixos, com capacidade para 1.000 unidades, que transporta carga geral, e se enquadra como transporte rodoviário de carga lotação de alto desempenho. Optou-se por considerar um veículo com esta capacidade genérica para simplificar os custos apresentados, de tal forma que o valor do termo independente b ficasse na mesma unidade de custo (R\$/unidade).

4.3. Síntese

A Tabela 4.4 sintetiza as características das instâncias construídas para o conjunto de dados analisado. Cada instância se refere a um valor diferente de custo fixo de operação das instalações e foram rotuladas de acordo com esse custo. Por exemplo, a instância 1-9 representa o Conjunto 1, quando o custo fixo mensal é de R\$ 9, enquanto a instância 2-4500 representa o Conjunto 2, quando o custo fixo mensal é de R\$ 4.500 A utilização de diferentes valores de custo fixo permite avaliar situações nas quais a empresa tem pouco (ou quase nenhum) custo para operar o local candidato, tanto quanto situações nas quais a operação do local candidato é mais onerosa.

Por fim, para calcular a matriz de distância entre locais candidatos e pontos de demanda (ou agrupamentos), foi utilizada a biblioteca OSMnx (BOEING, 2017), em linguagem de programação Python. Buscando minimizar os erros no cálculo da matriz de distância entre locais candidatos e pontos de venda, foi considerada a distância em linha reta entre cada par de pontos, corrigida com base na curvatura da superfície terrestre.

Tabela 4.4 – Características das instâncias solucionadas

| | Instância | Pontos de Demanda | Número de Municípios | Custo Fixo (R\$) | Custo de Transporte (R\$/km/unidade) |
|------------|------------|-------------------|----------------------|------------------|---|
| Conjunto 1 | 1-9 | | | 9 | |
| | 1-90 | | | 90 | |
| | 1-900 | | | 900 | |
| | 1-1800 | | | 1.800 | |
| | 1-4500 | 5.574 | 650 | 4.500 | |
| | 1-9000 | | | 9.000 | |
| | 1-18000 | | | 18.000 | |
| | 1-45000 | | | 45.000 | $0,047 + 0,0010206x$, se x for uma distância de até 40 km, e $0,094 + 0,0020412x$, se x for uma distância maior que 40 km |
| | 1-90000 | | | 90.000 | |
| | Conjunto 2 | 2-9 | | | 9 |
| 2-90 | | | | 90 | |
| 2-900 | | | | 900 | |
| 2-1800 | | | | 1.800 | |
| 2-4500 | | 4.460 | 585 | 4.500 | |
| 2-9000 | | | | 9.000 | |
| 2-18000 | | | | 18.000 | |
| 2-45000 | | | | 45.000 | |
| 2-90000 | | | | 90.000 | |

5. EXPERIMENTOS COMPUTACIONAIS

Este capítulo apresenta os resultados dos experimentos computacionais que foram realizados no âmbito da presente pesquisa. Aplicando a metodologia proposta no Capítulo 3, este capítulo busca:

- Definir o impacto, na precisão e qualidade das soluções, da aplicação de um método de agrupamento de pontos de demanda em um modelo de otimização de problemas de localização de instalações não capacitado (FCFL);
- Identificar as regiões no espaço nas quais pode existir diferença na alocação da demanda e quanto essa diferença pode representar localmente;
- Identificar características das instâncias que tornam um método de agrupamento de pontos de demanda mais adequado para a simplificação da modelagem de problemas de localização de instalações.

O capítulo se subdivide em três seções, nas quais são tratados (i) os agrupamentos obtidos através da aplicação dos métodos apresentados para o conjunto de dados; (ii) os resultados obtidos para cada instância; e (iii) a análise da qualidade das soluções obtidas para os modelos com pontos de demanda agregados.

5.1. Formação de agrupamentos para a modelagem

Os agrupamentos de pontos de demanda a serem atendidos pelas instalações que se pretende localizar são determinados através dos métodos apresentados anteriormente na Seção 3.2. Nas próximas subseções, são apresentados, para os Conjuntos 1 e 2, os agrupamentos obtidos aplicando-se os Métodos Ma1 e Ma2. Para todas as instâncias, cada agrupamento é representado por um ponto que se localiza no local que corresponde ao baricentro.

Para formar os agrupamentos, foi utilizado o módulo *K-Means* da biblioteca Scikit-Learn (PEDREGOSA *et al.*, 2011), utilizando a versão 3.8 da linguagem de programação Python, além das bibliotecas Pandas e Numpy.

5.1.1. Conjunto 1

Para o Conjunto 1 são criados 650 agrupamentos, uma vez que o número de agrupamentos é igual ao número de municípios existentes no conjunto. Ao aplicar o Método Ma1, todos os pontos de demanda de um município são representados por um único ponto localizado no baricentro desses pontos. A Figura 5.1 mostra a localização dos 650 baricentros para os pontos de demanda do Conjunto 1. A área de cada circunferência é proporcional à

demanda de cada município. Observa-se que as circunferências de maior área estão associadas às regiões metropolitanas, e às respectivas capitais dos estados. Há uma maior presença de cidades com maior demanda próximas à faixa litorânea, do que no interior, o que é esperado, tendo em vista a distribuição geográfica da população brasileira.

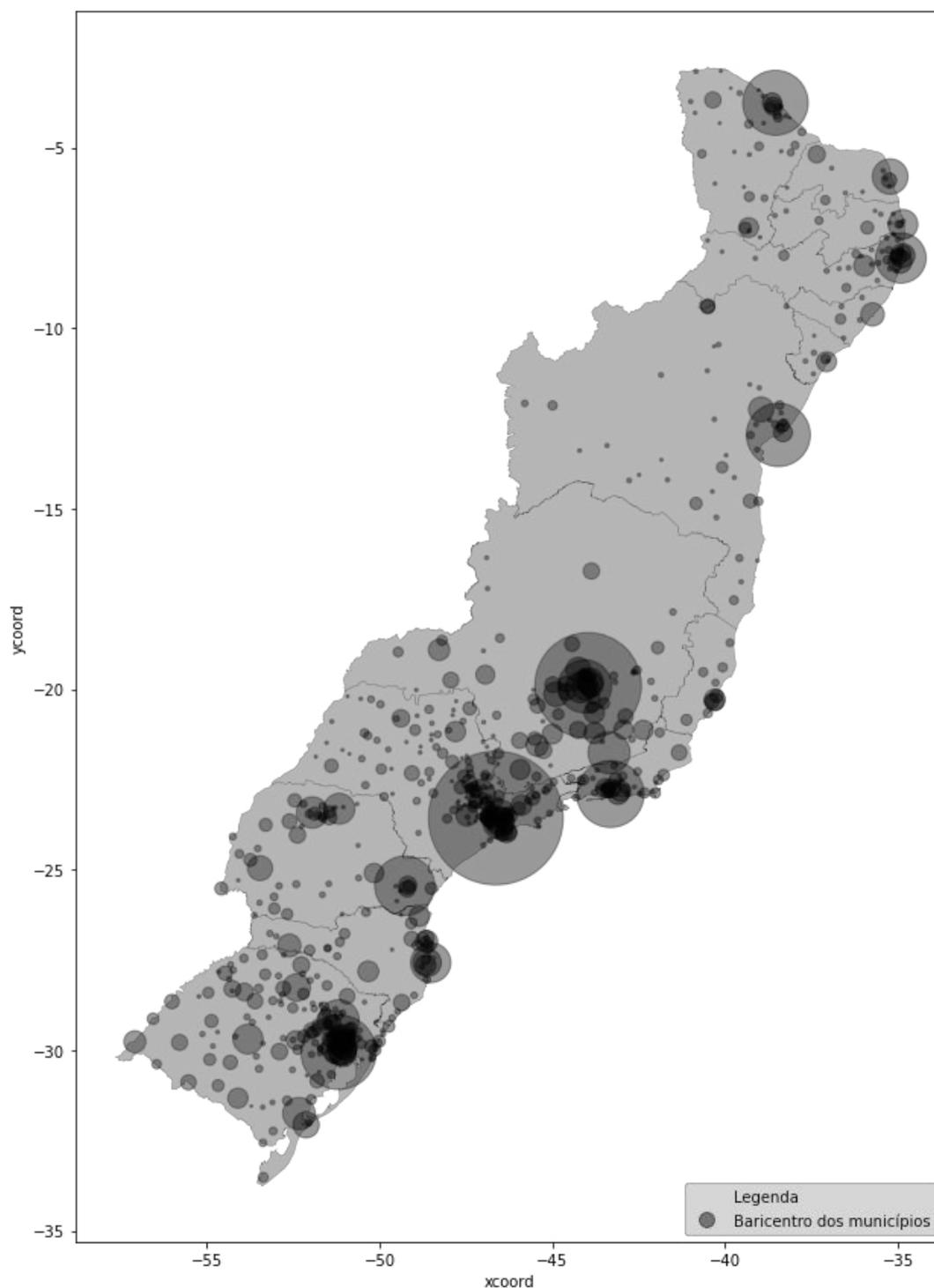


Figura 5.1 – Localização dos baricentros dos municípios e suas respectivas demandas para os pontos de demanda do Conjunto 1

Por sua vez, o Método Ma2 emprega o algoritmo *K-Means* para a definição dos agrupamentos, em que o parâmetro K correspondente ao número total de municípios assume o valor de 650. A Figura 5.2 mostra a localização dos baricentros dos agrupamentos obtidos através do Método Ma2, sendo que a área das circunferências representa a demanda de cada agrupamento.

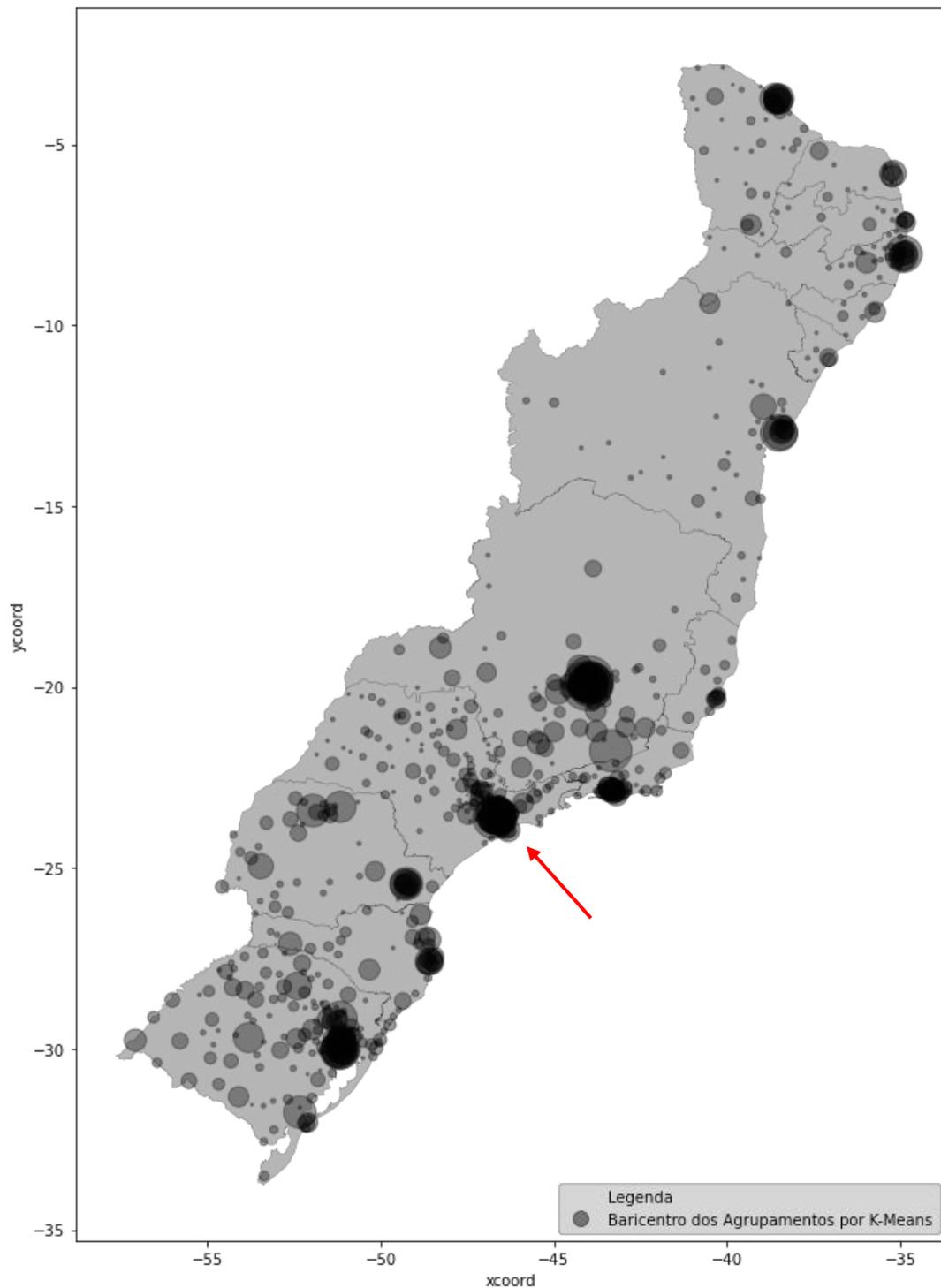


Figura 5.2 – Localização dos baricentros dos agrupamentos pelo método Ma2 e suas respectivas demandas para os pontos de demanda do Conjunto 1

Ao analisarmos conjuntamente as Figuras 5.1 e 5.2, observa-se que as diferenças com relação ao Método Ma1 são mais perceptíveis em localidades com maior demanda, como na RM de São Paulo, indicada pela seta em vermelho na figura, que possui circunferências de menor área que aquelas da Figura 5.1. Isso significa que os agrupamentos formados pelo Método Ma2 possuem menor demanda se comparados aos do Método Ma1, em decorrência do método de formação de agrupamento utilizado, uma vez que o Método Ma2 forma agrupamentos com menor número de pontos de demanda, com base em sua localização geográfica, o que possibilita que exista mais de um agrupamento em um mesmo município.

A Tabela 5.1 mostra o tamanho dos modelos matemáticos resultantes a partir do Conjunto 1, assim como a quantidade de linhas, colunas e valores diferentes de zero das matrizes correspondentes. O tamanho dos modelos matemáticos resolvidos é decorrente do número de instalações e do número de pontos de demanda, ou seja, são 150 locais candidatos nos modelos agregado e desagregado, e 5.547 pontos de demanda no modelo desagregado e 650 nos dois modelos agregados, que corresponde ao número de agrupamentos. Nota-se que a quantidade de variáveis no modelo desagregado é 8,57 vezes maior que nos modelos agregados, o que é esperado, uma vez que a quantidade de variáveis nos modelos agregados é menor. Deve-se observar que essa diminuição de tamanho do problema espera-se poder afetar favoravelmente os tempos de processamento dos modelos agregados.

Tabela 5.1 – Tamanho dos modelos do Conjunto 1

| | Desagregado | Método Ma1 e Ma2 |
|----------------------------|-------------|------------------|
| Tamanho | 150 x 5.547 | 150 x 650 |
| Linhas | 841.674 | 98.150 |
| Colunas | 836.250 | 97.650 |
| Valores diferentes de zero | 2.508.300 | 292.500 |

Analisando as localizações dos baricentros resultantes dos agrupamentos pelos Métodos Ma1 e Ma2, observa-se que muitos resultam os mesmos, como pode ser visto na Figura 5.3. Dos 650 agrupamentos obtidos pelo Método Ma2, 418 também resultaram do Método Ma1, o que significa que para alguns municípios, especialmente aqueles que possuem menor número de pontos de demanda e são mais afastados das grandes regiões metropolitanas, dois métodos diferentes podem resultar nos mesmos agrupamentos, especialmente no caso de demanda mais rarefeita (menos pontos) e mais dispersa; em outras palavras, o *K-Means* acaba determinando agrupamentos que coincidem com a divisão municipal. Os agrupamentos que

não são identificados como sendo os mesmos estão possivelmente associados a regiões com maior quantidade de pontos de demanda, devido ao emprego do algoritmo *K-Means* pelo Método Ma2 ao invés de respeitar os limites de município.

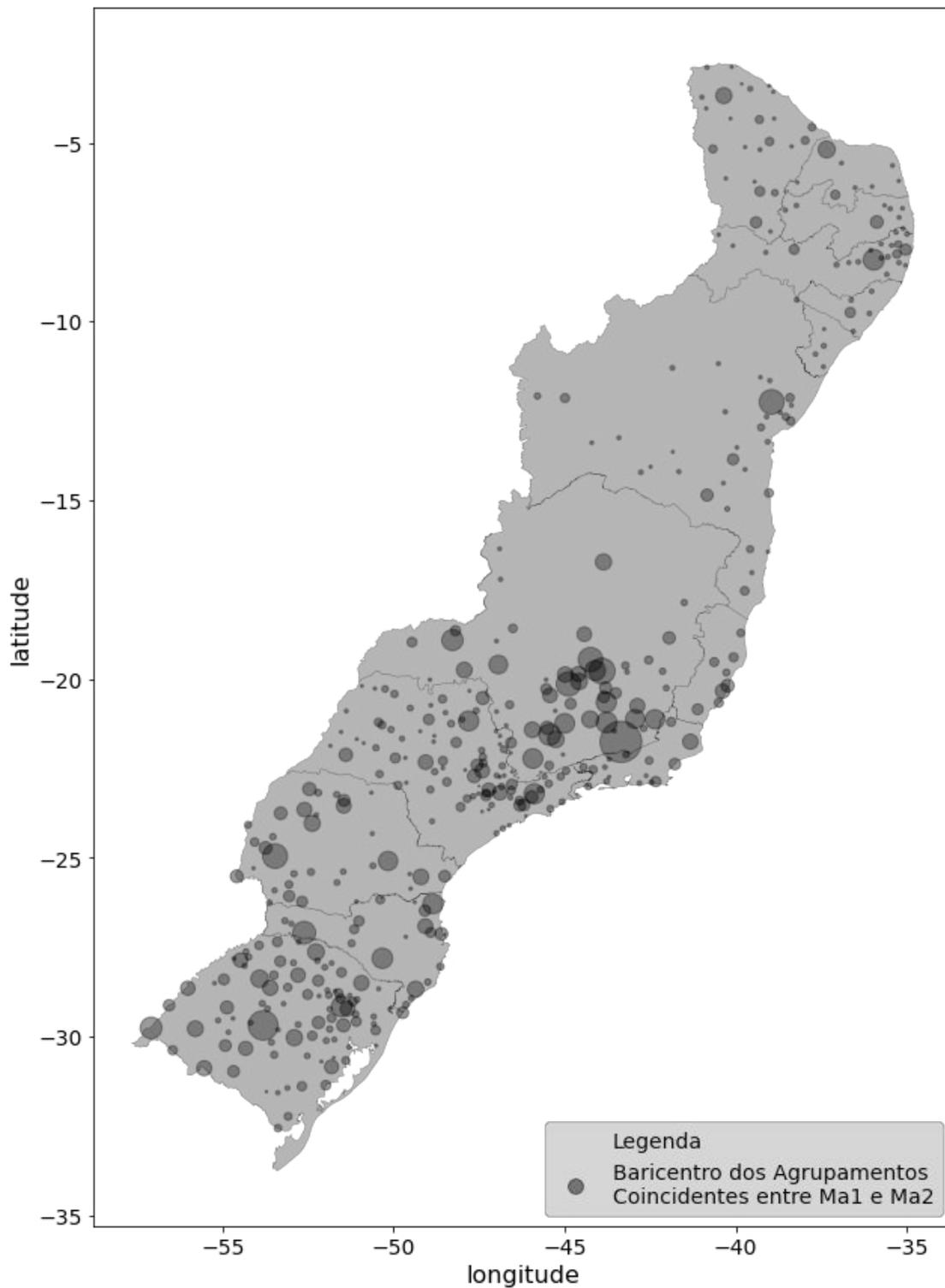


Figura 5.3 – Localização do baricentro dos agrupamentos coincidentes entre os métodos Ma1 e Ma2 para o Conjunto 1

5.1.2. Conjunto 2

No que diz respeito ao Conjunto 2, existem 585 agrupamentos de pontos de demanda nos Métodos Ma1 e Ma2. A Figura 5.4 mostra a localização dos baricentros dos agrupamentos dos pontos de demanda do Conjunto 2 pelo Método Ma1. A área de cada circunferência é proporcional a demanda de cada município. É possível ver que as circunferências referentes às capitais dos estados presentes no conjunto de dados são bem maiores que as demais, fato que está de acordo com a maneira na qual a população brasileira encontra-se distribuída.

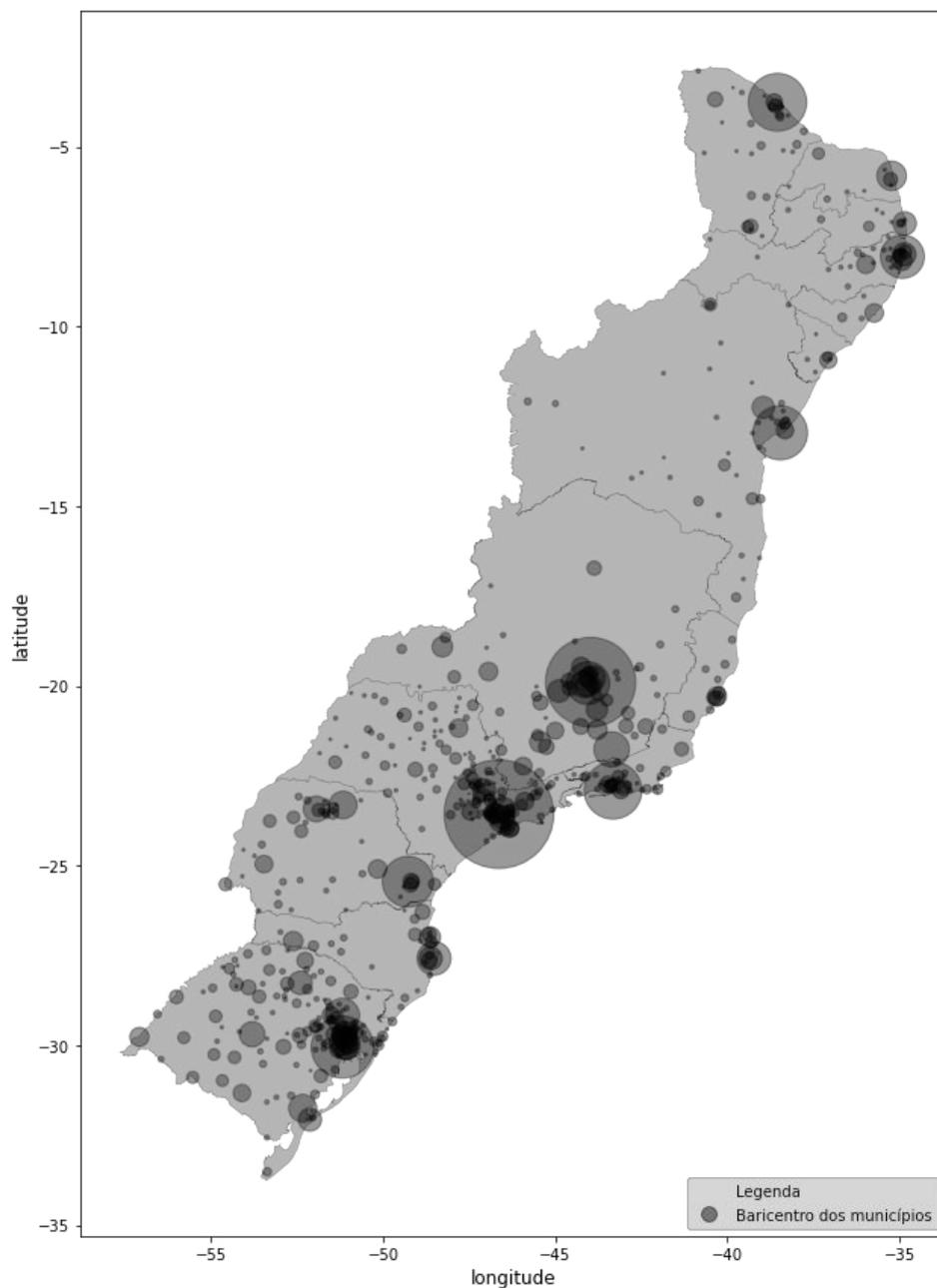


Figura 5.4 – Localização dos baricentros dos agrupamentos pelo método Ma1 e suas respectivas demandas para os pontos de demanda do Conjunto 2

Por sua vez, a localização dos baricentros dos agrupamentos dos pontos de demanda do Conjunto 2 aplicando o Método Ma2 são mostradas na Figura 5.5. Para o Conjunto 2, o valor de K considerado foi de 585, mesmo número de municípios. Observa-se que as áreas das circunferências são mais próximas umas às outras, o que pode significar que neste método, os agrupamentos possuem uma demanda mais próxima entre si que no Método Ma1. Um exemplo da diferença de tamanho das circunferências nas capitais ocorre em Belo Horizonte, indicada pela seta vermelha na figura. Se comparada à mesma região na Figura 5.4, pode se observar que os círculos aparentam mesmo tamanho, sem a sombra de uma circunferência muito maior que as demais, que representa a demanda da capital.

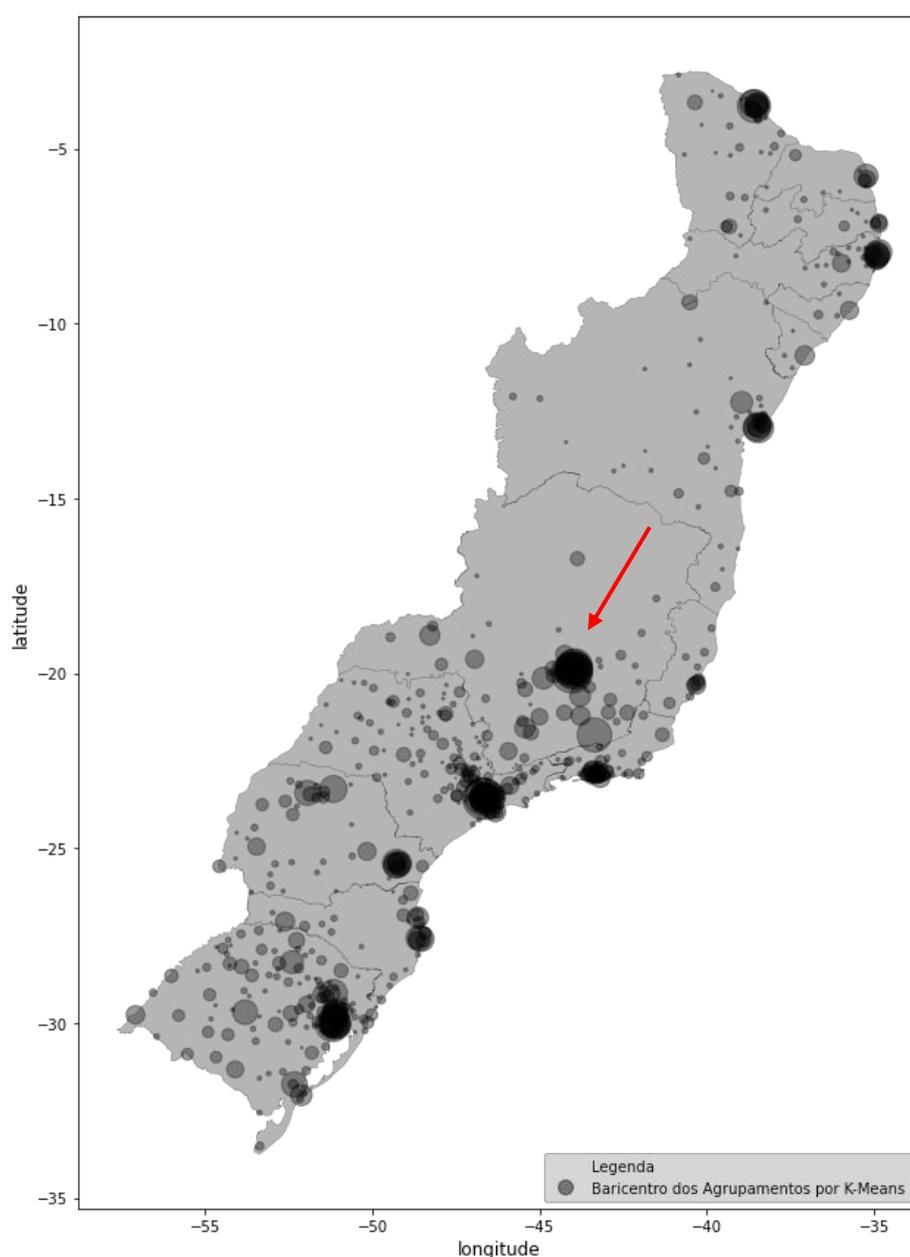


Figura 5.5 – Localização dos baricentros dos agrupamentos pelo método Ma2 e suas respectivas demandas para os pontos de demanda do Conjunto 2

A Tabela 5.2 mostra o tamanho dos modelos agregados e desagregado, bem como a quantidade de linhas, colunas e valores diferentes de zero em cada um. Para todos os modelos analisados, observa-se que o número de locais candidatos a instalação é 150. No modelo desagregado, há 4.460 pontos de demanda, enquanto nos modelos agregados, há 585 pontos de demanda, correspondente ao número de municípios do Conjunto 2. É possível ver que o modelo desagregado tem 7,62 vezes mais variáveis e restrições que os modelos agregados, o que pode, posteriormente, influenciar diretamente no tempo de processamento.

Tabela 5.2 – Tamanho dos modelos do Conjunto 2

| | Desagregado | Método Ma1 e Ma2 |
|----------------------------|-------------|------------------|
| Tamanho | 150 x 4.460 | 150 x 585 |
| Linhas | 673.460 | 88.335 |
| Colunas | 669.150 | 87.900 |
| Valores diferentes de zero | 2.007.000 | 263.250 |

Com relação aos agrupamentos, tanto pelo Método Ma1 (Figura 5.4), quanto pelo Método Ma2 (Figura 5.5), pode-se observar que muitos dos agrupamentos resultam idênticos em ambos os métodos. Dos 585 agrupamentos formados, 380 são comuns a ambos os métodos, o que sugere que o Método Ma2 somente forma agrupamentos diferentes do Método Ma1 em algumas regiões do espaço, associadas à grandes cidades e/ou regiões metropolitanas.

A Figura 5.6 mostra a localização dos baricentros dos agrupamentos que são comuns a ambos os modelos. Observa-se que muitos representam cidades com menor número de pontos de demanda, ou mais afastadas de outros centros urbanos, o que justifica a formação do mesmo agrupamento por dois métodos diferentes. Assim como na Figura 5.3, observa-se que as localidades correspondentes a megacidades e regiões metropolitanas como São Paulo, Rio de Janeiro e Belo Horizonte apresentam poucos, ou nenhum agrupamento correspondente, dado o maior número de pontos de demanda nessas localidades. Nestes casos, essa diferença é decorrente da aplicação do algoritmo *K-Means*, que busca minimizar as distâncias médias entre pontos, fazendo com que pontos mais próximos uns aos outros no mesmo município ou em vizinhos, como no caso de conurbações, se encontrem em um mesmo agrupamento.

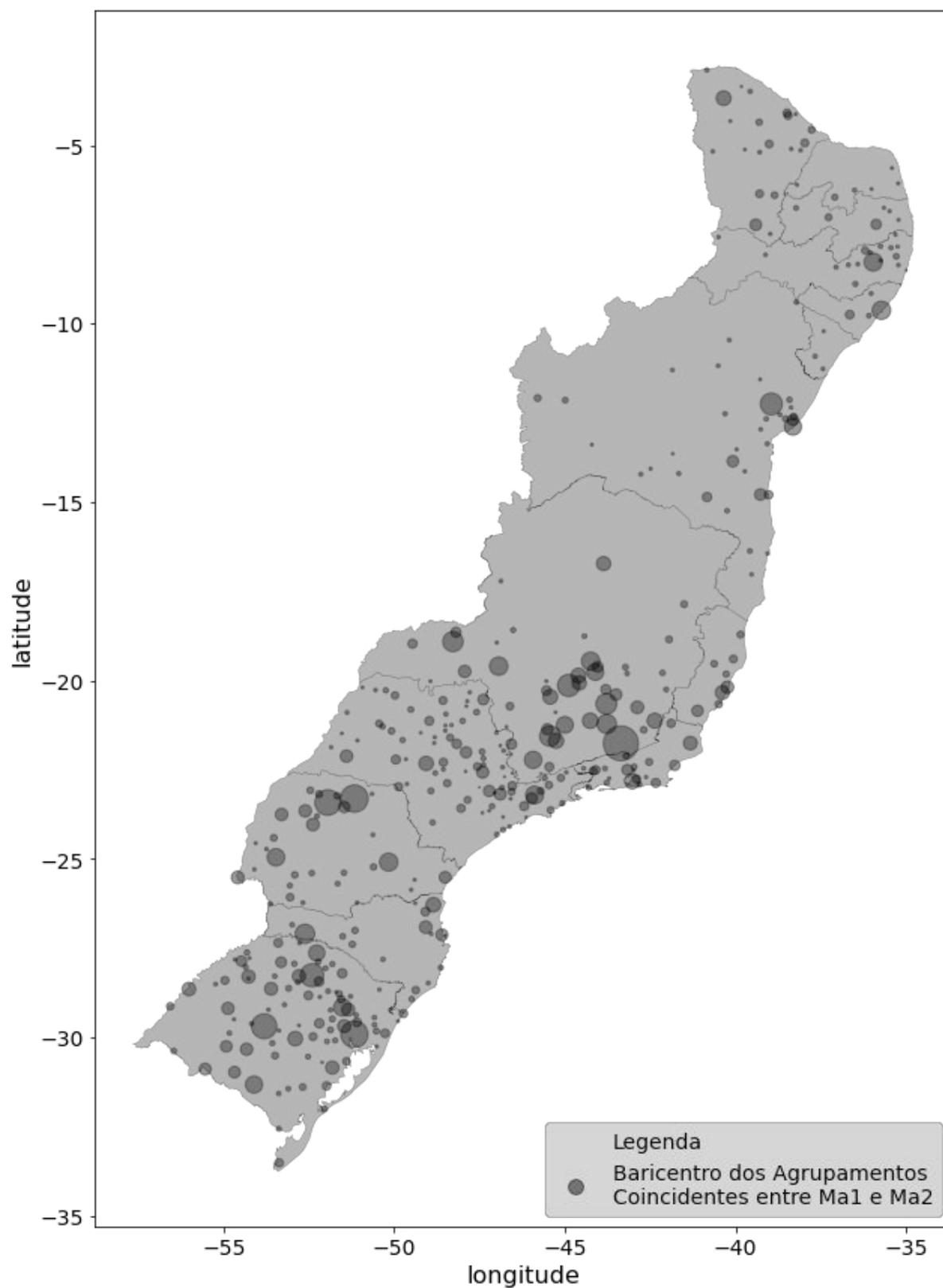


Figura 5.6 – Localização do baricentro dos agrupamentos coincidentes entre os métodos Ma1 e Ma2 para o Conjunto 2

5.2. Resultados dos modelos de otimização aplicados ao FCFL com e sem agregação da demanda

Nesta seção, são apresentadas as soluções obtidas para os modelos matemáticos de programação linear inteira que resolvem o FCFL com pontos de demanda agregados utilizando os dois métodos que estão sendo avaliados nesta pesquisa. De agora em diante, os modelos matemáticos cujos dados de demanda são agregados pelos Métodos Ma1 e Ma2 são chamados de Modelos Ma1 e Ma2, respectivamente. Também são apresentadas as soluções para os modelos que resolvam o FCFL com pontos de demanda desagregados (doravante modelo desagregado), o que permite a comparação dos resultados entre modelos com pontos de demanda agregados e desagregados. Foi utilizado o programa de otimização Gurobi (GUROBI, 2022) e linguagem de programação Python (versão 3.8) para a modelagem matemática de cada instância de cada conjunto de dados, como detalhado a seguir.

5.2.1. Conjunto 1

Na Tabela 5.3 são mostrados os valores resultantes da função objetivo, diferenças no valor da FO entre os modelos agregados e o desagregado e o número de instalações abertas na solução ótima para as instâncias do Conjunto 1. Como esperado, o número de instalações abertas nas soluções ótimas das instâncias diminuiu, uma vez que o custo fixo mensal aumenta. O Modelo Ma1 é aquele que apresenta as maiores diferenças com relação ao modelo desagregado, sendo essas diferenças 0,43% e 0,24%, nas instâncias 1-9 e 1-90, respectivamente. Nas instâncias 1-900 e 1-4500, é possível observar que o modelo Ma1 diferencia-se do modelo desagregado em 0,01%, enquanto nas instâncias 1-9000 a 1-90000, não existem diferenças no valor da FO entre modelo Ma1 e desagregado. Também na solução do Modelo Ma1 para essas instâncias, observa-se uma considerável diferença no número de instalações abertas na solução ótima do modelo, sendo 15 instalações abertas a menos na instância 1-9, que representa a maior diferença no número de instalações abertas entre um modelo agregado e o desagregado. A diferença no número de instalações abertas pode ser decorrente do método de agrupamento de pontos de demanda utilizado.

Tabela 5.3 – Informações sobre as instâncias do Conjunto 1

| Instância / Modelo | 1-9 | | | 1-90 | | | 1-900 | | |
|-----------------------------|-------------|------------|------------|-------------|------------|--------------|-------------|--------------|------------|
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 69.870,58 | 70.171,78 | 69.891,68 | 75.778,89 | 75.958,03 | 75.780,89 | 107.286,45 | 107.286,69 | 107.316,10 |
| Diferença FO % | - | 0,43% | 0,03% | - | 0,24% | 0,00% | - | 0,00% | 0,03% |
| # Instalações | 122 | 107 | 119 | 83 | 80 | 82 | 37 | 37 | 37 |
| Tempo de processamento | 26 | 1 | 1 | 189 | 1 | 1 | 425 | 1 | 1 |
| Nós explorados | 168 | 0 | 0 | 1294 | 0 | 0 | 384 | 0 | 0 |
| Iterações do <i>simplex</i> | 8.855 | 173 | 206 | 27.925 | 547 | 716 | 48.533 | 2.219 | 2.452 |
| Instância / Modelo | 1-1800 | | | 1-4500 | | | 1-9000 | | |
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 127.962,48 | 127.977,07 | 127.962,48 | 171.926,72 | 171.926,72 | 171.926,72 | 222.405,04 | 222.405,04 | 222.405,04 |
| Diferença FO % | - | 0,01% | 0,00% | - | 0,00% | 0,00% | - | 0,00% | 0,00% |
| # Instalações | 28 | 28 | 28 | 18 | 18 | 18 | 12 | 12 | 12 |
| Tempo de processamento | 162 | 1 | 1 | 181 | 1 | 1 | 378 | 1 | 1 |
| Nós explorados | 1.457 | 0 | 0 | 1.738 | 0 | 0 | 4.561 | 0 | 0 |
| Iterações do <i>simplex</i> | 55.717 | 3.208 | 3.413 | 81.833 | 4.983 | 5.293 | 283.492 | 6.806 | 6.851 |
| Instância / Modelo | 1-18000 | | | 1-45000 | | | 1-90000 | | |
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 299.313,99 | 299.313,99 | 299.313,99 | 456.279,80 | 456.279,80 | 456.279,80 | 580.209,97 | 580.209,97 | 580.209,97 |
| Diferença FO % | - | 0,00% | 0,00% | - | 0,00% | 0,00% | - | 0,00% | 0,00% |
| # Instalações | 9 | 9 | 9 | 5 | 5 | 5 | 4 | 4 | 4 |
| Tempo de processamento | 159 | 1 | 1 | 430 | 2 | 2 | 582 | 2 | 2 |
| Nós explorados | 1.604 | 0 | 0 | 2.998 | 0 | 0 | 3.910 | 0 | 0 |
| Iterações do <i>simplex</i> | 156.286 | 8.616 | 8.948 | 373.587 | 13.192 | 13.643 | 506.408 | 17.357 | 17.623 |

Analisando as soluções obtidas através do Modelo Ma2 (Tabela 5.3), é possível ver que a maior diferença no valor da FO com relação ao modelo desagregado é de 0,03% nas instâncias 1-9 e 1-900, enquanto nas demais instâncias, essa diferença é próxima (instância 1-90) ou igual a zero. Nas soluções ótimas do Modelo Ma2 para as instâncias 1-9 e 1-90, observa-se que o modelo agregado possui duas instalações abertas a menos que o modelo desagregado na instância 1-9, e uma a menos na instância 1-90. Essa diferença no número de instalações abertas reforça a tese de que o método utilizado para agrupar pontos de demanda influencia no número de instalações abertas, especialmente para as instâncias que apresentaram menor custo fixo mensal e maior número de instalações abertas.

Na Tabela 5.3 também são mostrados o tempo de processamento, o número de nós explorados e de iterações do *simplex* necessárias para obter as soluções ótimas das instâncias do Conjunto 1. O tempo de processamento do modelo desagregado é, em média, 281 segundos (aproximadamente 5 minutos), enquanto nos modelos agregados o tempo é, aproximadamente, 1 segundo para todas as instâncias, um indício de que reduzir o tamanho do problema pode diminuir consideravelmente o tempo de processamento, o que sugere não existir diferenças entre os modelos agregados com relação ao tempo de processamento. Apesar do tempo de processamento do problema desagregado ser pequeno, ainda assim pode-se perceber que o agrupamento dos pontos de demanda diminui o tempo de processamento, o que se pensarmos em problemas com um número ainda maior de variáveis, seja mais expressivo.

Ambas as informações sobre o número de nós explorados e de iterações do *simplex* são fornecidas pelo programa de otimização. O número de nós explorados está diretamente associado ao número de planos de *branch-and-cut* necessários para obter as soluções. Em média, para o modelo desagregado, foi necessário explorar 2.013 nós para obter a solução ótima, enquanto nos modelos agregados, nenhum plano de *branch-and-cut* foi explorado. Essa diferença é consistente com relação ao tamanho dos modelos, ou seja, um modelo maior, que possua maior número de variáveis e de restrições, precisa de um número maior de iterações para chegar a sua solução ótima que um modelo menor. O fato de os modelos agregados não calcularem planos do *branch-and-cut* e terem apenas o nó raiz explorado sugere que os Modelos Ma1 e Ma2 também podem obter soluções ótimas para os problemas, com menor tempo de processamento, especialmente para as instâncias de maior custo fixo, como 1-18000 a 1-90000.

O número de iterações do *simplex* pode ser diretamente associado à complexidade do problema. Em média, o modelo desagregado requer 171.404 iterações para obter sua solução

ótima, enquanto os Modelos Ma1 e Ma2, precisam em média de 6.345 e 6.572 iterações, respectivamente. Como esperado, o modelo desagregado necessita de um número muito maior de iterações para obter sua solução se comparado aos modelos agregados, possivelmente em decorrência de seu tamanho. Destaca-se o número de iterações do *simplex* para os Modelos Ma1 e Ma2. Seria esperado que por terem o mesmo tamanho, o número de iterações dos modelos agregados fosse o mesmo, contudo, o que se vê, é que o Modelo Ma1 necessita de 3,45% menos iterações, em média, que o Modelo Ma2 para obter soluções ótimas, o que sugere que, para instâncias de maior custo fixo e menor número de instalações abertas, o Método Ma1 é mais eficiente que o Método Ma2, especialmente nas instâncias de maior custo fixo, como 1-18000 a 1-90000, nas quais ambos os modelos agregados apresentam mesmo valor da FO que o modelo desagregado.

Para o conjunto 1, pode-se concluir que o Modelo Ma2 é mais adequado para agrupar os pontos de demanda nas instâncias de menor custo fixo mensal e maior número de instalações abertas, tais como 1-9 a 1-900, uma vez que o Modelo Ma2 apresenta diferenças menores no valor da FO com relação ao modelo desagregado. Nessas instâncias, nota-se que o método de agrupamento da demanda utilizado influencia na abertura de instalações mais próximas aos pontos de demanda, especialmente no Modelo Ma2, fazendo com que suas soluções sejam mais próximas ao modelo desagregado. À medida que observamos a diminuição no número de instalações abertas nas soluções ótimas das instâncias, percebemos que o agrupamento da demanda já não produz mais efeitos no valor da função objetivo, de tal forma que nas instâncias 1-18000 a 1-90000, os modelos agregados apresentam mesmo valor da FO que o modelo desagregado. Isso sugere que somente quando as instâncias possuem custo fixo muito pequeno ou que muitas instalações abertas são necessárias, o agrupamento da demanda afeta o valor da FO.

5.2.2. Conjunto 2

O Conjunto 2 é uma variante do Conjunto 1 na qual alguns pontos de demanda do Conjunto 1 foram sorteados, e tiveram suas demandas multiplicadas por um valor entre 0,60 e 1,40. No Conjunto 2, existem 4.460 pontos de demanda, localizados em 585 municípios. Para a solução do Conjunto 2, os locais candidatos são os mesmos utilizados na solução do Conjunto 1

Tabela 5.4 – Informações sobre as instâncias do Conjunto 2

| Instância / Modelo | 2-9 | | | 2-90 | | | 2-900 | | |
|-----------------------------|-------------|------------|--------------|-------------|--------------|------------|-------------|------------|------------|
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 49.615,16 | 49.819,17 | 49.616,10 | 55.118,82 | 55.191,36 | 55.118,82 | 82.236,44 | 82.236,44 | 82.236,44 |
| Diferença FO % | - | 0,41% | 0,00% | - | 0,00% | 0,00% | - | 0,00% | 0,00% |
| # Instalações | 117 | 103 | 117 | 76 | 76 | 76 | 33 | 33 | 33 |
| Tempo de processamento | 21 | 2 | 1 | 51 | 1 | 1 | 163 | 1 | 1 |
| Nós explorados | 209 | 0 | 0 | 1.284 | 0 | 0 | 1.655 | 0 | 0 |
| Iterações do <i>simplex</i> | 5.668 | 195 | 236 | 32.979 | 612 | 813 | 43.423 | 2.400 | 2.631 |
| Instância / Modelo | 2-1800 | | | 2-4500 | | | 2-9000 | | |
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 99.888,29 | 99.888,29 | 99.888,29 | 137.400,95 | 137.413,37 | 137.417,77 | 182.234,75 | 182.663,03 | 182.234,75 |
| Diferença FO % | - | 0,00% | 0,00% | - | 0,01% | 0,01% | - | 0,24% | 0,00% |
| # Instalações | 22 | 22 | 22 | 15 | 14 | 15 | 10 | 10 | 10 |
| Tempo de processamento | 156 | 1 | 1 | 127 | 1 | 1 | 135 | 1 | 1 |
| Nós explorados | 1.828 | 0 | 0 | 1.707 | 0 | 0 | 1.758 | 0 | 0 |
| Iterações do <i>simplex</i> | 59.445 | 3.395 | 3.671 | 95.226 | 5.197 | 5.520 | 120.268 | 7.124 | 7.291 |
| Instância / Modelo | 2-18000 | | | 2-45000 | | | 2-90000 | | |
| | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 | Desagregado | Modelo Ma1 | Modelo Ma2 |
| Valor da FO | 248.724,71 | 248.724,71 | 248.724,71 | 364.562,14 | 364.562,14 | 364.562,14 | 483.963,76 | 483.963,76 | 483.963,76 |
| Diferença FO % | - | 0,00% | 0,00% | - | 0,00% | 0,00% | - | 0,00% | 0,00% |
| # Instalações | 8 | 8 | 8 | 4 | 4 | 4 | 3 | 3 | 3 |
| Tempo de processamento | 419 | 1 | 2 | 261 | 2 | 2 | 150 | 2 | 2 |
| Nós explorados | 5.742 | 0 | 0 | 3.477 | 0 | 0 | 1.562 | 0 | 0 |
| Iterações do <i>simplex</i> | 221.939 | 9.068 | 9.335 | 370.581 | 13.283 | 13.482 | 521.360 | 20.495 | 20.976 |

Com relação às instâncias do Conjunto 2, os valores da FO, das diferenças nos valores da FO entre modelos agregados e desagregado e o número de instalações abertas são mostrados na Tabela 5.4. Como esperado, à medida que os custos fixos aumentam, o número de instalações abertas nas soluções ótimas diminui. Observa-se que nenhuma diferença entre modelos agregados e o desagregado excede 1%.

O Modelo Ma1 é aquele que apresenta a maior diferença com relação ao modelo desagregado: 0,41% na instância 2-9. Também na instância 2-9, o Modelo Ma1 apresenta a maior diferença no número de instalações abertas se comparado ao modelo desagregado, apresentando 14 instalações abertas a menos na solução ótima. Por sua vez, a solução do Modelo Ma2 na instância 2-9 mostra que a diferença no valor da FO com relação ao modelo desagregado é próxima a zero, e menor que a do Modelo Ma1. Observa-se também que são abertas 2 instalações a menos na solução ótima se comparado ao modelo desagregado. Essas diferenças entre os Modelos Ma1 e Ma2 indicam que os diferentes métodos de agrupamento utilizados podem influenciar no número de instalações abertas nas soluções ótimas dos modelos, em especial, naqueles que apresentam menor custo fixo e maior número de instalações abertas.

Na Tabela 5.4 também são mostrados os tempos de processamento, número de nós explorados e de iterações do *simplex* necessárias para obter a solução ótima para as instâncias do Conjunto 2. Observa-se que o tempo de processamento dos modelos agregados varia em torno de 1 segundo, enquanto no modelo desagregado, é necessário, em média, 165 segundos (aproximadamente 3 minutos) para obter a solução ótima. Essa diferença entre modelos agregados e desagregado indica novamente que uma diminuição no tamanho do modelo pode levar a uma diminuição considerável no tempo de processamento.

É possível ver também na Tabela 5.4, que nos modelos agregados são necessários, em média, 2.136 nós para obter a solução ótima, enquanto nos modelos desagregados, a solução é encontrada no nó raiz, o que significa que nenhum plano do *branch-and-cut* foi calculado. Essa diferença considerável no número de nós sugere que os modelos agregados sejam mais fáceis de resolver, além de demandarem menor tempo de processamento.

Por fim, com relação às iterações do *simplex*, é possível ver na Tabela 5.4 que enquanto o Modelo Ma1 necessita, em média, 6.683 iterações para obter sua solução ótima, o Modelo Ma2 necessita de 7.106 iterações. Diferentemente do modelo desagregado, que necessita 163.432 iterações, um número 23 vezes maior de iterações. Essa diferença no número de

iterações é esperada, uma vez que o modelo desagregado apresenta maior número de variáveis e restrições. Destaca-se que, apesar dos modelos agregados apresentarem o mesmo tamanho, o Modelo Ma1 requer, em média, 3,5% menos iterações para ser solucionado que o Modelo Ma2, o que sugere que o Modelo Ma1 seja mais eficiente, especialmente para instâncias de maior custo fixo, como 2-18000 a 2-90000, nas quais ambos os modelos agregados não apresentam diferenças com relação ao modelo desagregado.

Com relação ao Conjunto 2, também é possível notar que o Modelo Ma2 é mais adequado para agrupar os pontos de demanda, especialmente nas instâncias de menor custo fixo mensal e maior número de instalações abertas, como 2-9 a 2-9000, uma vez que apresenta diferenças menores no valor da FO com relação ao modelo desagregado. Nessas instâncias, percebemos novamente que o agrupamento da demanda modifica o conjunto de instalações abertas nas soluções de cada instância, fazendo com que, no Modelo Ma2 existam mais instalações abertas que no Modelo Ma1, que se encontram mais próximas dos pontos de demanda. Observa-se também que à medida que o custo fixo de operação aumenta, diminuindo o número de candidatos abertos nas soluções ótimas, os efeitos do agrupamento da demanda não se mostram relevantes, uma vez que o valor da FO dos modelos agregados é o mesmo do modelo desagregado para as instâncias 2-18000 a 2-90000, reforçando que os efeitos do agrupamento da demanda são perceptíveis apenas em instâncias de menor custo fixo mensal e maior número de instalações abertas.

5.3. Análise comparativa dos métodos de agregação propostos

Nesta seção, são analisadas duas instâncias selecionadas de cada conjunto, nas quais ambos os modelos agregados apresentam diferença no valor da FO com relação ao modelo desagregado. Para o Conjunto 1, foram escolhidas as instâncias 1-9 e 1-900, por serem as instâncias nas quais o valor da FO do Modelo Ma1 (agregação por município) é maior que o do Modelo Ma2 (0,43% e 0,03%, respectivamente) e cujo valor da FO do Modelo Ma2 (*K-Means*) é maior que o do Modelo Ma1 (0,03% e 0,01%, respectivamente) e os dois modelos agregados são diferentes do modelo desagregado. Já para o Conjunto 2, as duas instâncias analisadas, 2-9 e 2-4500, são as únicas nas quais o valor da FO do Modelo Ma2 é diferente do modelo desagregado, sendo que a instância 2-9 também é aquela na qual o valor da FO do Modelo Ma1 apresenta a maior diferença (0,41%) com relação ao modelo desagregado.

Nas análises dos dois conjuntos, os candidatos abertos nas soluções ótimas são representados por triângulos, para os quais azul indica que é um candidato aberto apenas no

modelo desagregado, laranja indica um candidato aberto apenas no Modelo Ma1, e magenta, um candidato aberto apenas no Modelo Ma2. Para representar as diferenças percentuais no custo de transporte em decorrência da alocação da demanda a outras instalações, utiliza-se uma escala de cores vermelha e verde, sendo que vermelho indica que os custos de transporte do modelo com pontos de demanda agregados são menores que aqueles do modelo desagregado, enquanto a cor verde indica o contrário. As alocações da demanda que indicam a partir de qual candidato aberto cada ponto de demanda é atendido são detalhadas no APÊNDICE B.

5.3.1. Conjunto 1

Inicialmente, buscou-se detectar regiões no espaço onde pode existir diferenças de alocação entre modelos e para tal, utiliza-se o algoritmo DBSCAN (ESTER *et al.*, 1996), como detalhado anteriormente na seção 3.4. Para executar o algoritmo, é necessário definir o parâmetro *Eps*, que representa a maior distância entre pontos dentro de um mesmo agrupamento, e *MinPts*, a menor quantidade de pontos necessária para que um agrupamento seja formado.

Com relação ao parâmetro *MinPts*, que indica o menor número de pontos que formam um agrupamento, foi considerado o valor 3, também utilizado em Rahmah e Sitanggang (2016), e que possibilita a identificação de agrupamentos dentro de um mesmo município que estejam mais afastados uns dos outros.

Para o cálculo do parâmetro *Eps*, foram utilizados todos os pontos de demanda que estejam em municípios que possuem um número maior ou igual ao percentil 75 do número de pontos de demanda, ou seja, para o Conjunto 1, todos os pontos de demanda em municípios com 6 ou mais pontos foram utilizados no cálculo do parâmetro de distância. Optou-se por remover parte dos municípios do cálculo do parâmetro *Eps* dado que esses pontos possivelmente encontram-se em agrupamentos em comum entre os Modelos Ma1 e Ma2, e não nas regiões onde a alocação de demanda entre modelos é diferente.

O parâmetro *Eps* foi determinado através do ponto de inflexão da curva que associa a distância com relação aos vizinhos mais próximos. A Figura 5.7 mostra a curva (em azul) que associa a distância com relação aos vizinhos mais próximos e o número de pontos de demanda. É possível ver que o ponto de inflexão é próximo a 2 km, marcado pela linha tracejada laranja. Esse valor indica a maioria dos vizinhos mais próximos se encontram a uma distância aproximada de 2 km, e é adequado para o Conjunto 1, sabendo-se que a distância média entre pontos desse conjunto é de 1,41 km.

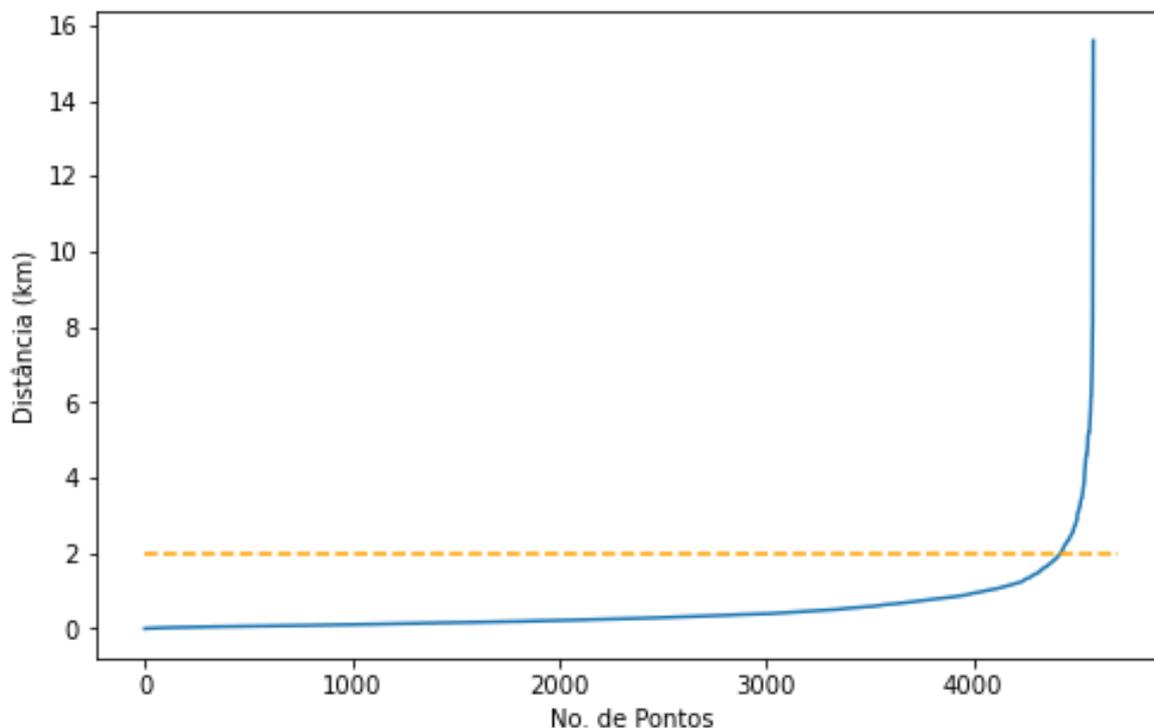


Figura 5.7 – Cálculo do parâmetro *Eps* para o Conjunto 1

Dos 161 municípios analisados pelo DBSCAN, foram observados 13 municípios nos quais há mais de um agrupamento, 17 municípios nos quais os agrupamentos envolvem pontos de demanda em mais de uma localidade, e 13 municípios que possuem um ou mais agrupamentos em seu território ou em que os pontos de demanda foram agrupados com outros de outros municípios. Para os demais municípios, todos os pontos de um mesmo município correspondem ao mesmo agrupamento obtido através do DBSCAN. Nota-se que o número de municípios analisados pelo DBSCAN é diferente daquele que representa o número de agrupamentos diferentes aos Modelos Ma1 e Ma2, e isso é consequência da utilização municípios somente com 6 pontos de demanda ou mais. Apesar de não usar as informações dos pontos de demanda em agrupamentos que não se repetem nos Modelos Ma1 e Ma2, o DBSCAN acaba por identificar os mesmos agrupamentos, o que reforça o indício de que são nas regiões do espaço com maior número de pontos de demanda que podem acontecer diferenças no conjunto de instalações abertas e logo a partir de qual delas a demanda de cada ponto é atendida. Ou seja, as regiões do espaço nas quais pode acontecer diferença na alocação da demanda são as mesmas que possuem muitos pontos de demanda, o que sugere que em conurbações, com regiões metropolitanas, o agrupamento de pontos pelo Modelo Ma1 pode não ser o mais adequado.

A Figura 5.8 mostra a localização dos agrupamentos do DBSCAN considerando os municípios que possuem ou cujos pontos de demanda se encontram mais de um agrupamento. As cores das localidades representam os possíveis casos de agrupamentos formados pelo DBSCAN, sendo o Caso 2 (cor laranja) aquele no qual existe mais de um agrupamento em um mesmo município, o Caso 3 (cor azul) aquele no qual existem agrupamentos com pontos de demanda localizados em mais de um município, enquanto o Caso 4 (cor rosa), é aquele no qual existem tanto agrupamentos de pontos de demanda pertencentes somente ao município, quanto agrupamentos com pontos de demanda em mais de um município.

Na Figura 5.8, observa-se que os agrupamentos se encontram principalmente nas capitais dos respectivos estados, e em algumas cidades maiores no interior de São Paulo, o que é esperado, uma vez que cidades maiores têm mais pontos de demanda. Também se percebe que capitais como Rio de Janeiro, Vitória e Porto Alegre possuem mais de um agrupamento em seu território, o que sugere que existem agrupamentos menores dentro do próprio município e que podem não ser levados em consideração no Modelo Ma1 ao agrupar os pontos de demanda apenas por município. Em outras capitais, como Belo Horizonte e São Paulo, é possível ver que os agrupamentos são formados tanto por pontos de demanda em municípios vizinhos, quanto no próprio município, dado o elevado número de pontos de demanda, e as localizações relativas dos municípios. Ou seja, as regiões identificadas pelo DBSCAN nas quais pode acontecer alocação incorreta da demanda, correspondem as regiões nas quais os agrupamentos dos Modelos Ma1 e Ma2 não são os mesmos.

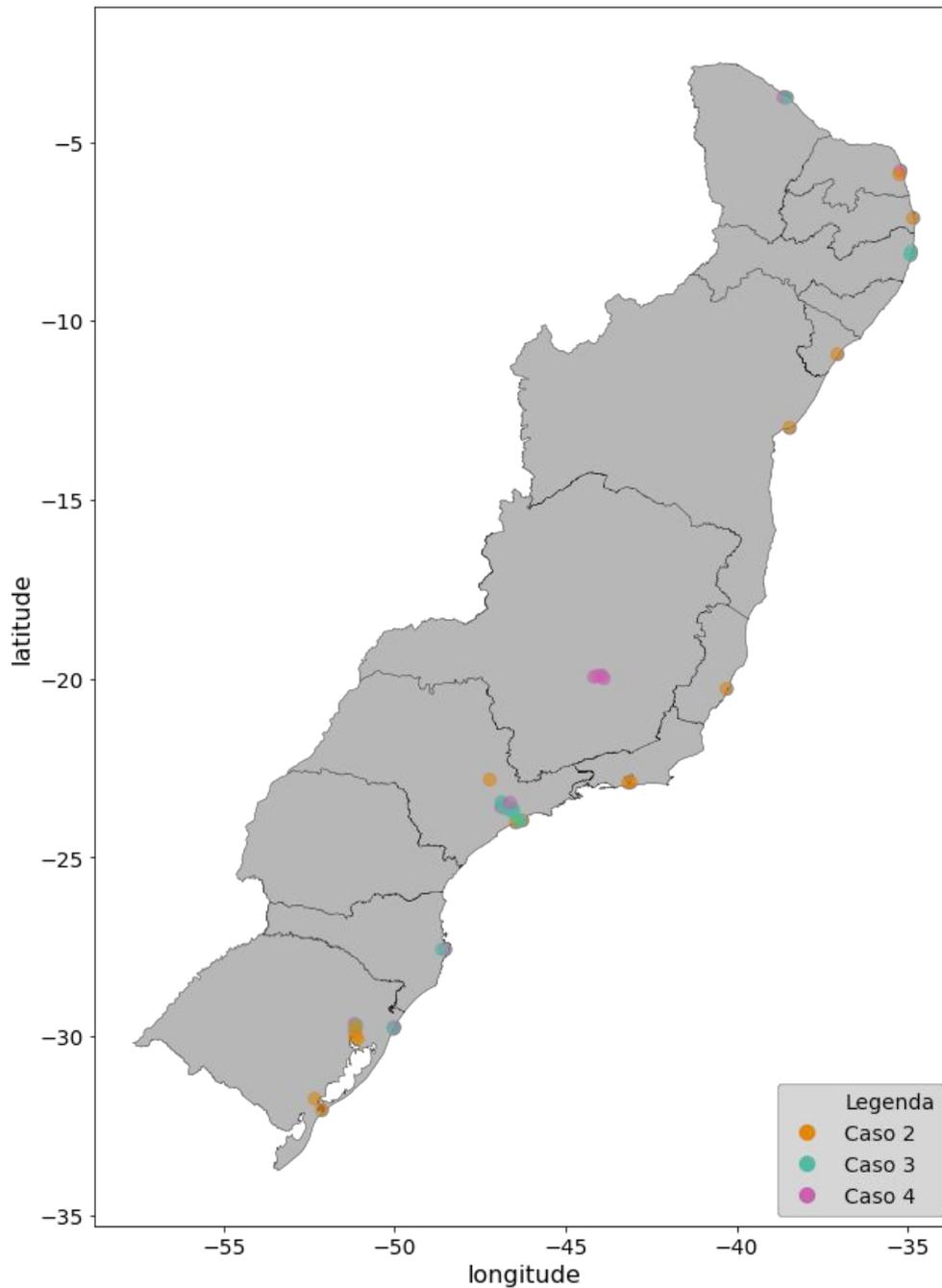


Figura 5.8 – Agrupamentos do DBSCAN para o Conjunto 1

O DBSCAN permite que os pontos de demanda de um mesmo município se encontrem em agrupamentos diferentes. A Figura 5.9 mostra alguns dos agrupamentos do DBSCAN para a Região Metropolitana de São Paulo. Observa-se que dentro de um mesmo município, como São Paulo, é possível existir mais de um agrupamento, e que dada a proximidade entre os pontos de demanda de São Paulo com outros municípios, tais como Osasco, Santo André e São Bernardo do Campo, o agrupamento que os representa é o mesmo. Diferentemente do que acontece com os pontos de demanda em Mogi das Cruzes, cidade na qual todos os pontos de demanda do município encontram-se no mesmo agrupamento.

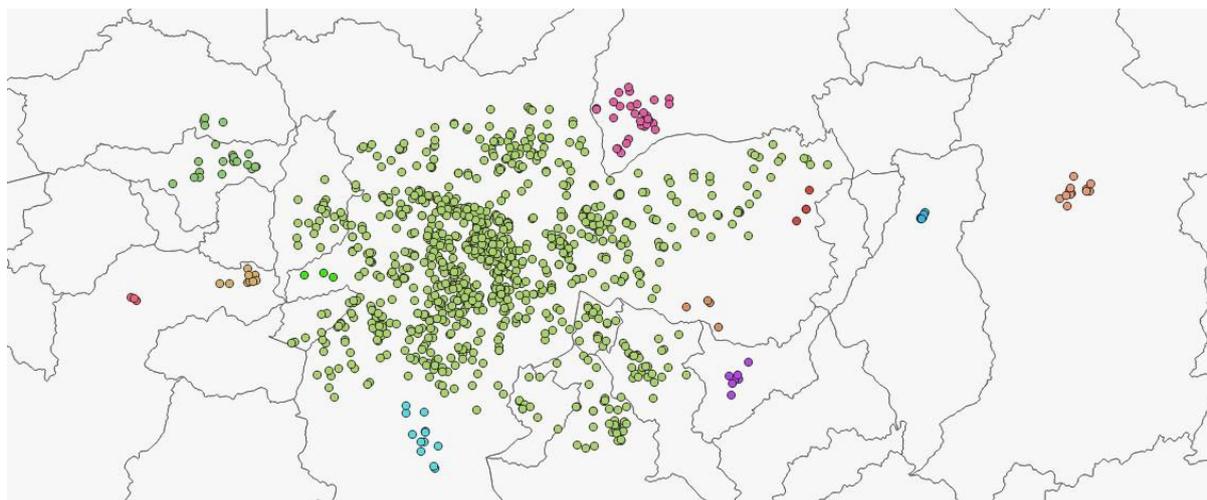


Figura 5.9 – Exemplo de agrupamentos do DBSCAN

Instância 1-9

Na instância 1-9, Modelo Ma1 apresenta 0,43% de diferença no valor da FO com relação ao modelo desagregado. Na solução de Ma1, foram identificados 107 candidatos abertos, 15 a menos que no modelo agregado. Dentre os 107 candidatos abertos em Ma1, 104 também são abertos no modelo desagregado, o que significa que a diferença no valor da FO é decorrência do número e da localização das instalações abertas no Modelo Ma1.

A Figura 5.10 mostra os locais nos quais os candidatos abertos nos modelos desagregado e Ma1 são diferentes. É possível ver que o Modelo Ma1 possui instalações abertas somente em Minas Gerais e no Rio de Janeiro, fazendo com que os pontos de demanda nessas localidades sejam atendidos a partir de candidatos diferentes daqueles determinados a partir do modelo desagregado. Nos outros estados, não existem candidatos abertos apenas no modelo agregado, o que significa que a demanda dos pontos nessas regiões (São Paulo, Porto Alegre e Salvador) é atendida por candidatos abertos que são comuns aos dois modelos, o que leva ao aumento no valor da FO.

As diferenças nos custos de transporte por município para o Modelo Ma1 são mostradas na Figura 5.11. São observadas diferenças entre -2,46% e 18,46%, o que significa que existem locais nos quais os custos de transporte no Modelo Ma1 são menores e maiores, respectivamente, que os do modelo desagregado. Em Minas Gerais, é possível ver que um dos municípios é colorido em vermelho, o que é esperado uma vez que na solução do Modelo Ma1 existe um candidato aberto mais próximo aos pontos de demanda desse município que o aberto no modelo desagregado, o que pode ser confirmado através da Figura 5.10.

Por sua vez, nas localidades dos estados da Bahia, Rio Grande do Sul e São Paulo, onde há apenas instalações do modelo desagregado, observa-se que os custos de transporte são todos positivos. Como esperado, esses pontos são atendidos a partir de candidatos abertos mais distantes no Modelo Ma1 que no modelo desagregado, o que justifica o aumento nos custos de transporte.

Por último, no estado do Rio de Janeiro existe uma única instalação aberta no Modelo Ma1 mostrada na Figura 5.11, todavia, o que se observa, tanto na cidade do Rio de Janeiro, quanto na vizinha Itaguaí, os custos de transporte são maiores no Modelo Ma1 quando comparado ao modelo desagregado uma vez que o número de instalações abertas é menor. Destaca-se a cidade de Seropédica, colorida em verde mais claro, que deveria apresentar custo de transporte maior que as outras por ser mais distante. Neste caso, a demanda do município é atendida a partir de uma instalação aberta que é comum às soluções dos modelos desagregado e Ma1, a qual é mais próxima ao local, o que diminui, portanto, as diferenças nos custos de transporte para esse município.

O Modelo Ma2 apresenta diferença de 0,03% no valor da FO com relação ao modelo desagregado. Também foi mostrado na Tabela 5.4 que existem 3 candidatos abertos a menos em sua solução ótima quando comparado ao modelo desagregado. Dentre os 119 candidatos abertos na solução do Modelo Ma2, 117 também estão presentes na solução do modelo desagregado. A presença de maior número de instalações abertas e comuns aos modelos Ma2 e desagregado podem justificar a menor diferença no valor da FO entre esses modelos quando comparados ao Modelo Ma1.

A Figura 5.12 mostra a localização dos candidatos que abrem somente nos modelos desagregado e Ma2. É possível ver que as regiões destacadas nas quais esses candidatos se encontram correspondem às capitais dos estados, anteriormente identificadas pelo DBSCAN como regiões onde poderia acontecer diferença na alocação da demanda. Observa-se que o Modelo Ma2 possui instalações únicas somente no Rio Grande do Sul, o que pode justificar as diferenças nos custos de transporte.

A Figura 5.13 mostra as diferenças no custo de transporte por município entre os modelos desagregado e Ma2. As diferenças se encontram entre -1,37% e 1,36%, o que indica que existem localidades nas quais o Modelo Ma2 apresenta valores menores ou maiores de custo de transporte que o modelo desagregado. Nas localidades mostradas de Minas Gerais e São Paulo, as diferenças nos custos de transporte são todas positivas, o que significa que o

Modelo Ma2 não abriu instalações diferentes daquelas que também foram abertas no modelo desagregado, fazendo com que essas instalações atendam a um conjunto diferente de pontos de demanda e aumentando, portanto, o custo de transporte.

Já no Rio Grande do Sul, nota-se que o município de Viamão apresenta cor vermelha, decorrente do menor custo de transporte do Modelo Ma2, o que já é esperado, uma vez que uma das instalações abertas no modelo agregado é mais próxima do município que as demais. Todavia, percebe-se também que os demais municípios apresentam diferenças positivas, como consequência da abertura de uma (não duas) instalação mais distante que aquelas do modelo desagregado para atender a demanda.

Para a instância 1-9, pode-se observar que todas as regiões nas quais existem pontos de demanda alocados a instalações diferentes nos modelos agregados foram identificadas anteriormente pelo DBACAN. Observa-se também que o Modelo Ma2 além de apresentar uma diferença no valor da FO de 0,03% se comparada ao Modelo Ma1 no qual a diferença de 0,43%, também apresenta menores diferenças nos custos de transporte (varia entre -1,37% e 1,36% no Modelo Ma2, e entre -2,46% e 18,46% no Modelo Ma1), em decorrência do maior número de instalações abertas em sua solução ótima. Essas evidências sugerem que o Modelo Ma2 mesmo não apresentando mesmo valor de FO que o modelo desagregado é mais adequado para agrupar os pontos de demanda da instância 1-9, ao se tratar de uma instância com maior número de candidatos abertos e menor custo fixo.

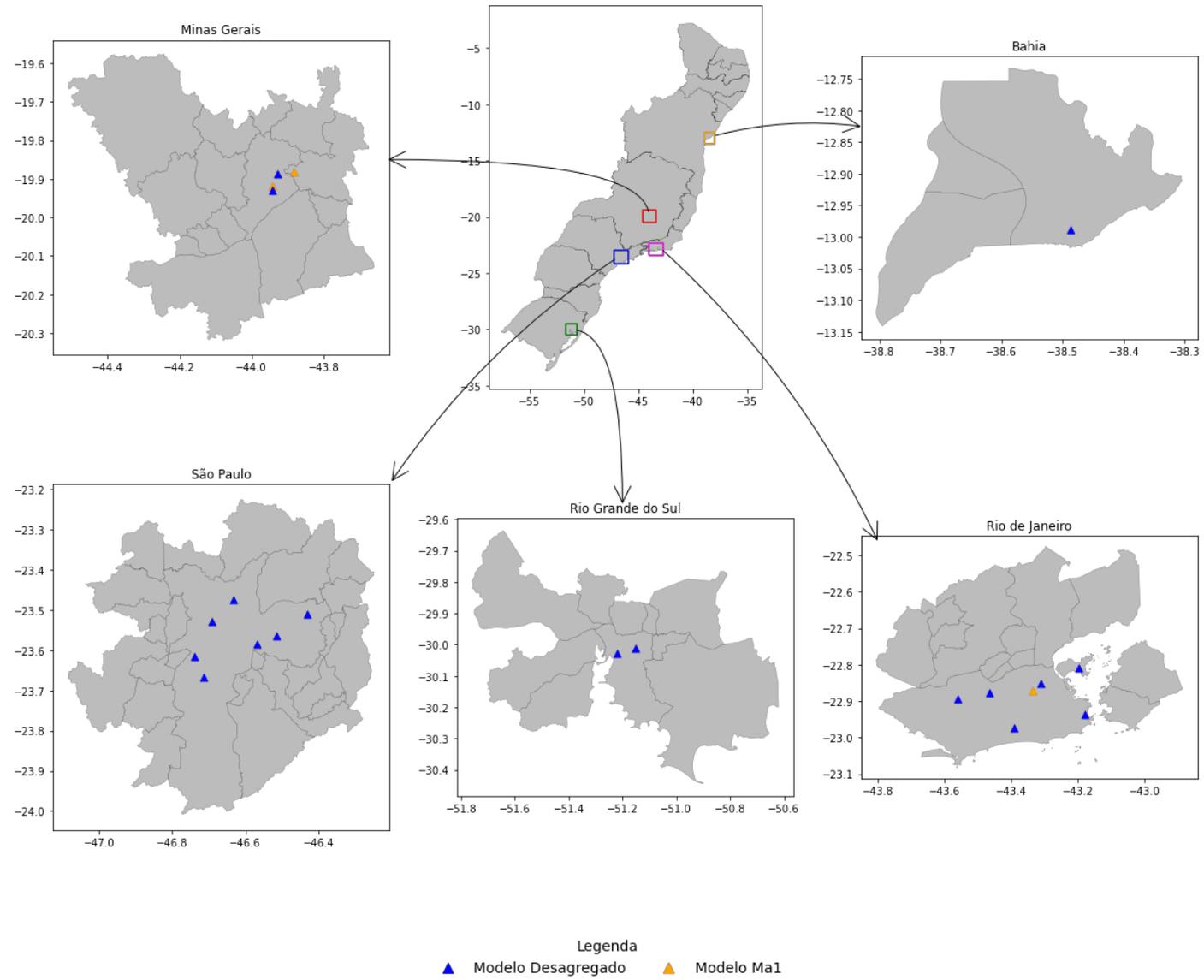


Figura 5.10 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 1-9

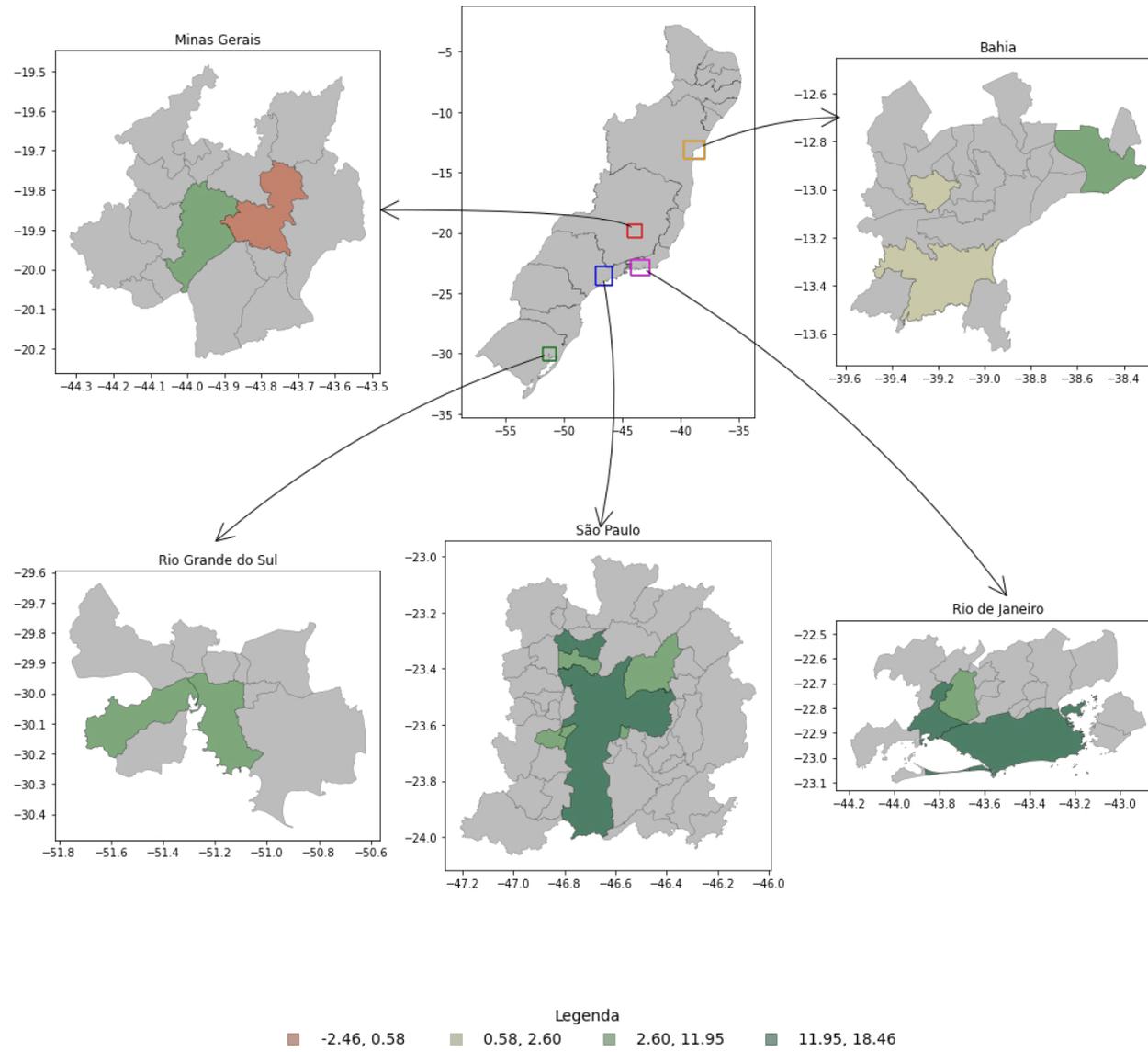


Figura 5.11 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 1-9

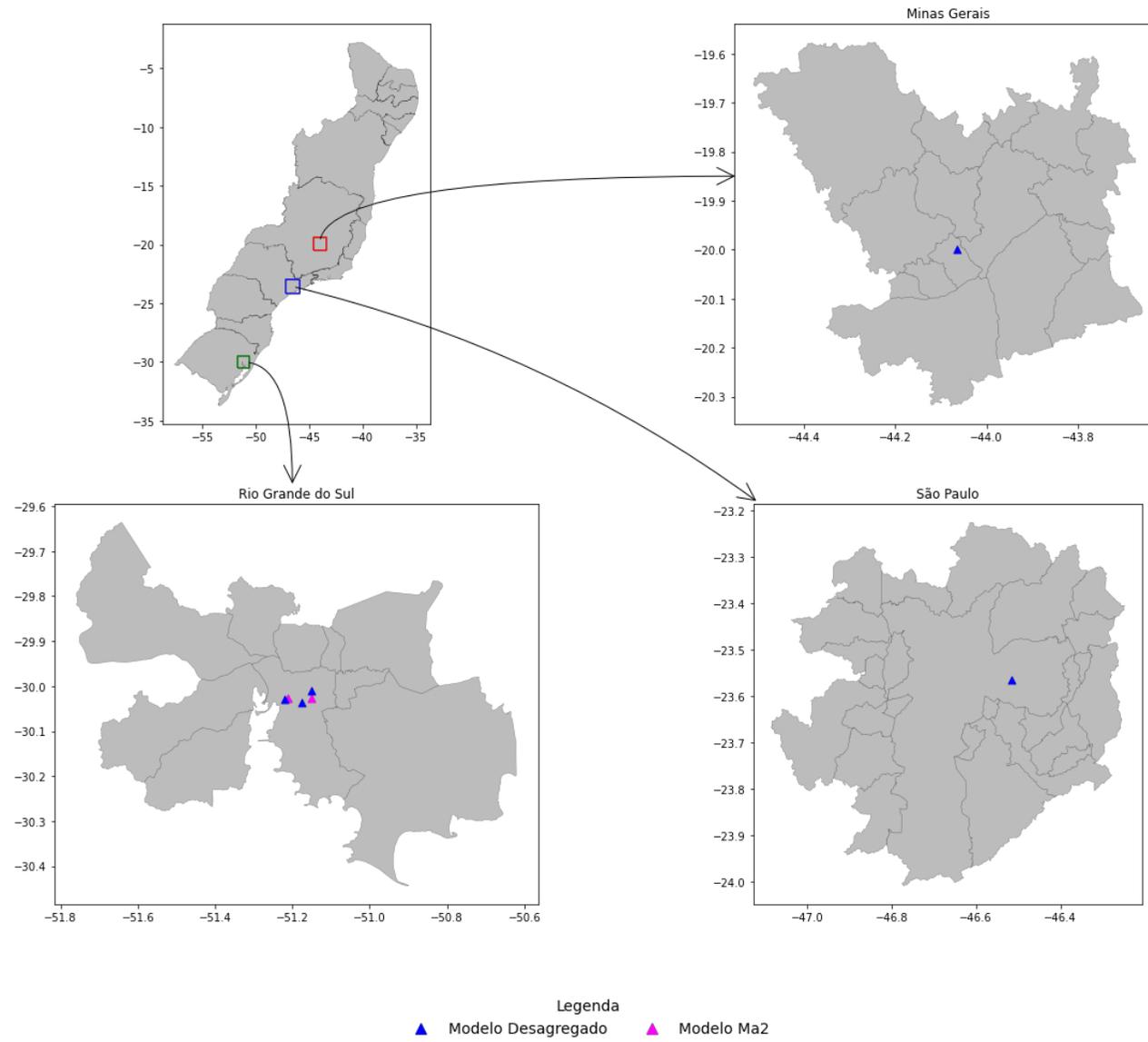


Figura 5.12 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 1-9

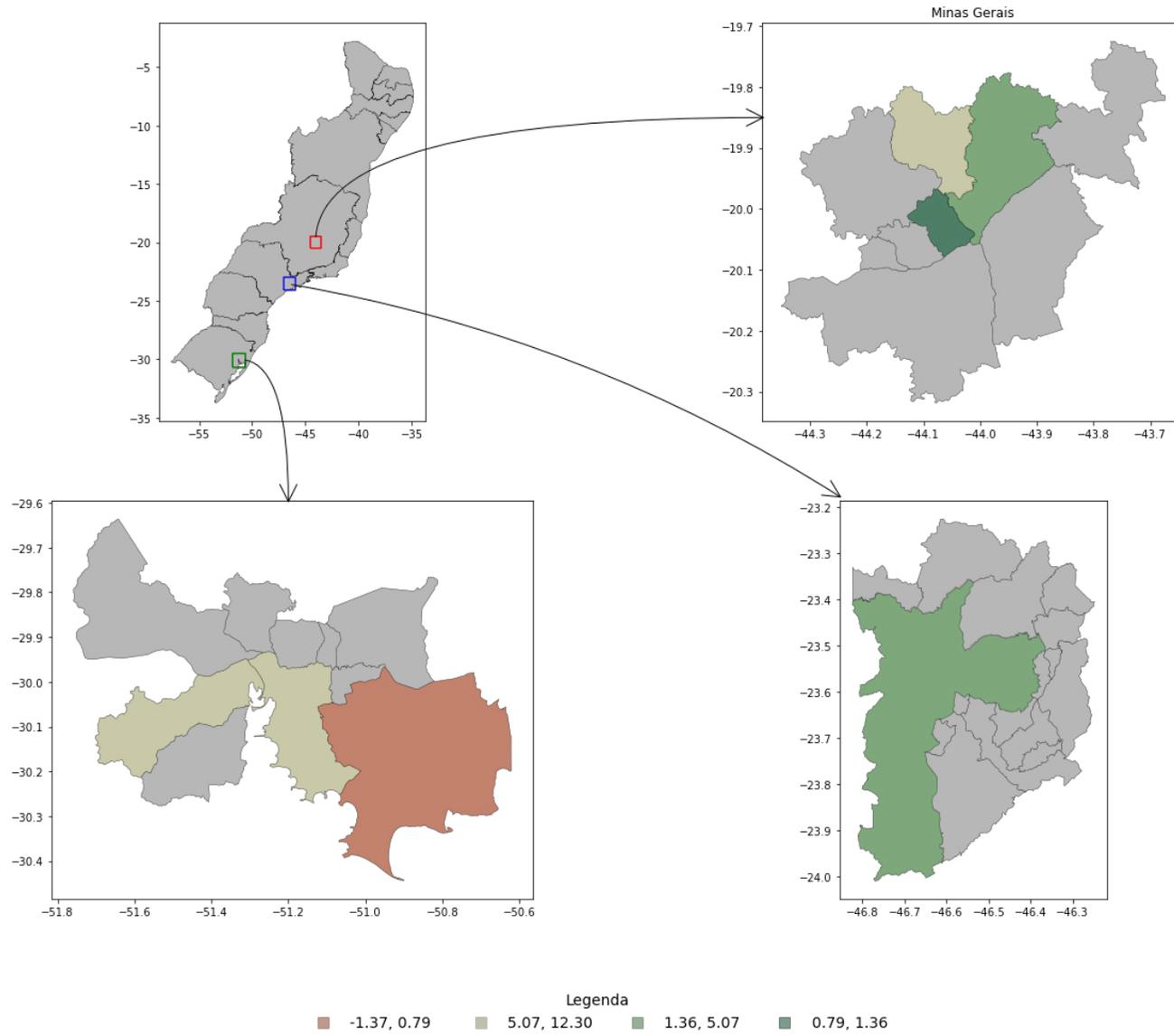


Figura 5.13 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 1-9

Instância 1-900

Na solução da instância 1-900 pelo Método Ma1, foi observada uma diferença no valor da FO próxima a zero (equivalente a R\$ 0,24). Não há diferença no número de instalações abertas nos modelos Ma1 e desagregado, porém existe 1 instalação aberta no Modelo Ma1 que é diferente daquelas abertas no modelo desagregado, ou seja, ocorre uma troca no conjunto de candidatos abertos. A Figura 5.14 mostra a localização das instalações diferentes nos modelos desagregado e Ma1. Ambas se encontram no município de São Paulo, porém em decorrência do método de agrupamento empregado, a instalação aberta no Modelo Ma1 encontra-se mais à nordeste que aquela aberta no modelo desagregado, o que ocasionou a diferença de alocação de demanda.

As diferenças nos custos de transporte por município entre os modelos desagregado e Ma1 são mostradas na Figura 5.15. Observa-se que as diferenças variam entre -3,02% e 3,39%, o que significa que, ao abrir um outro candidato para atender a demanda, o Modelo Ma1 fez com que alguns pontos de demanda fossem atendidos por uma instalação mais próxima a eles, enquanto outros foram atendidos por uma instalação mais distante, e por isso existem tais diferenças de custo. Como o candidato aberto no Modelo Ma1 encontra-se à nordeste do modelo desagregado, é possível ver que os municípios mais a norte e a leste apresentam cores vermelhas, enquanto os municípios ao sul e a oeste apresentam cores verdes.

Já no Modelo Ma2, observa-se uma diferença de 0,03% com relação ao valor da FO do modelo desagregado, ainda que o número de candidatos abertos na solução dos dois modelos seja o mesmo. Isso ocorre porque existem dois candidatos abertos que distinguem as duas soluções. A Figura 5.16 mostra a localização dos candidatos que são abertos somente nos modelos desagregado e Ma2. Apenas no Rio de Janeiro e em São Paulo existem diferenças nas instalações abertas.

Na Figura 5.17 são mostradas as diferenças nos custos de transporte entre os municípios nos modelos desagregado e Ma2. As diferenças nos custos de transporte variam entre -3,02% e 63,77%, em decorrência da localização dos candidatos abertos. Com relação ao estado de São Paulo, observa-se que o conjunto de municípios presentes é o mesmo mostrado anteriormente na Figura 5.15.

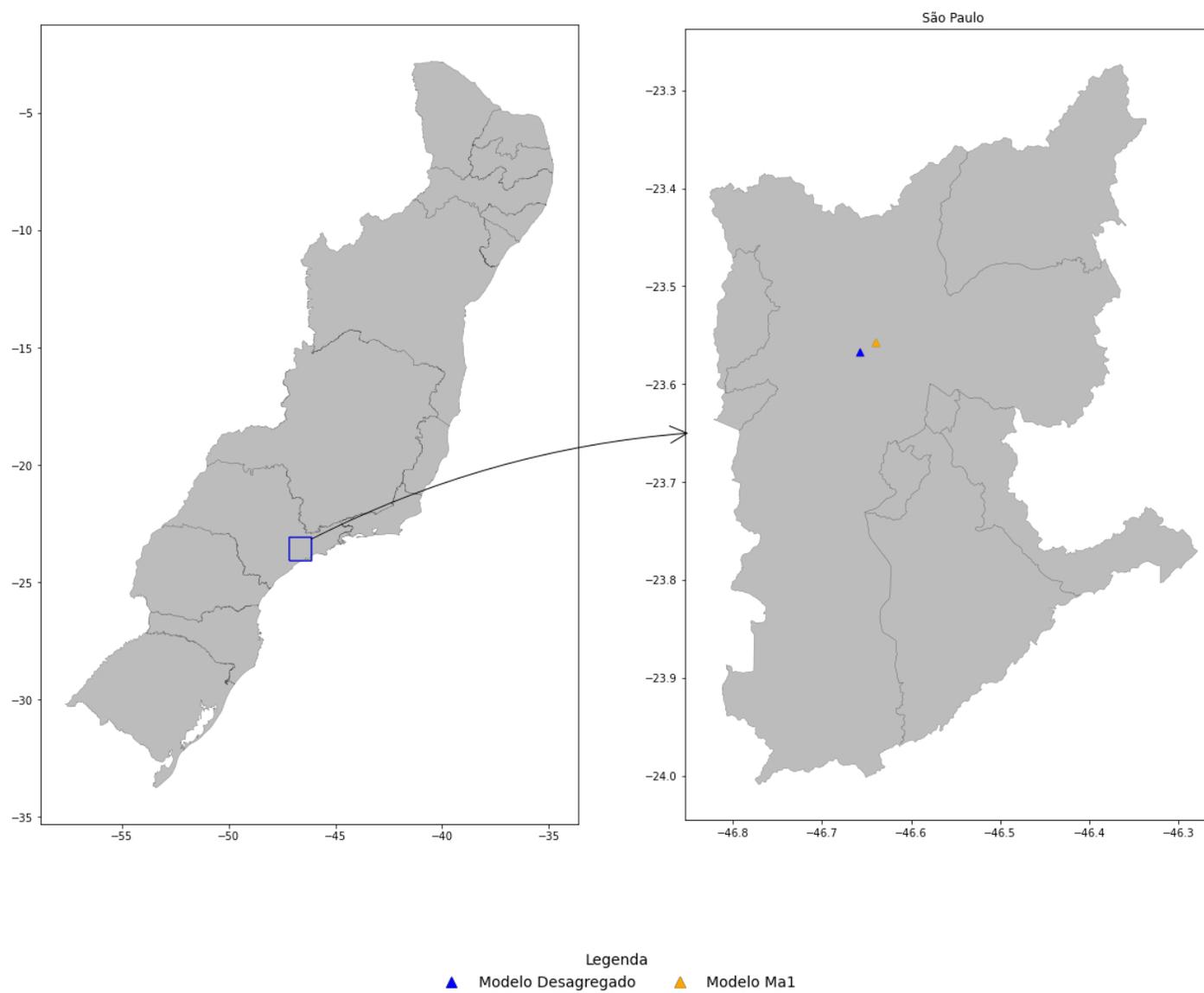


Figura 5.14 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 1-900

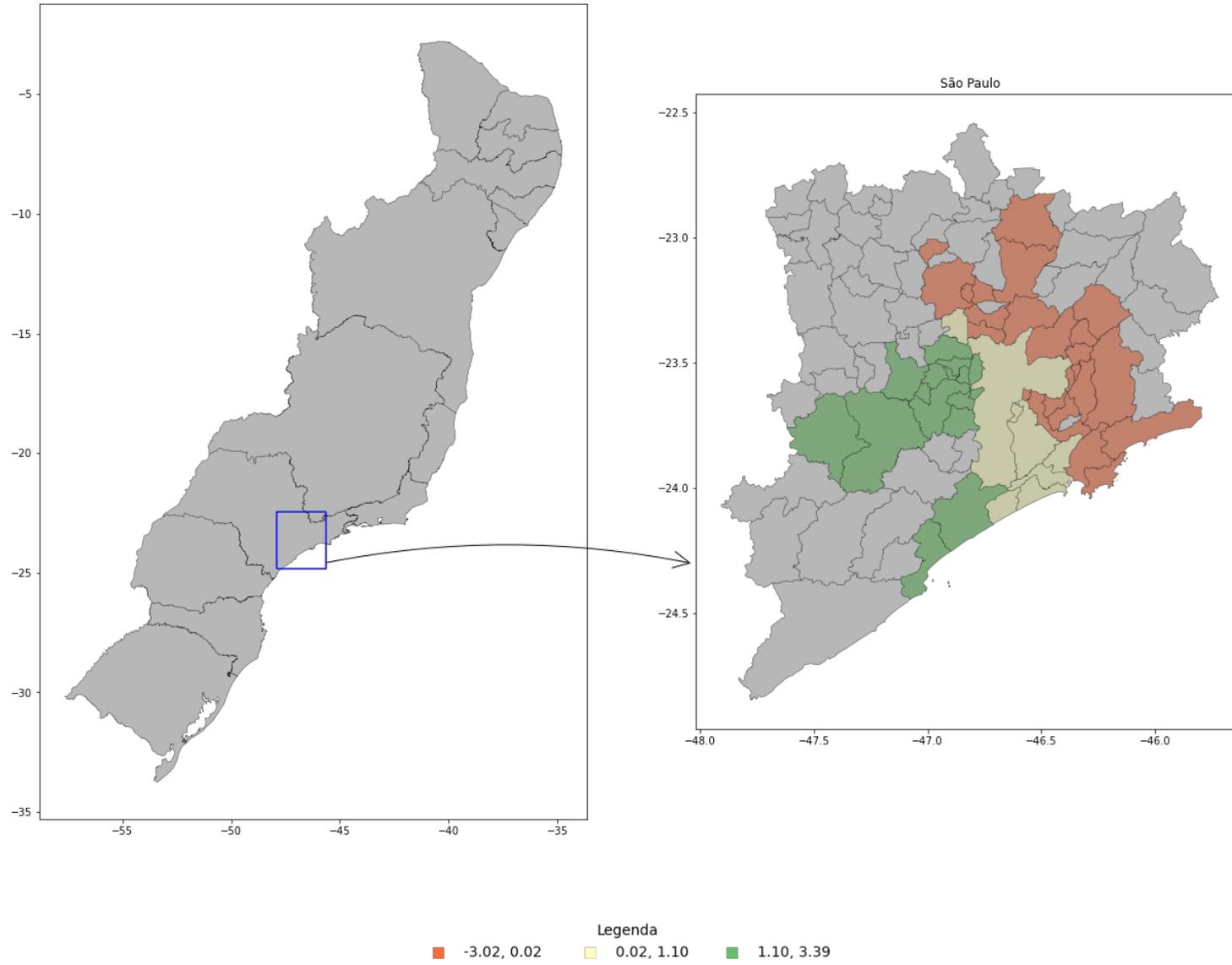


Figura 5.15 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 1-900

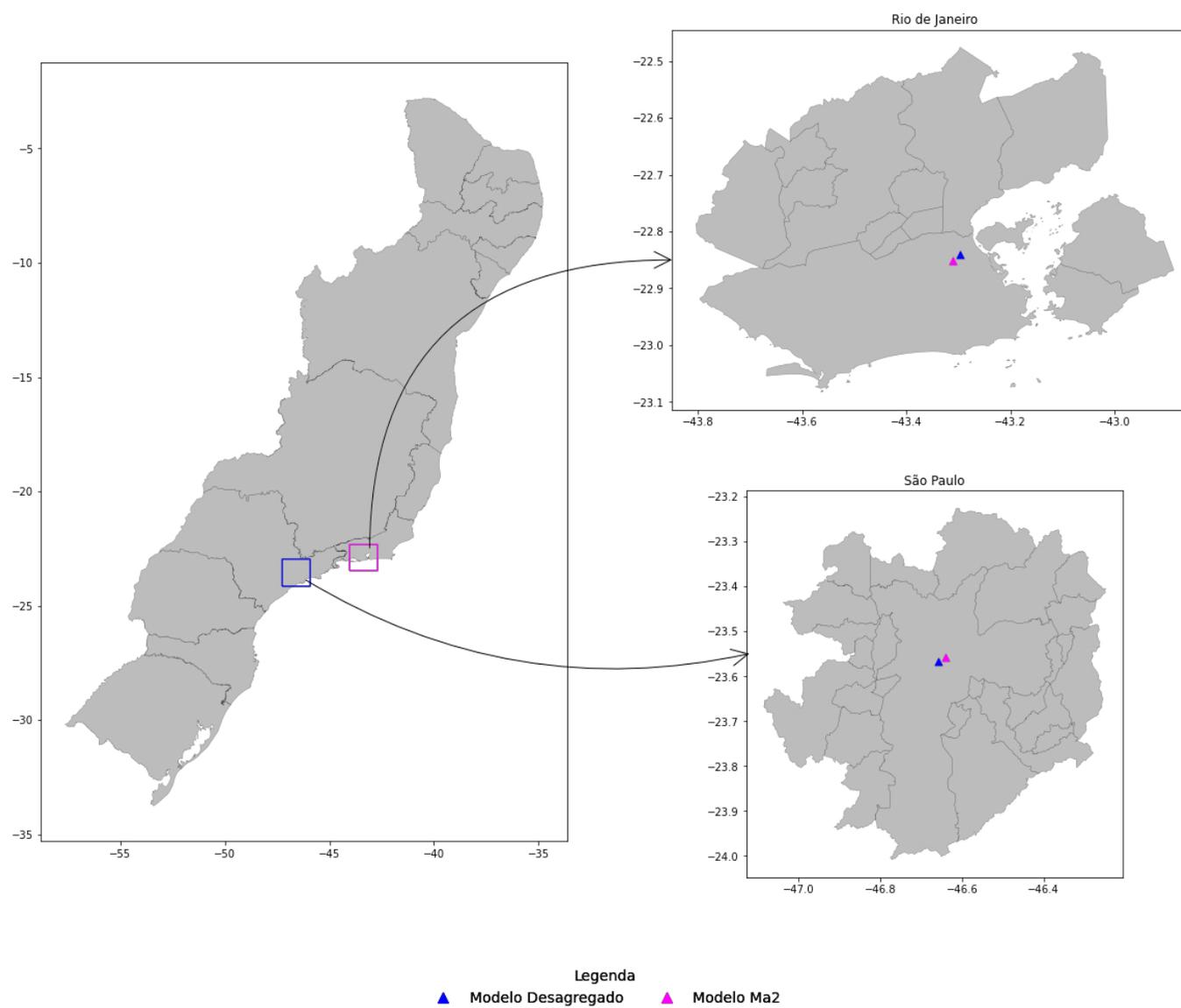


Figura 5.16 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 1-900

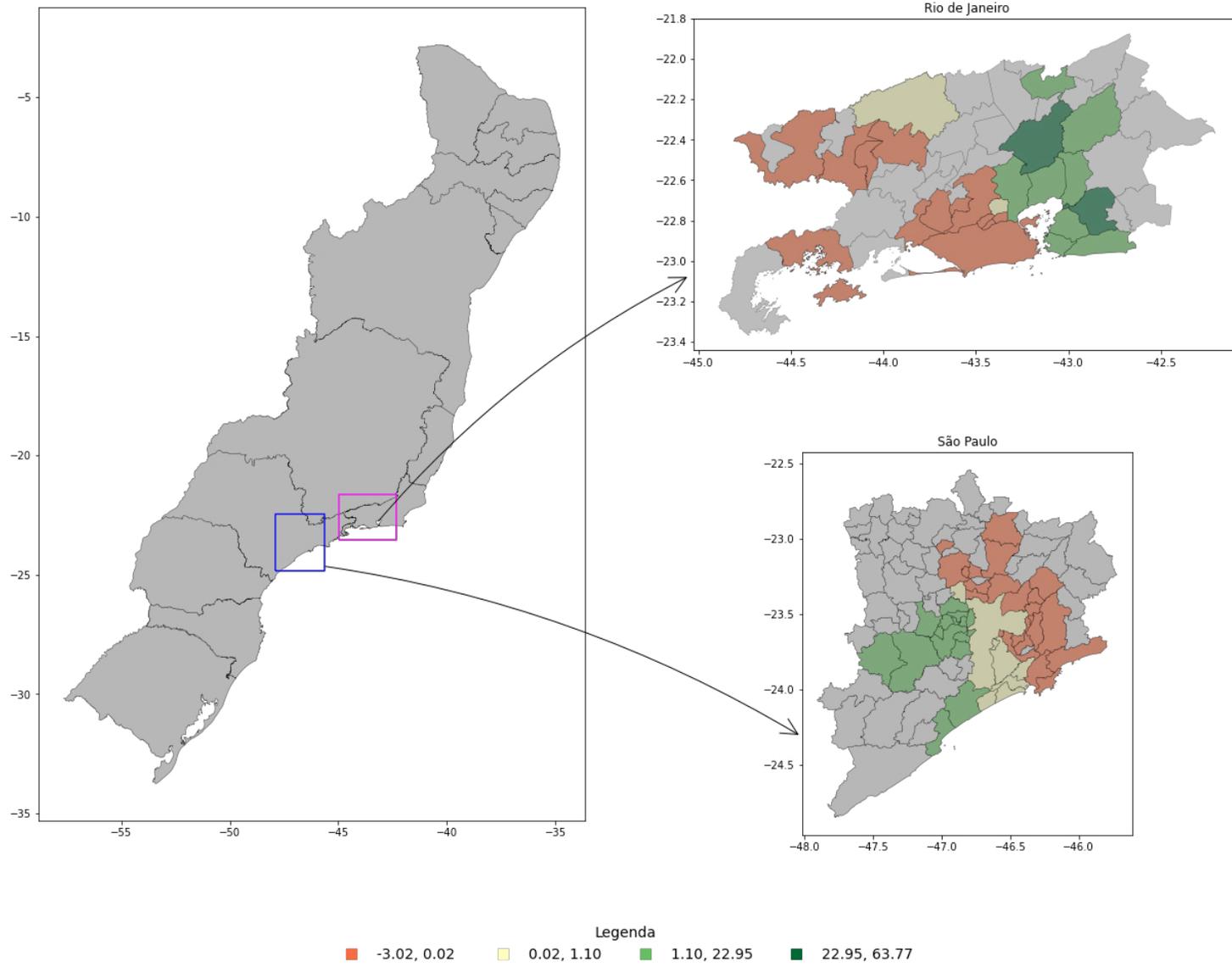


Figura 5.17 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 1-900

Já no município do Rio de Janeiro, pelo fato da instalação aberta se encontrar a sudoeste da aberta no modelo desagregado, observa-se que os municípios a leste e a norte estão coloridos de verde, uma vez que os custos de transporte no Modelo Ma2 aumentaram. Diferentemente dos municípios à oeste, que apresentam colorações vermelhas, indicando que a instalação aberta se encontra mais próxima deles. Ao contrário do que é esperado, alguns municípios mais afastados da capital, tais como Resende, apresentam custo de transporte menor no modelo agregado, o que pode significar que a instalação que atende esses pontos no modelo desagregado além de não ser aquela localizada no município do Rio de Janeiro é mais distante da instalação a partir da qual a demanda é atendida no Modelo Ma2.

Para a instância 1-900, observamos que os pontos de demanda que são alocados a instalações diferentes nos modelos agregado e desagregado também se encontram em regiões identificadas pelo DBSCAN. O Modelo Ma2 apresentou as maiores diferenças no valor da FO com relação ao modelo desagregado (0,03% se comparado ao valor próximo a zero do Modelo Ma1), assim como as maiores diferenças nos custos de transporte (entre -3,02% e 63,77% no Modelo Ma2, enquanto no Modelo Ma1 os custos de transporte variam entre -3,02% e 3,39%), indicativo de que as instalações abertas se encontram mais afastadas dos pontos de demanda no Modelo Ma2. Ou seja, nesta instância, o agrupamento de pontos de demanda pelo Modelo Ma1 está associado a resultados mais próximos do modelo desagregado.

Na Tabela 5.5 são mostradas as distâncias médias e maiores distâncias entre pontos de demanda e candidatos abertos mais próximos na solução ótima dos modelos agregados e desagregado. É possível ver que quanto maior o custo fixo e menor o número de candidatos abertos, maiores as distâncias entre pontos de demanda e candidatos abertos mais próximos (D_{max}), o que é esperado, tendo em vista que o menor número de candidatos abertos pode se encontrar mais distante de inúmeros pontos de demanda, especialmente aqueles de menor demanda.

Observa-se também na Tabela 5.5 que as distâncias médias dos pontos de demanda até as instalações mais próximas às quais são alocados (D_{avg}) não apresentam crescimento linear em cada instância. Nas instâncias 1-9, 1-90, 1-900 e 1-1800, nas quais ao menos um modelo agregado apresenta valor da FO maior que o modelo desagregado, as distâncias médias dos modelos agregados são maiores que no modelo desagregado. Por exemplo, na instância 1-9, o Modelo Ma1, que apresenta uma diferença de 0,43% no valor da FO com relação ao modelo desagregado, apresenta pontos de demanda mais distantes do candidato aberto mais próximo que o Modelo Ma2, que apresenta uma diferença de 0,03% com relação ao modelo

desagregado. A maior distância média do Modelo Ma1 se comparada ao Modelo Ma2 é esperada tendo em vista o menor número de candidatos abertos na solução ótima.

Tabela 5.5 – Distâncias médias e maiores distâncias (em metros) entre os pontos de demanda e os candidatos abertos mais próximos para os modelos que representam as instâncias do Conjunto 1

| Instância | | Desagregado | Método Ma1 | Método Ma2 |
|-----------|-------------------------------|-------------|------------|------------|
| 1-9 | Distância Média (D_{avg}) | 1.057,89 | 1.090,10 | 1.062,16 |
| | Maior distância (D_{max}) | 289,20 | 289,20 | 289,20 |
| 1-90 | Distância Média (D_{avg}) | 1.124,74 | 1.138,42 | 1.128,38 |
| | Maior distância (D_{max}) | 289,20 | 289,20 | 289,20 |
| 1-900 | Distância Média (D_{avg}) | 1.158,14 | 1.158,14 | 1.158,12 |
| | Maior distância (D_{max}) | 647,21 | 647,21 | 647,21 |
| 1-1800 | Distância Média (D_{avg}) | 1.173,70 | 1.173,73 | 1.173,70 |
| | Maior distância (D_{max}) | 652,00 | 652,01 | 652,00 |
| 1-4500 | Distância Média (D_{avg}) | 1.207,74 | 1.207,74 | 1.207,74 |
| | Maior distância (D_{max}) | 802,44 | 802,44 | 802,44 |
| 1-9000 | Distância Média (D_{avg}) | 1.130,81 | 1.130,81 | 1.130,81 |
| | Maior distância (D_{max}) | 802,44 | 802,44 | 802,44 |
| 1-18000 | Distância Média (D_{avg}) | 1.183,65 | 1.183,65 | 1.183,65 |
| | Maior distância (D_{max}) | 802,44 | 802,44 | 802,44 |
| 1-45000 | Distância Média (D_{avg}) | 1.274,66 | 1.274,66 | 1.274,66 |
| | Maior distância (D_{max}) | 898,96 | 898,96 | 898,96 |
| 1-90000 | Distância Média (D_{avg}) | 1.092,77 | 1.092,77 | 1.092,77 |
| | Maior distância (D_{max}) | 1.018,73 | 1.018,73 | 1.018,73 |

5.3.2. Conjunto 2

Para o Conjunto 2, também busca-se identificar regiões no espaço nas quais é possível existir diferenças de alocação entre os modelos agregado e desagregado, utilizando o algoritmo DBSCAN (Ester *et al.*, 1996). Para executar o algoritmo, são definidos os parâmetros Eps , que representa a maior distância entre pontos dentro de um mesmo agrupamento, e $MinPts$, que é associado a menor quantidade de pontos necessária para que um agrupamento seja formado.

Para o parâmetro $MinPts$, foi indicado o número 3, também utilizado em Rahmah e Sitanggang (2016), e no Conjunto 1, anteriormente apresentado, o que possibilita também identificar agrupamentos dentro de um mesmo município que estejam mais afastados uns dos outros. Já para estimar o parâmetro Eps , foram utilizados todos os pontos de demanda que estão

em municípios que possuem um número maior ou igual ao percentil 75 do número de pontos de demanda, ou seja, para o Conjunto 2, todos os pontos de demanda em municípios com 5 ou mais pontos foram utilizados no cálculo do parâmetro de distância. Foram considerados somente municípios com 5 ou mais pontos de demanda por estarem mais vezes associados aos agrupamentos que não são comuns aos Modelos Ma1 e Ma2.

O parâmetro *Eps* foi determinado através do ponto de inflexão da curva que associa a distância com relação aos vizinhos mais próximos. A curva em azul na Figura 5.18 associa a distância com relação aos vizinhos mais próximos e o número de pontos de demanda. Observa-se que o ponto de inflexão é aproximadamente 1,70 km, indicado pela linha tracejada em laranja, ou seja, a maioria dos vizinhos mais próximos encontra-se a uma distância de aproximadamente 1,70 km, um número adequado para o Conjunto 2, uma vez que a distância média entre os pontos de demanda nesses municípios analisados é de 1,28 km.

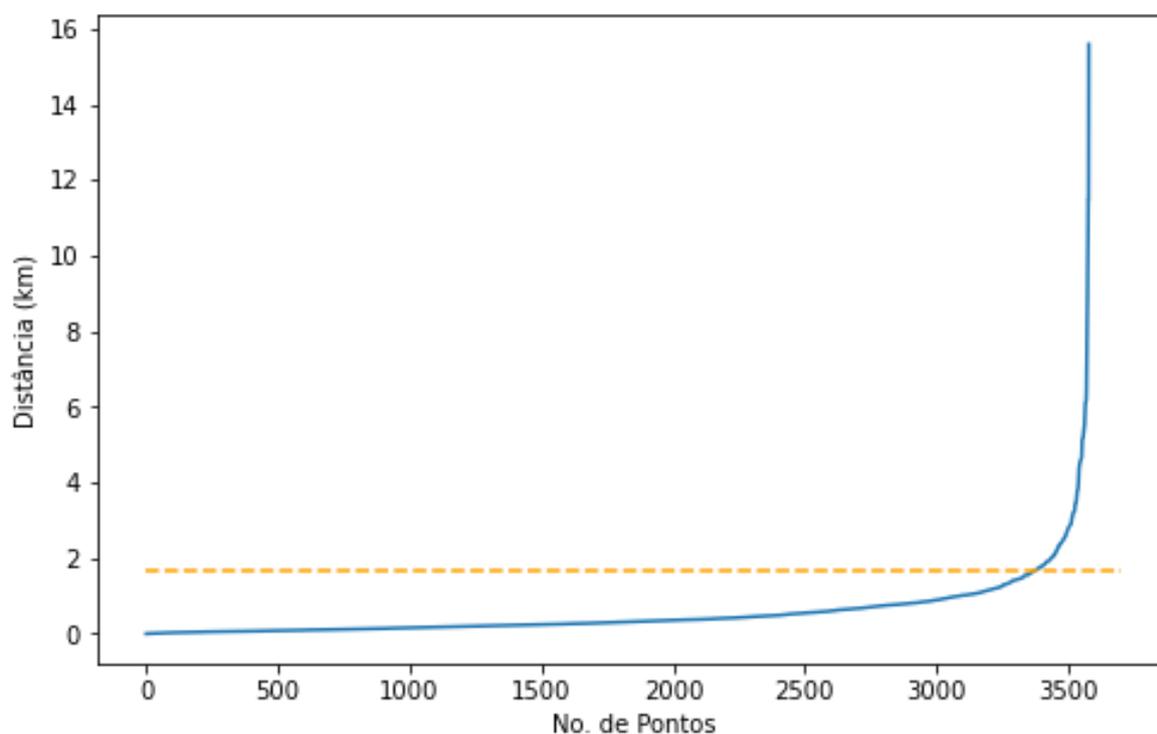


Figura 5.18 – Cálculo do parâmetro *Eps* para o Conjunto 2

Dos 136 municípios analisados pelo DBSCAN, foram observados 18 municípios nos quais há mais de um agrupamento, 10 municípios nos quais os agrupamentos correspondem a mais de uma localidade e 10 municípios que possuem um ou mais agrupamentos em seu território ou em que os pontos de demanda estão agrupados em mais de um município. Para os demais municípios, seus pontos de demanda foram agrupados por município utilizando o DBSCAN.

O número de municípios nos quais os pontos de demanda utilizados no algoritmo DBSCAN é diferente do número de municípios que não representam os mesmos agrupamentos dos Modelos Ma1 e Ma2. Isso é decorrente da consideração de municípios com 5 ou mais pontos de demanda, uma vez que existem municípios em regiões metropolitanas que possuem menor número de pontos de demanda. Todavia, destaca-se que não existe perda na identificação de regiões com possíveis alocações diferentes da demanda pelo uso de menor número de municípios no algoritmo DBSCAN.

A localização dos agrupamentos por DBSCAN considerando os municípios que possuem ou se encontram em mais de um agrupamento são mostrados na Figura 5.19. As cores das localidades representam os possíveis casos de agrupamentos formados pelo DBSCAN, sendo o Caso 2 (cor laranja) aquele no qual existe mais de um agrupamento em um mesmo município, o Caso 3 (cor azul) aquele no qual existem agrupamentos com pontos de demanda localizados em mais de um município, enquanto o Caso 4 (cor rosa), é aquele no qual existem tanto agrupamentos de pontos de demanda pertencentes somente ao município, quanto agrupamentos com pontos de demanda em mais de um município.

Na Figura 5.19, são vistos apenas agrupamentos nas capitais e em algumas cidades maiores do interior de São Paulo. Algumas capitais, como Porto Alegre, Rio de Janeiro e Salvador são indicadas pela cor laranja, o que significa que existe mais de um agrupamento do DBSCAN nesses municípios, porém seus pontos de demanda não formam agrupamentos com municípios vizinhos. Diferentemente de Belo Horizonte, Florianópolis e São Paulo, nas quais as capitais formam agrupamentos com municípios vizinhos, além de apresentarem outros agrupamentos somente com pontos de demanda em seu território. Esses agrupamentos são esperados, uma vez que o valor do parâmetro *Eps* e do conjunto de pontos de demanda que compõem o Conjunto 2 possibilita que agrupamentos com pontos de demanda em mais de um município sejam encontrados.

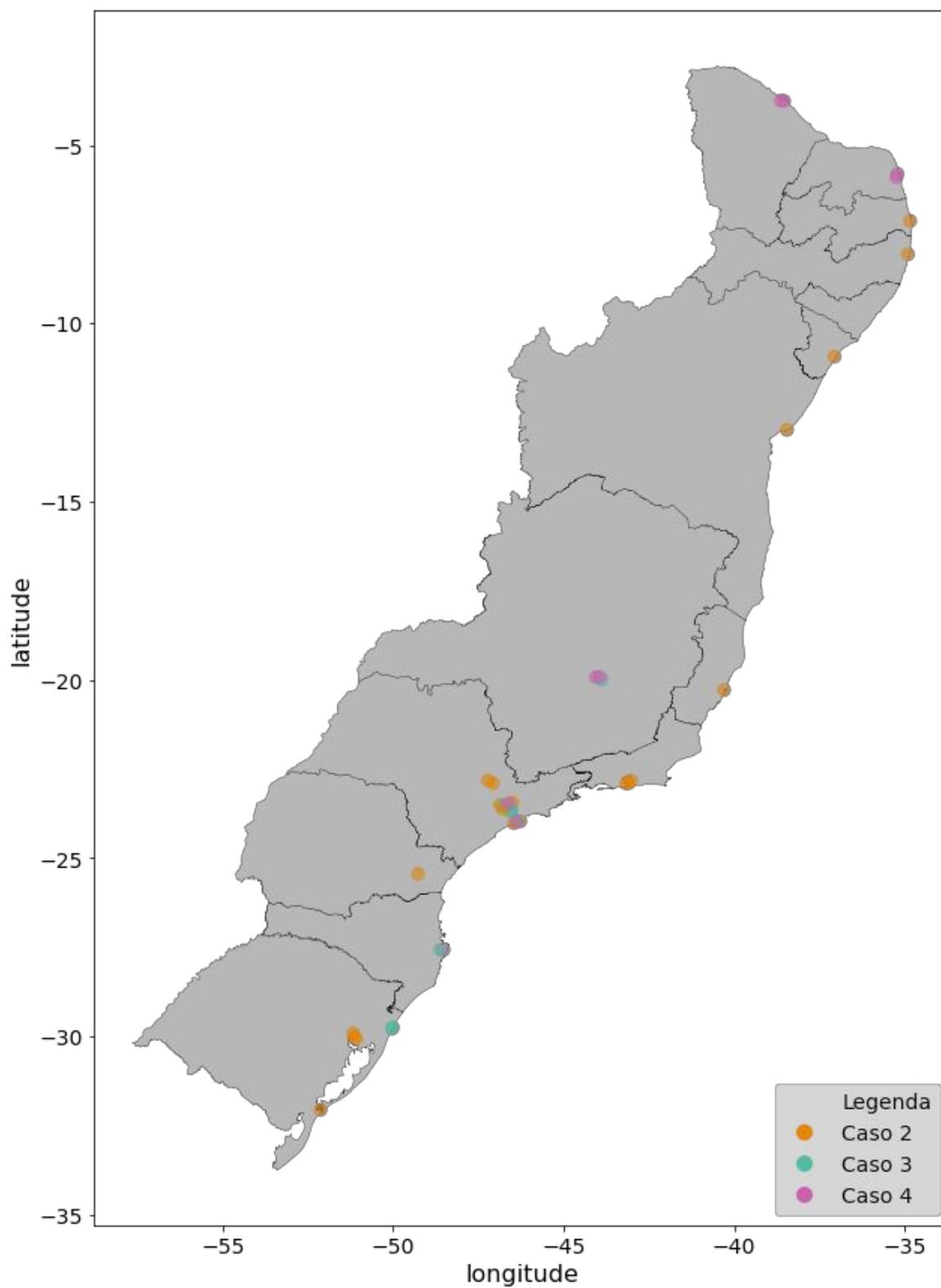


Figura 5.19 – Agrupamentos do DBSCAN para o Conjunto 2

Instância 2-9

Na instância 2-9, ambos os modelos agregados apresentam diferença no valor da FO com relação ao modelo desagregado. Para o Modelo Ma1, essa diferença é de 0,41%. Na

solução ótima desse modelo, são identificadas 103 abertas, das quais 99 também se encontram na solução do modelo desagregado. As regiões nas quais existem diferenças na abertura de candidatos, são mostradas na Figura 5.20. No Modelo Ma1, existem candidatos abertos em Minas Gerais, Rio de Janeiro, Rio Grande do Sul e São Paulo. Na Bahia, há somente candidatos abertos do modelo desagregado. Essa diferença no número e na localização dos candidatos abertos faz com que os pontos de demanda sejam atendidos a partir de candidatos diferentes no modelo agregado, levando à diferença no valor da FO entre o modelo desagregado e o Ma1.

Na Figura 5.21 são mostradas as diferenças nos custos de transporte por município para o Modelo Ma1. São observadas diferenças entre -6,73% e 18,47%, que indicam localidades nas quais o Modelo Ma1 apresenta custos menores e maiores que o modelo desagregado, respectivamente. Na Bahia, não existem instalações abertas unicamente no Modelo Ma1, e por isso a diferença entre custo de transporte entre dois modelos é positiva. O município mais distante, indicado em verde (Valença), é aquele que possui maior diferença de custos de transporte com relação ao modelo desagregado justamente pela maior distância.

Minas Gerais, Rio de Janeiro e Rio Grande do Sul, são alguns dos estados nos quais existem mais instalações abertas unicamente no modelo desagregado que no Modelo Ma1. Nestes estados, é possível ver que mesmo abrindo uma instalação, os custos de transporte do modelo agregado ainda são maiores que no modelo desagregado, como no caso da cidade do Rio de Janeiro, na qual o modelo agregado apresenta custos de transporte 13,13% maiores que no modelo desagregado. Novamente, as diferenças nos custos de transporte são decorrentes da maior proximidade ou distância entre candidatos abertos e pontos de demanda.

No Modelo Ma1, São Paulo é o único estado no qual existem localidades com custo de transporte menor no modelo agregado se comparado ao custo do modelo desagregado. Essa diferença é decorrente da abertura de uma instalação em Osasco, mais próxima de alguns pontos de demanda que as demais abertas em São Paulo no modelo desagregado. Por se encontrar mais a oeste que os candidatos do modelo desagregado, os custos são menores nos municípios a oeste (cor vermelha), que a leste (cor verde) do candidato aberto.

Por sua vez, o Modelo Ma2 apresenta uma diferença de 0,03% no valor da FO com relação ao modelo desagregado. Das 117 instalações abertas no Modelo Ma2, 114 também são abertas no modelo desagregado. As instalações que são diferentes nos dois modelos encontram-se em Minas Gerais, Rio Grande do Sul e São Paulo, como mostrado na Figura 5.22. Essa

diferença na localização dos candidatos abertos pode ser decorrente do agrupamento de pontos de demanda do Modelo Ma2.

Na Figura 5.23 são mostradas as diferenças nos custos de transportes para os municípios com ao menos um ponto de demanda alocado a instalações diferentes no Modelo Ma2. É possível ver que as diferenças nos custos de transporte variam entre -23,25% e 15,76%. Em Minas Gerais, pelo fato de existir uma instalação aberta no Modelo Ma2, os custos de transporte no modelo agregado são menores que no modelo desagregado.

Nos estados do Rio Grande do Sul e São Paulo são observadas tanto diferenças positivas quanto negativas nos custos de transporte. Isso é esperado, uma vez que existem tanto instalações que são abertas ou no modelo desagregado ou no Modelo Ma2 nessas regiões, fazendo com que os custos de transporte para os pontos de demanda mais próximos a essas instalações no Modelo Ma2 sejam menores, e aumentando os custos de transporte para os pontos mais afastados. Em São Paulo, no município de Osasco onde o candidato aberto no Modelo Ma2 se localiza, como esperado, os custos de transporte no modelo Ma2 são menores que no modelo desagregado. Em outros municípios, como Cajamar e Itapevi, dada a distância entre os pontos de demanda e os candidatos abertos que são diferentes entre o Modelo Ma2 e o desagregado, é esperado que as diferenças no custo de transporte sejam maiores que nos vizinhos São Paulo e Barueri, porém é possível ver pela cor (verde claro, se comparado ao verde escuro) que os custos de transporte são menores, o que sugere que esses pontos de demanda sejam atendidos por outra instalação aberta como resultado do modelo desagregado, senão aquelas abertas na cidade de São Paulo.

Na instância 2-9, foi possível ver que o algoritmo DBSCAN identificou as regiões nas quais a alocação dos pontos de demanda é diferente nos modelos desagregado e agregados. O Modelo Ma1 resultou 14 candidatos abertos a menos que os Modelos Ma2 e o desagregado, o que pode justificar a maior diferença no valor da FO com relação ao modelo desagregado. Porém, o Modelo Ma2 foi aquele no qual os custos de transporte mais variaram com relação ao modelo desagregado (entre -23,25% e 15,76% no Modelo Ma2, e entre -6,73% e 18,47% no Modelo Ma1), o que significa que os candidatos abertos se encontram ou mais próximos ou mais afastados dos pontos de demanda que os candidatos do Modelo Ma1. Para a instância 2-9, com maior número de instalações abertas em decorrência do menor custo fixo, o Modelo Ma2 é mais adequado com agrupamento de pontos de demanda, pela menor diferença no valor da FO com relação ao modelo desagregado (0,03% em comparação a 0,41%), ainda que as diferenças de custos de transportes sejam maiores que no modelo Ma1.

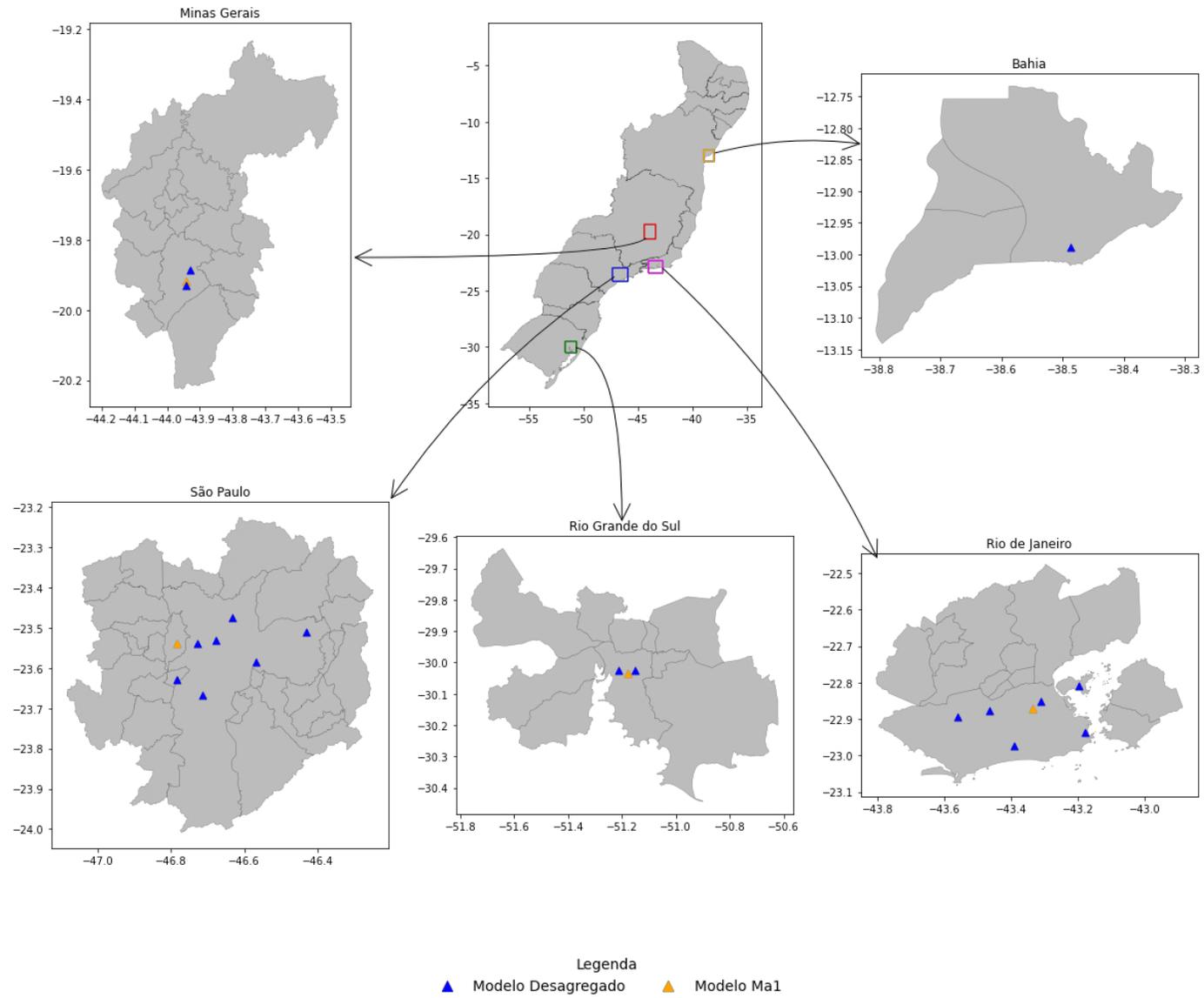


Figura 5.20 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 2-9

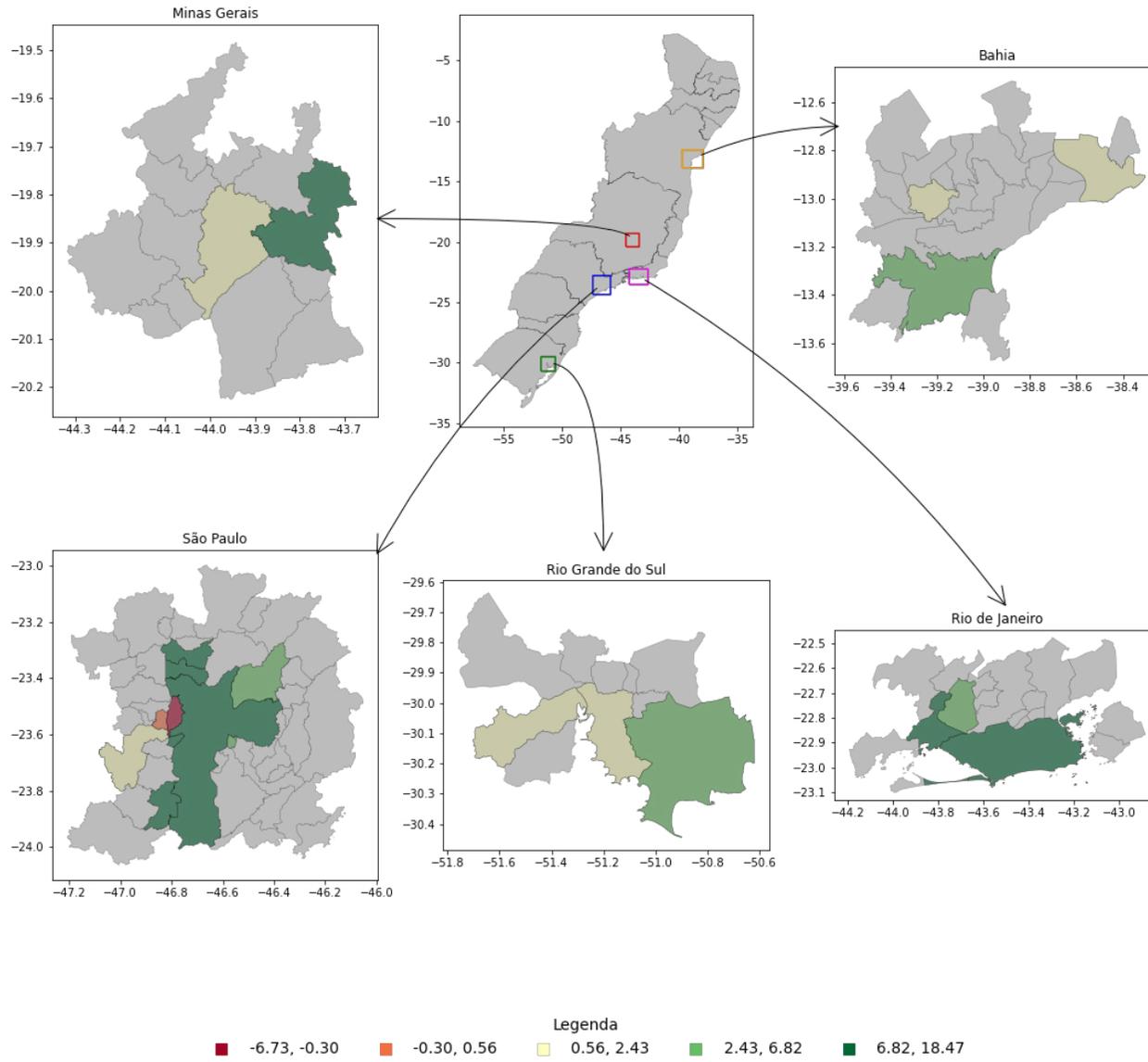


Figura 5.21 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 2-9

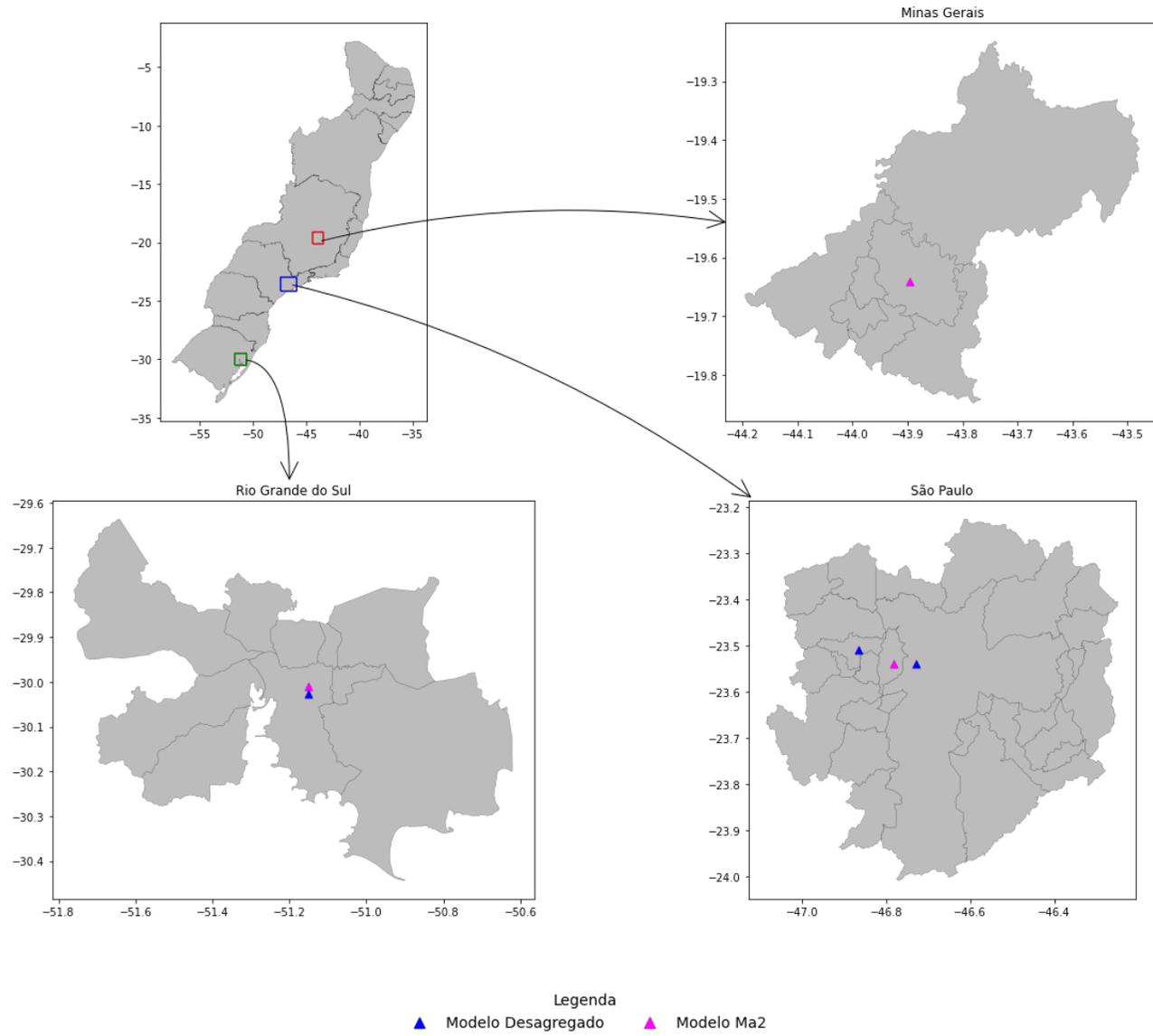


Figura 5.22 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 2-9

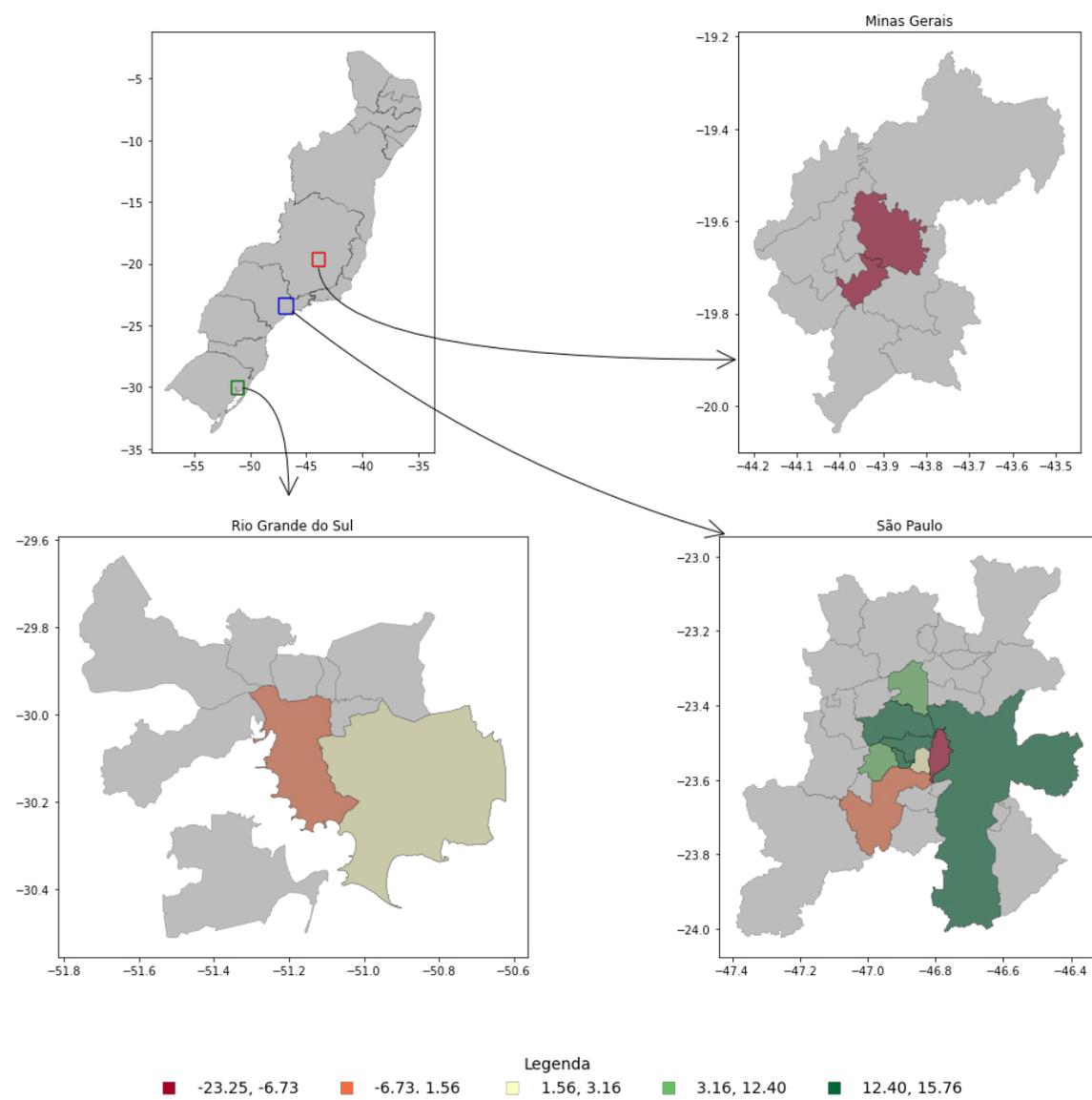


Figura 5.23 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 2-9

Instância 2-4500

Na instância 2-4500 ambos os modelos agregados apresentam uma diferença de 0,01% com relação ao valor da FO do modelo desagregado. Para o Modelo Ma1, foi observada 1 instalação aberta a menos que no modelo desagregado, de forma que dentre os 14 candidatos abertos na solução ótima, 13 são comuns ao modelo desagregado. A Figura 5.24 mostra a localização das instalações que são abertas somente nos modelos desagregado ou Ma1. Nota-se que a instalação aberta no Modelo Ma1 se encontra na Paraíba, enquanto aquelas abertas no modelo desagregado se encontram em Pernambuco e no Rio Grande do Norte.

Os custos de transporte para cidades nos estados de Alagoas, Bahia, Paraíba, Pernambuco e Rio Grande do Norte são mostradas na Figura 5.25. É possível ver que os custos de transporte variam entre -84,99% e 662,90%, o que significa que existem municípios mais próximos ao candidato aberto no Modelo Ma1 que possuem custos de transporte menores que o modelo desagregado, assim como municípios mais distantes e com custos de transporte maiores. É a diferença no número e na localização dos candidatos abertos que leva o Modelo Ma1 a apresentar diferença no valor da FO com relação ao modelo desagregado.

Por sua vez, o Modelo Ma2 apresenta o mesmo número de instalações abertas na solução ótima que o modelo desagregado e, dos 15 candidatos abertos no modelo Ma2, 14 também se encontram abertos no modelo desagregado. Nessa instância, a solução ótima do Modelo Ma2 indica a abertura de uma instalação diferente na cidade de São Paulo, como mostrado na Figura 5.26. É possível ver que a instalação que abriu no Modelo Ma2 encontra-se mais a sul e a leste que a instalação do modelo desagregado.

Com relação à diferença nos custos de transporte, a Figura 5.27 mostra diferenças entre -7,71% e 110,31% para municípios nos estados de Minas Gerais e São Paulo. Em decorrência da localização do candidato aberto no Modelo Ma2, observa-se que os custos de transporte nos municípios mais a sul e a leste (cor vermelha) são menores no modelo Ma2 se comparado ao modelo desagregado, enquanto em alguns municípios a norte e a oeste (cor verde), o custo de transporte do modelo agregado é maior. Diferentemente do esperado, nos municípios destacados pela cor bege, as diferenças nos custos de transporte são menores que em municípios mais próximos. Isso acontece porque nessas localidades, os candidatos abertos nos modelos Ma2 e desagregado são igualmente afastados, o que leva a uma diferença menor nesses municípios.

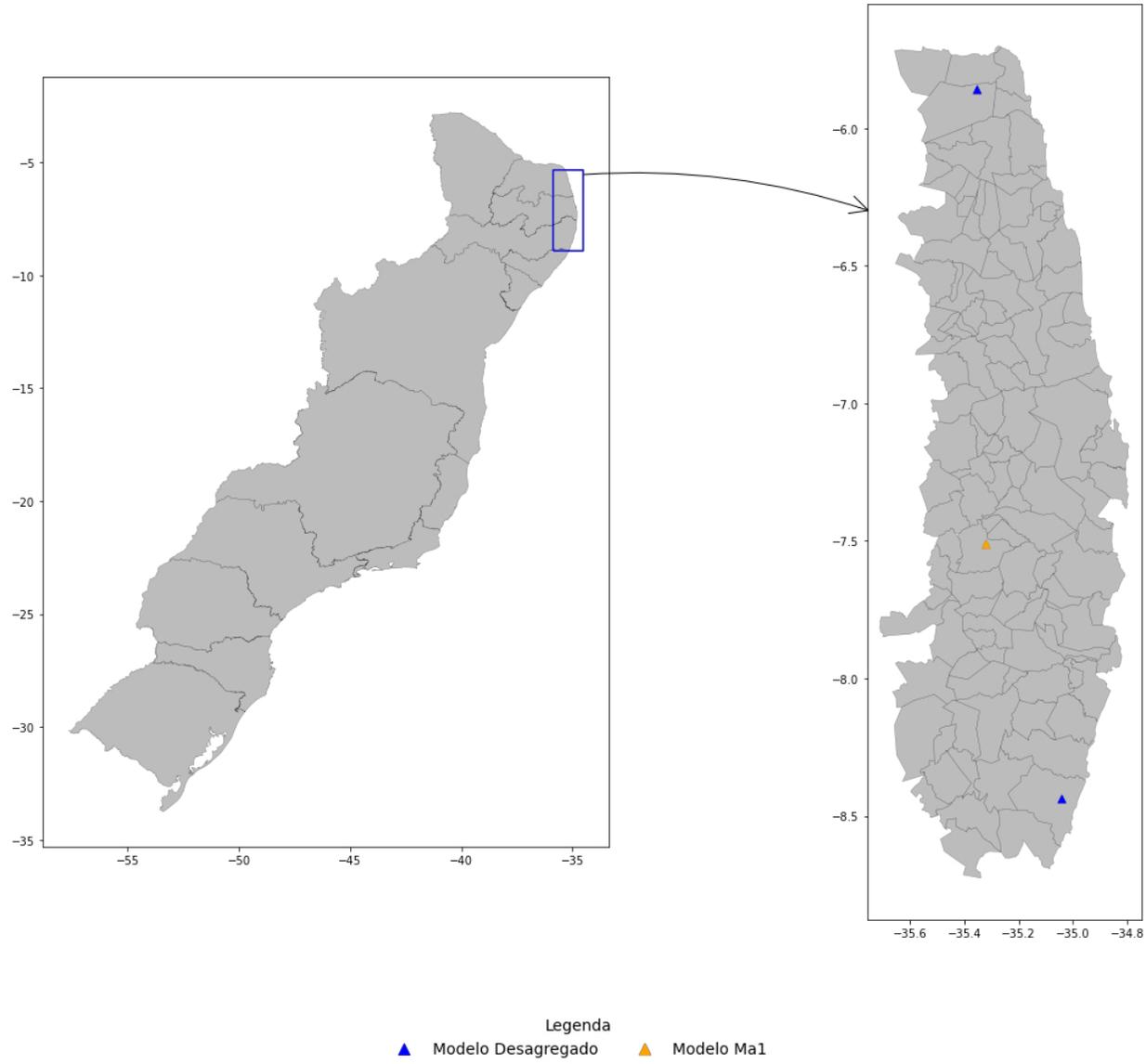


Figura 5.24 – Candidatos diferentes entre os modelos desagregado e Ma1 para a instância 2-4500

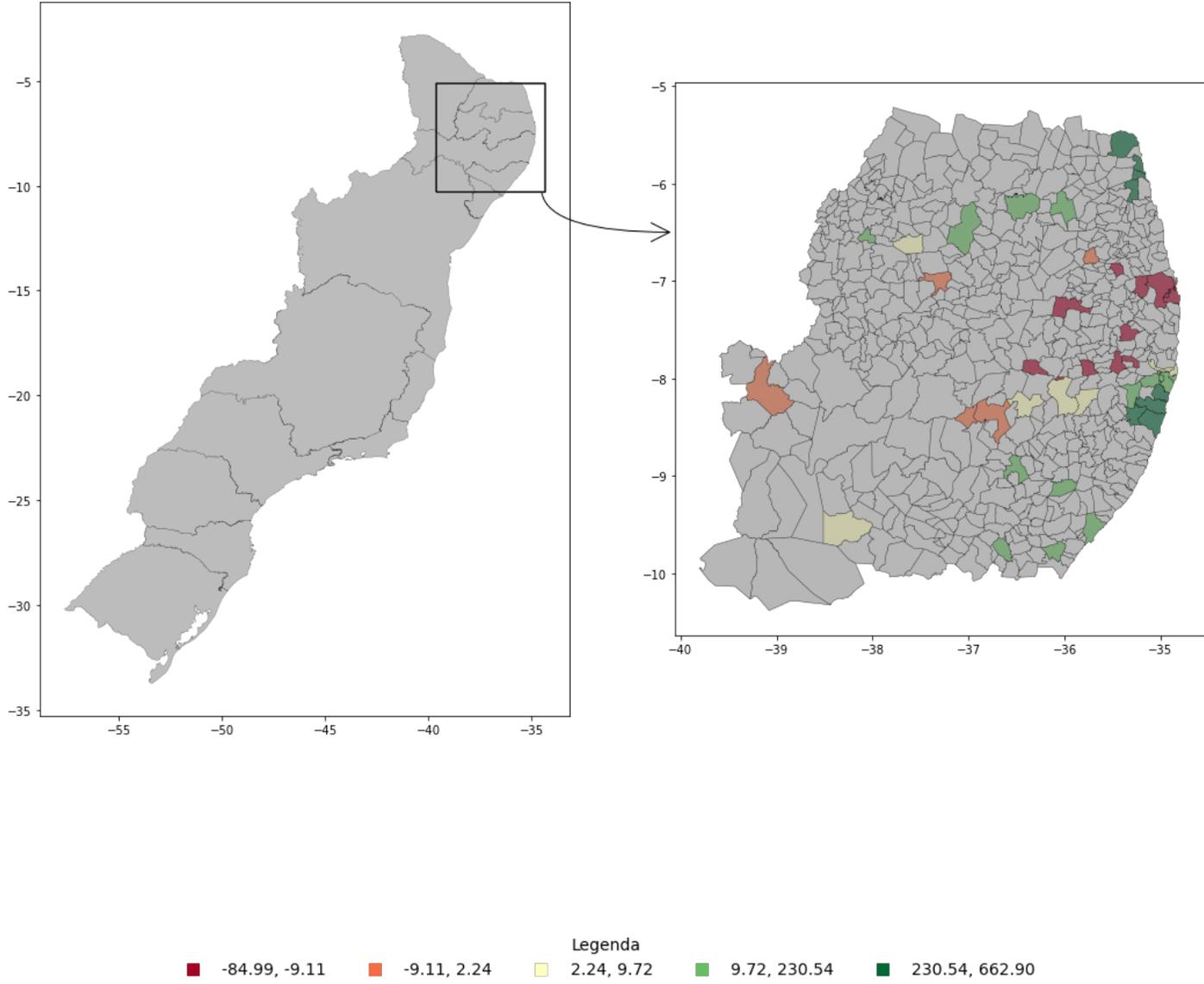


Figura 5.25 – Diferença de custo de transporte entre os modelos desagregado e Ma1 para a instância 2-4500

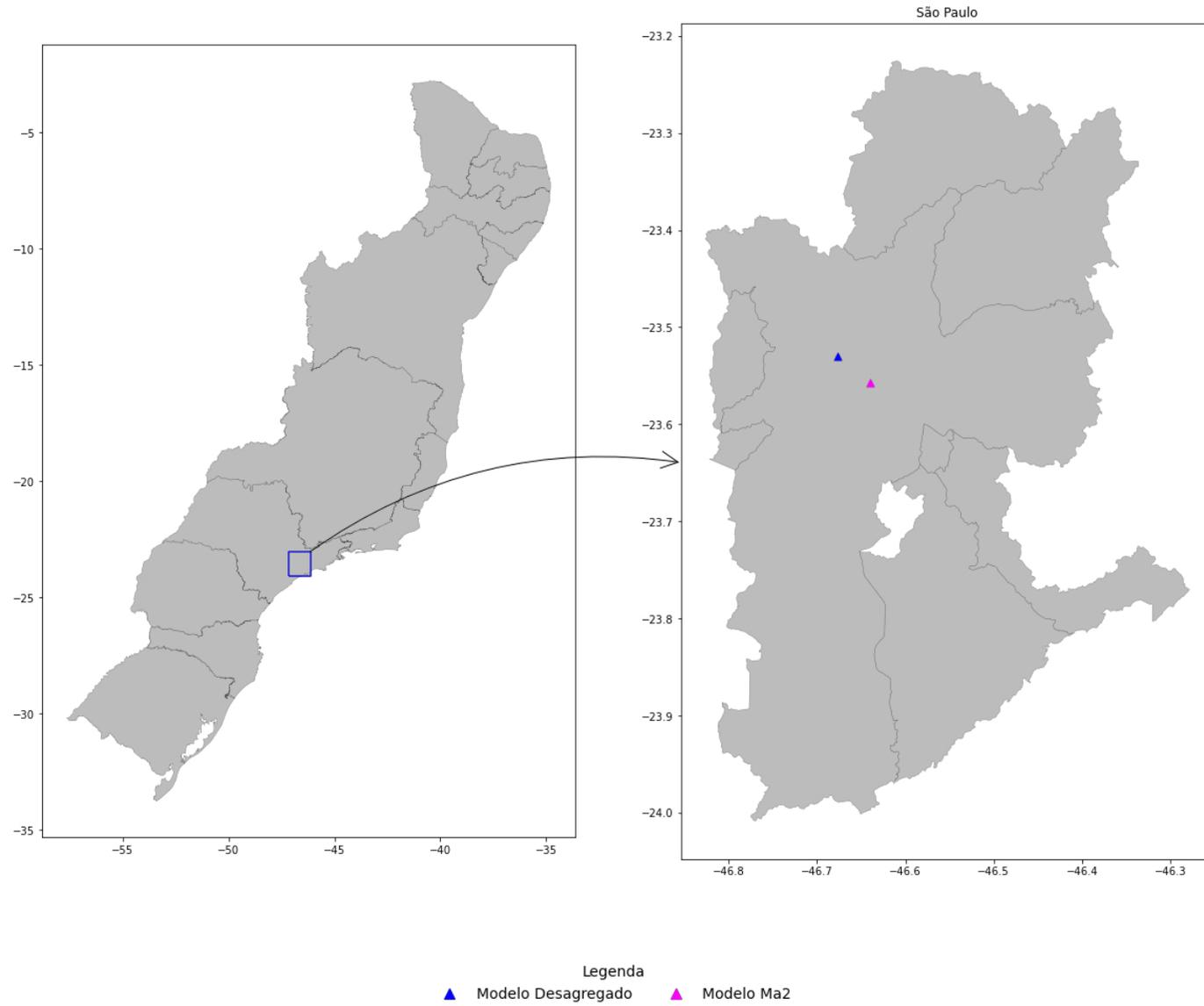


Figura 5.26 – Candidatos diferentes entre os modelos desagregado e Ma2 para a instância 2-4500

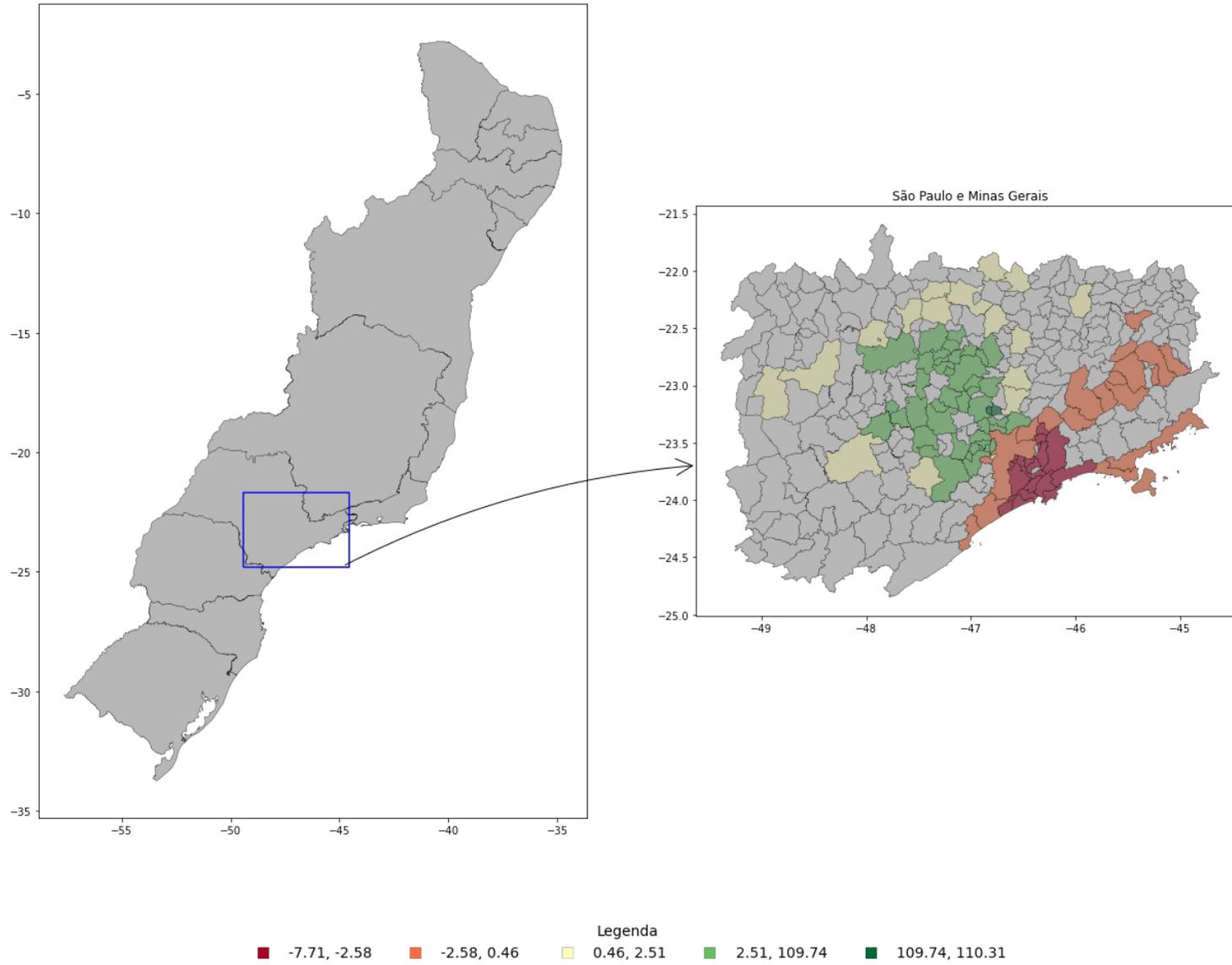


Figura 5.27 – Diferença de custo de transporte entre os modelos desagregado e Ma2 para a instância 2-4500

Para a instância 2-4500, o algoritmo DBSCAN identificou apenas parte as localidades nas quais os pontos de demanda foram alocados a instalações diferentes nos modelos agregados e desagregado, em especial no Modelo Ma2. O Modelo Ma1 apresenta um valor de FO mais próximo ao do modelo desagregado, e maiores diferenças nos custos de transporte se comparado ao Modelo Ma2 (entre -84,99% e 662,90% no Modelo Ma1, e entre -7,71% e 110,31% no Modelo Ma2), ou seja, no Modelo Ma1, as instalações abertas estão ou muito mais próximas ou muito mais afastadas que no Modelo Ma2. Ainda que o Modelo Ma2 apresente menores diferenças nos custos de transporte, o Modelo Ma1 é mais indicado para agrupar os pontos de demanda na instância 2-4500, com aproximadamente 10% do número de candidatos abertos e um custo fixo igual a metade do custo base, uma vez que a diferença no valor da FO com relação ao modelo desagregado é menor.

Por fim, para as soluções ótimas de cada modelo são calculadas as distâncias médias e as maiores distâncias entre pontos de demanda e candidatos abertos mais próximos. A Tabela 5.6 mostra essas distâncias para as instâncias do Conjunto 2. É possível ver que as maiores distâncias entre pontos de demanda e candidatos abertos (D_{max}) aumenta, à medida que o número de candidatos abertos diminui, o que é esperado, tendo em vista que o menor número de candidatos pode se encontrar mais distante de diversos pontos de demanda, em especial aqueles já mais afastados dos candidatos abertos.

Com relação as distâncias médias (D_{avg}), observa-se na Tabela 5.6 que não há aumento linear nessa distância, à medida que aumentam os custos fixos. É possível ver nas instâncias 2-9, 2-90, 2-4500 e 2-9000, nas quais o valor da FO de ao menos um dos modelos agregados é diferente do modelo desagregado, que as distâncias médias não são as mesmas para todos os modelos. Essa diferença na distância média é decorrente da abertura de instalações diferentes no modelo agregado. Nas instâncias 2-9, 2-90 e 2-9000, o Modelo Ma1 é aquele que apresenta maiores distâncias médias e maior diferença no valor da FO com relação ao modelo desagregado, diferentemente da instância 2-4500, na qual esse modelo apresenta a menor distância média entre pontos de demanda e candidatos abertos. Isso significa que para a instância 2-4500, os pontos de demanda estão mais próximos aos candidatos abertos no Modelo Ma1 que no modelo desagregado, e isso pode ser decorrente da abertura de instalações diferentes no modelo agregado, e não necessariamente implica em um menor valor da FO.

Tabela 5.6 – Distâncias médias e maiores distâncias (em metros) entre os pontos de demanda e os candidatos abertos mais próximos para os modelos que representam as instâncias do Conjunto 2

| Instância | | Desagregado | Método Ma1 | Método Ma2 |
|-----------|-------------------------------|-------------|------------|------------|
| 2-9 | Distância Média (D_{avg}) | 1.053,86 | 1.087,42 | 1.055,11 |
| | Maior distância (D_{max}) | 404,10 | 404,10 | 404,10 |
| 2-90 | Distância Média (D_{avg}) | 1.126,99 | 1.127,23 | 1.126,99 |
| | Maior distância (D_{max}) | 581,93 | 581,93 | 581,93 |
| 2-900 | Distância Média (D_{avg}) | 1.159,90 | 1.159,90 | 1.159,90 |
| | Maior distância (D_{max}) | 652,01 | 652,01 | 652,01 |
| 2-1800 | Distância Média (D_{avg}) | 1.163,05 | 1.163,05 | 1.163,05 |
| | Maior distância (D_{max}) | 888,37 | 888,37 | 888,37 |
| 2-4500 | Distância Média (D_{avg}) | 1.191,71 | 1.131,81 | 1.191,71 |
| | Maior distância (D_{max}) | 891,38 | 891,38 | 891,38 |
| 2-9000 | Distância Média (D_{avg}) | 1.180,93 | 1.181,75 | 1.180,93 |
| | Maior distância (D_{max}) | 891,38 | 891,38 | 891,38 |
| 2-18000 | Distância Média (D_{avg}) | 1.208,67 | 1.208,67 | 1.208,67 |
| | Maior distância (D_{max}) | 985,92 | 985,92 | 985,92 |
| 2-45000 | Distância Média (D_{avg}) | 1.088,56 | 1.088,56 | 1.088,56 |
| | Maior distância (D_{max}) | 1.018,73 | 1.018,73 | 1.018,73 |
| 2-90000 | Distância Média (D_{avg}) | 1.211,40 | 1.211,40 | 1.211,40 |
| | Maior distância (D_{max}) | 1.054,03 | 1.054,03 | 1.054,03 |

6. CONSIDERAÇÕES FINAIS

Esta pesquisa tratou de analisar o impacto na qualidade da solução de problemas de localização de instalações ao agrupar a demanda. Este é um assunto relevante dentro da modelagem de problemas de localização, haja vista que mesmo com a melhoria dos computadores e dos programas de otimização, a obtenção de soluções ótimas para esses problemas ainda pode requerer bastante tempo. Enxergamos uma oportunidade no agrupamento de pontos de demanda como representação adequada do problema original, com menor tempo de processamento.

Este trabalho compreendeu a análise de instâncias do FCFL baseadas em distribuições reais B2B ao invés de utilizar instâncias consagradas da literatura. Foi possível observar que ainda que se trate de instâncias baseadas em um sistema B2B, atualmente já existem problemas muito maiores apresentados na literatura e para os quais os programas de otimização, como o Gurobi, têm capacidade de obter soluções ótimas. Ressalta-se também que diferentemente do que as pesquisas mais recentes sobre os problemas de instalações com custo fixo que têm utilizado principalmente heurísticas para a obtenção de soluções ótimas ou quase-ótimas, foi utilizado um método exato para a obtenção de soluções, o que possibilitou uma análise mais precisa e acurada dos resultados e da comparação relativa do impacto da agregação de demanda nas configurações otimizadas resultantes.

Foram escolhidos e investigados dois métodos de formação de agrupamentos: Ma1, que agrupa os pontos de demanda com relação à localização geográfica do município, e Ma2, que utiliza o algoritmo *K-Means* para obter um número K de agrupamentos igual ao número de municípios. O racional por trás da utilização de métodos de agrupamento tão diferentes justifica-se no fato de que representar a demanda de um município por apenas um ponto, localizado no seu respectivo baricentro, prática usual em aplicações reais de projeto de malha, pode não levar em conta a existência de outros subgrupos dentro do agrupamento principal e levaria a uma diferença maior que 1% no valor da FO se comparada ao modelo desagregado.

As instâncias utilizadas pra avaliar os métodos propostos representam problemas reais, baseadas na localização de pontos de venda de empresas farmacêuticas, que se encaixam em um sistema de distribuição B2B. No total, foram criados dois conjuntos de dados dos quais as 18 instâncias do FCFL foram baseadas.

Para as instâncias analisadas, foi possível observar que, em ambos os modelos agregados os tempos de processamento e o número de operações do *simplex* necessárias para

obter a solução ótima diminuíram, como consequência da diminuição do tamanho do problema. Dentre os resultados das instâncias de cada conjunto, é possível concluir que o Modelo Ma2 deve ser escolhido para agrupar os pontos de demanda, uma vez que seus resultados são iguais ou mais próximos aos resultados do modelo desagregado, se comparado ao Modelo Ma1. Além disso, mesmo para instâncias nas quais ambos os modelos agregados apresentam soluções diferentes do modelo desagregado, as instalações abertas no Modelo Ma2 encontram-se mais próximas aos pontos de demanda que aquelas abertas no Modelo Ma1. Destaca-se que ao aumentar o custo fixo mensal das instalações, os efeitos de agregação não se mostraram significantes, de maneira que ao utilizar o Modelo Ma2, garantimos um valor da FO mais próximo ao problema desagregado para todas as instâncias, se comparado ao Modelo Ma1.

Como sugestão de extensão desta pesquisa, sugere-se a utilização de instâncias associadas a um problema de p -medianas ou ao problema de localização de instalações capacitadas com custo fixo, para analisar o impacto das modelagens na qualidade dos resultados dos modelos de otimização. Este trabalho utilizou como base dois modelos para formação de agrupamentos: a localização geográfica dos pontos de demanda por município, e os agrupamentos indicados pelo algoritmo *K-Means*. Esses são apenas dois métodos de formação de agrupamento dentre diversos outros, como métodos de encadeamento ou até mesmo o algoritmo DBSCAN, para os quais é possível analisar a qualidade dos métodos de formação de agrupamentos. Vale ressaltar que dentre os diferentes métodos de formação de agrupamentos existentes, o DBSCAN é o menos utilizado para agrupar pontos de demanda. Outros estudos futuros permitiriam ir mais além e estudar a forma como a demanda se distribui no espaço e como isso altera a envoltória convexa do problema agregado e leva a abertura de um número diferente de candidatos, em especial, nos modelos com pontos de demanda agregados; além de comparar se o agrupamento de pontos de demanda permite a substituição do uso de heurísticas para a obtenção de soluções ótimas para algumas instâncias.

REFERÊNCIAS

- AARDAL, K. Capacitated facility location: Separation algorithms and computational experience. **Mathematical Programming, Series B**, v. 81, n. 2, p. 149–175, 1998.
- AARDAL, K.; POCHET, Y.; WOLSEY, L. A. Capacitated Facility Location: Valid Inequalities and Facets. **Mathematics of Operations Research**, v. 20, n. 3, p. 562–582, 1995.
- ANDERSSON, G.; FRANCIS, R.; NORMAK, T.; RAYCO, M. Aggregation method experimentation for large-scale network location problems. **Location Science**, v. 6, n. 1–4, p. 25–39, 1 maio 1998.
- ANKERST, M.; BREUNIG, M.; KRIEGEL, H.; SANDER, J. OPTICS: Ordering Points to Identify the Clustering Structure. **Association for Computing Machinery**, v. 28, n. 2, 1999.
- ARENALES, M.; ARMENTANO, V.; MORABITO, R.; YANASSE, H. **Pesquisa operacional**. Rio de Janeiro: Campus Elsevier, 2007
- AURENHAMMER, F. Voronoi Diagrams - A Survey of a Fundamental Data Structure. **ACM Computing Surveys (CSUR)**, v. 23, n. 3, p. 345–405, 1991.
- BALLOU, R. **Gerenciamento da cadeia de suprimentos/logística empresarial**. Porto Alegre: Bookman, 2006.
- BARAHONA, F.; CHUDAK, F. A. Near-optimal solutions to large-scale facility location problems. **Discrete Optimization**, v. 2, n. 1, p. 35–50, 2005.
- BARCELO, J., HALLEFJORD, Å., FERNANDEZ, E., JÖRNSTEN, K. Lagrangean relaxation and constraint generation procedures for capacitated plant location problems with single sourcing. **OR Spektrum**, v.12, p. 79–88, 1990.
- BEASLEY, J. OR-Library: Distributing Test Problems by Electronic Mail. **Journal of the Operational Research Society**, v. 41, p. 1069–1072, 1990.
- BOEING, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. **Computers, Environment and Urban Systems**, v. 65, p. 126–139, 1 set. 2017.
- BRASIL. Resolução nº 5.949, de 13 de julho de 2021. Altera o Anexo II da Resolução ANTT nº 5.867, de 14 de janeiro de 2020, em razão do disposto nos parágrafos 1º e 2º do artigo 5º da Lei nº 13.703, de 8 de agosto de 2018. **Diário Oficial [da] República Federativa do Brasil**: seção 1, Brasília, DF, p. 63, 14 jul. 2021.

BUENO, A. **Preço médio do aluguel de salas e conjuntos comerciais acelera em maio.**

São Paulo, 2019. Disponível em <https://fipezap.zapimoveis.com.br/preco-medio-do-aluguel-de-salas-e-conjuntos-comerciais-acelera-em-maio/>. Acesso em 07 mar. 2022.

BURKEY, M. L.; BHADURY, J.; EISELT, H. A. Voronoi diagrams and their uses. In: EISELT, H. A.; MARIANOV, V. (Eds.). **Foundations of Location Science**. New York: Springer, 2011.

CEBECAUER, M.; BUZNA, L. A versatile adaptive aggregation framework for spatially large discrete location-allocation problems. **Computers and Industrial Engineering**, v. 111, p. 364–380, 2017.

CHURCH, R. L.; REVELLE, C. S. The Maximal Covering Facility Location. **Papers of the Regional Science Association**, v. 32, p. 101–118, 1974

CORNUÉJOLS, G.; NEMHAUSER, G.; WOLSEY, L. The uncapacitated facility location problem. In: MIRCHANDANI, P.; FRANCIS, R. (Eds.) **Discrete location theory**. Wiley-Interscience, New York, 1990

CORTELLA, M. S. **Não nascemos prontos!** Provocações filosóficas. 19 ed. Rio de Janeiro: Ed. Vozes, 2006.

CURRENT, J. R.; SCHILLING, D. A. Elimination of Source A and B Errors in p-Median Location Problems. **Geographical Analysis**, v. 19, n. 2, p. 95–110, 1987.

DASKIN, M. S. **Network and Discrete Location**. 2 ed. Hoboken, New Jersey: John Wiley & Sons, Ltd, 2013.

DASKIN, M. S. et al. Aggregation effects in maximum covering models. **Annals of Operations Research**, v. 18, p. 115–140, 1989.

DEKLE, J.; MARIEL S. LAVIERI, S.; MARTIN, E.; EMIR-FARINAS, H.; FRANCIS, R. A Florida County Locates Disaster Recovery Centers. **INFORMS Journal on Applied Analytics**, v. 35, n. 2, p. 133-139, 2005

DREZNER, Z.; HAMACHER, H. **Facility Location**. Springer-Verlag, Berlin Heidelberg, 2002

ERKUT, E.; BOZKAYA, B. Analysis of aggregation errors for the p-median problem. **Computers and Operations Research**, v. 26, n. 10–11, p. 1075–1096, 1999.

- ESNAF, Ş.; KÜÇÜKDENİZ, T. A fuzzy clustering-based hybrid method for a multi-facility location problem. **Journal of Intelligent Manufacturing**, v. 20, n. 2, p. 259–265, 2009.
- ESTER, M.; KRIEGEL, H.; SANDER, J.; XU, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. **AAAI Press**. P. 226-231, 1996.
- EVERITT, B.; LANDAU, S.; LEESE, M.; STAHL, D. **Cluster Analysis**. Hoboken: Wiley, 2011.
- FARAHANI, R. Z.; HEKMATFAR, M. **Facility location: concepts, models, algorithms, and case studies**. Heidelberg London: Physica Springer distributor, 2009
- FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados**. 1 ed. Rio de Janeiro: Elsevier, 2017.
- FRANCIS, R. L.; LOWE, T. J. Comparative error bound theory for three location models: Continuous demand versus discrete demand. **Top**, v. 22, n. 1, p. 144–169, 2014.
- FRANCIS, R.; LOWE, T.; RAYCO, M.; TAMIR, A. Aggregation error for location models: Survey and analysis. **Annals of Operations Research**, v. 167, n. 1, p. 171–208, 2009.
- GALINDO, G.; BATTÀ, R. Prepositioning of supplies in preparation for a hurricane under potential destruction of prepositioned supplies. **Socio-Economic Planning Sciences**, v. 47, n. 1, p. 20–37, 2013.
- GALVÃO, R.; RAGGI, L. A Method For Solving To Optimality Uncapacitated Location Problems. **Annals of Operations Research**, v. 18, p. 225–244, 1989.
- GALVÃO, L. C.; NOVAES, A. G. N.; SOUSA DE CURSI, J. C.; SOUSA, J.C. A multiplicatively weighted Voronoi diagram approach to logistics districting. **Computers & Operations Research**, v. 33, n. 1, p. 93–114, 1 jan. 2006.
- GAO, X. A location-driven approach for warehouse location problem. **Journal of the Operational Research Society**, v. 72, n. 12, p. 2735–2754, 2021.
- GAREY, M.; JOHNSON, D. **Computers and intractability: a guide to the theory of NP-completeness**. Freeman, 1979
- GEOSPATIAL FOUNDATION PROJECT. QGIS Geographic Information System, 2020.
- GHIANI, G. et al. Capacitated location of collection sites in an urban waste management system. **Waste Management**, v. 32, n. 7, p. 1291–1296, 2012.

GOOGLE. **Geocoding API**, 2022. Disponível em:

<https://developers.google.com/maps/documentation/geocoding?hl=pt-br>. Acesso em 20 abr. 2022

GUAZZELLI, C. S. **Modelos e métodos para estudos de configurações de redes logísticas**.

2018. Tese (Doutorado em Ciências) – Escola Politécnica da Universidade de São Paulo, 2018. DOI 10.11606/T.3.2018.tde-13072018-112347. Disponível em:

<https://www.teses.usp.br/teses/disponiveis/3/3138/tde-13072018-112347/pt-br.php>. Acesso em 07 mar. 2022

GUNER, A. R.; SEVKLI, M. A Discrete Particle Swarm Optimization Algorithm for Uncapacitated Facility Location Problem. **Journal of Artificial Evolution and Applications**, v. 2008, p. 1–9, 2008.

GUROBI OPTIMIZATION, LLC. Gurobi Optimizer Reference Manual, 2022. Disponível em <https://www.gurobi.com>. Acesso em 07 mar. 2022

HAKLI, H.; ORTACAY, Z. An improved scatter search algorithm for the uncapacitated facility location problem. **Computers and Industrial Engineering**, v. 135, n. December 2017, p. 855–867, 2019.

HAKIMI, S. L. Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. **Operations Research**, v. 12, n. 3, p. 450–459, 1964.

HAKIMI, S. L. Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems. **Operations Research**, v. 13, n. 3, p. 462–475, 1965.

HAN, J.; KAMBER, M.; PEI, J. (Eds.). **Data Mining**. 3 ed. Boston: Morgan Kaufmann, 2012. p. 443–495.

HANSEN, P. et al. Primal-dual variable neighborhood search for the simple plant-location problem. **INFORMS Journal on Computing**, v. 19, n. 4, p. 552–564, 2007

HILLSMAN, E. L.; RHODA, R. Errors in measuring distances from populations to service centers. **The Annals of Regional Science**, v. 12, n. 3, p. 74–88, 1978.

HODGSON, M. J.; NEUMAN, S. A GIS approach to eliminating source C aggregation error in p-median models. **Location Science**, v. 1, n. 2, p. 155–170, 1993.

- HODGSON, M. J.; SHMULEVITZ, F.; KÖRKEL, M. Aggregation error effects on the discrete-space p-median model: The case of Edmonton, Canada. **Canadian Geographer**, v. 41, n. 4, p. 415–428, 1997.
- HOFER, K.; FLUCHER, S.; FELLENDORF, M.; SCHADLER, M.; HAFNER, N. Estimation of Changes in Customer's Mobility Behaviour by the Use of Parcel Lockers. *Transportation Research Procedia*, 47, 425–432, 2020.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Censo demográfico**. Características da população e dos domicílios. Resultado do universo. Rio de Janeiro, 2011. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/93/cd_2010_caracteristicas_populacao_domicilios.pdf . Acesso em 22 de setembro de 2021.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Grade estatística**. Rio de Janeiro, 2016.
- IRAWAN, C. A.; SALHI, S.; LUIS, M.; AZIZI, N. The continuous single source location problem with capacity and zone-dependent fixed cost: Models and solution approaches. **European Journal of Operational Research**, v. 263, n. 1, p. 94–107, 2017.
- IRAWAN, C. A.; SALHI, S. Aggregation and non aggregation techniques for large facility location problems-a survey. **Yugoslav Journal of Operations Research**, v. 25, n. 3, p. 313–341, 2015a.
- IRAWAN, C. A.; SALHI, S. Solving large p-median problems by a multistage hybrid approach using demand points aggregation and variable neighbourhood search. **Journal of Global Optimization**, v. 63, n. 3, p. 537–554, 2015b.
- KASHAN, M. H.; NAHAVANDI, N.; KASHAN, A. H. DisABC: A new artificial bee colony algorithm for binary optimization. **Applied Soft Computing Journal**, v. 12, n. 1, p. 342–352, 2012.
- KIN, B.; VERLINDE, S.; MACHARIS, C. Sustainable urban freight transport in megacities in emerging urban markets. **Sustainable Cities and Society**, v. 32, p. 31–41, 2017
- KHUMAWALA, B. An Efficient Branch and Bound Algorithm for the Warehouse Location Problem. **Management Science**, v. 18, n. 12, p.718–731, 1972
- KRARUP, J.; PRUZAN, P. M. The simple plant location problem: Survey and synthesis. **European Journal of Operational Research**, 1983.

KRATICA, J.; DUGOŠIJA, D.; SAVIĆ, A. A new mixed integer linear programming model for the multi level uncapacitated facility location problem. **Applied Mathematical Modelling**, v. 38, n. 7–8, p. 2118–2129, 2014.

KRATICA, J.; TOŠIĆ, D.; FILIPOVIĆ, V.; LJUBIĆ, I. Solving the simple plant location problem by genetic algorithm. **RAIRO Operations Research**, v. 35, p. 127–142, 2001.

LAPORTE, G.; NICKEL, S.; SALDANHA DA GAMA, F. S. **Location science**. Springer, 2015.

LEVIN, Y.; BEN-ISRAEL, A. A heuristic method for large-scale multi-facility location problems. **Computers and Operations Research**, v. 31, n. 2, p. 257–272, 2004.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. **Anais do the fifth Berkeley symposium on mathematical statistics and probability**, v. 1, n. 14, p. 281-297

MARCHAND, H. et al. Cutting planes in integer and mixed integer programming. **Discrete Applied Mathematics**, v. 123, n. 1–3, p. 397–446, 2002.

MARQUES, M. DO C.; DIAS, J. M. Simple dynamic location problem with uncertainty: A primal-dual heuristic approach. **Optimization**, v. 62, n. 10, p. 1379–1397, 2013.

MELO, M. T.; NICKEL, S.; SALDANHA DA GAMA, F. S. Dynamic multi-commodity capacitated facility location: A mathematical modeling framework for strategic supply chain planning. **Computers and Operations Research**, v. 33, n. 1, p. 181–208, 2006.

MELO, M. T.; NICKEL, S.; SALDANHA-DA-GAMA, F. Facility location and supply chain management - A review. **European Journal of Operational Research**, v. 196, n. 2, p. 401-412, 2009.

NOVAES, A. G. N.; SOUZA DE CURSI, J. E.; SILVA, A. C. L.; SOUZA, J. C. Solving continuous location–districting problems with Voronoi diagrams. **Computers & Operations Research**, v. 36, n. 1, p. 40–59, 1 jan. 2009.

OKABE, A.; BOOTS, B.; SUGUIHARA, K.; CHIU, S. N. **Spatial Tessellations: of Voronoi Diagrams Concepts and Applications**. Chichester, New York: Wiley, 2000

OKABE, A.; SUZUKI, A. Locational optimization problems solved through Voronoi diagrams. **European Journal of Operational Research**, v. 98, n. 3, p. 445–456, 1 maio 1997.

O'KELLY, M. E. A CLUSTERING APPROACH TO THE PLANAR HUB LOCATION PROBLEM. **Annals of Operations Research**, v. 40, p. 339–353, 1992.

PAUL, J. A.; MACDONALD, L. Location and capacity allocations decisions to mitigate the impacts of unexpected disasters. **European Journal of Operational Research**, v. 251, n. 1, p. 252–263, 2016.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, É. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, 12, 2825–2830, 2011.

PEETERS, D.; THOMAS, I. Distance predicting functions and applied location-allocation models. **Journal of Geographical Systems**, v. 2, n. 2, p. 167–184, 2000.

RAHMAH, N.; SITANGGANG, I. S. Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. **IOP Conference Series: Earth and Environmental Science**, v. 31, n. 1, 2016.

RAYCO, M. B.; FRANCIS, R. L.; LOWE, T. J. Error-bound driven demand point aggregation for the rectilinear distance p-center model. **Location Science**, v. 4, n. 4, p. 213–235, 1 dez. 1996

REVELLE, C. S.; EISELT, H. A.; DASKIN, M. S. A bibliography for some fundamental problem categories in discrete location science. **European Journal of Operational Research**, v. 184, n. 3, p. 817–848, 2008.

REVELLE, C. S.; SWAIN, R. W. Central facilities location. **Geographical Analysis**, p. 30–42, 1970

RODERO, R. **Gostaria de saber qual o tamanho mínimo de uma drogaria por partes**. São Paulo, 2018. Disponível em: <https://guiadafarmacia.com.br/perguntas/qual-o-tamanho-minimo-de-uma-drogaria-por-partes/>. Acesso em 31 março 2022.

SANKARAN, J. K. On solving large instances of the capacitated facility location problem. **European Journal of Operational Research**, v. 178, n. 3, p. 663–676, 2007.

SUGAR, C.; JAMES, G. Finding the number of clusters in a dataset. **Journal of the American Statistical Association**, v. 98, n. 436, p. 750–763, 2003.

SUN, M. Solving the uncapacitated facility location problem using tabu search. **Computers and Operations Research**, v. 33, n. 9, p. 2563–2589, 2006.

TOREGAS, C. et al. The Location of Emergency Service Facilities. **Operations Research**, v. 19, n. 6, p. 1363–1373, 1971.

VAN ROY, T. A Cross Decomposition Algorithm for Capacitated Facility Location. **Operations Research**, v. 34, n. 1, p.145-163, 1986.

WOLSEY, L. A. **Integer Programming**. New York: John Wiley & Sons, Ltd, 1998.

YIGIT, V.; AYDIN, M. E.; TURKBEY, O. Solving large-scale uncapacitated facility location problems with evolutionary simulated annealing. **International Journal of Production Research**, v. 44, n. 22, p. 4773–4791, 2006.

ZHAO, P.; BATTÀ, R. Analysis of centroid aggregation for the Euclidean distance p-median problem. **European Journal of Operational Research**, v. 113, n. 1, p. 147–168, 16 fev. 1999.

APÊNDICE A – DIAGRAMAS DE VORONOI

Segundo Okabe *et al.* (2000), o diagrama de Voronoi é um constructo geométrico, cuja origem é associada ao matemático alemão Gauss, ainda no século 19. Todavia, foi o matemático Georges Voronoi, no começo do século 20, quem formalizou o conceito de partição e associação do espaço ao ponto existente mais próximo. Posteriormente, outros estudiosos como Thiessen e Delauney também deixaram suas contribuições, de maneira que outros nomes tais como polígonos de Thiessen, ou tesselações de Delauney são utilizados para apresentar algum caso especial de diagramas de Voronoi.

O diagrama de Voronoi é um diagrama geométrico que possibilita aplicações em diversas áreas. Seja um espaço d -dimensional, tal que $2 \leq d \leq \infty$, no qual se localizam diversos pontos. A cada ponto, é designada uma porção do espaço para a qual este ponto em questão é o mais próximo no espaço. Para cada subdivisão do plano em regiões, dá-se o nome de polígonos de Voronoi, sendo que o conjunto de polígonos de Voronoi é denominado diagrama de Voronoi (BURKEY; BHADURY; EISELT, 2011). Ressalta-se, todavia, que para que sejam obtidos polígonos de Voronoi, é necessário existir ao menos dois pontos, do contrário, o único ponto existente é considerado o mais próximo de todo o espaço em questão. Um exemplo do diagrama de Voronoi para uma área irregular é apresentado na Figura A.1.

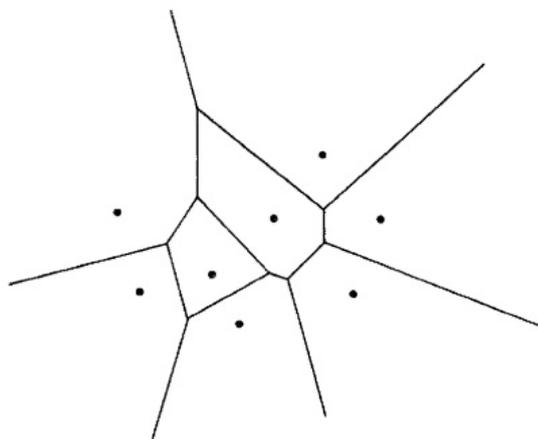


Figura A.1 – Diagrama de Voronoi para oito pontos em um plano. Fonte: Aurenhammer (1991)

A obtenção do diagrama de Voronoi mais simples para um conjunto de pontos utiliza a distância euclidiana para calcular a distância entre as regiões do espaço e o respectivo ponto. Ressalta-se, porém, que é possível utilizar outras medidas senão a euclidiana para o cálculo das distâncias, e este uso pode implicar em uma partição espacial diferente para um mesmo conjunto de pontos no espaço, como apresentado pela Figura A.2.

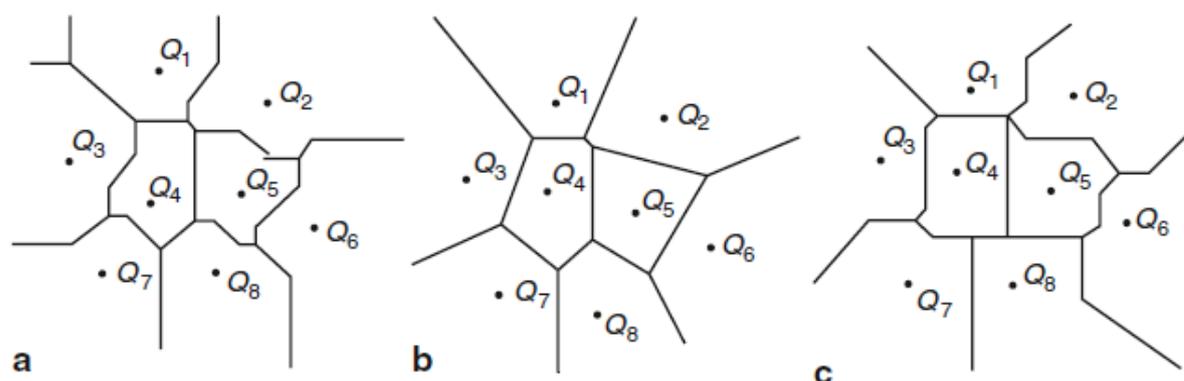


Figura A.2 – Diagramas de Voronoi obtidos utilizando (a) distância Manhattan, (b) distância euclidiana e (c) distância de Chebychev. Fonte: Burkey, Bhadury e Eiselt (2011)

Para aprofundar-se mais em questões teóricas e de modelagem dos polígonos de Voronoi, o leitor é direcionado a Okabe *et al.* (2000) e Aurenhammer (1991). Para conhecimento de aplicações, sugere-se a leitura dos artigos de Galvão *et al.* (2006), Novaes *et al.* (2009) e Okabe e Suzuki (1997), e do estudo de caso disponível em Burkey, Bhadury e Eiselt (2011).

APÊNDICE B – ALOCAÇÃO DE DEMANDA

Na seção 5.3 foram analisadas as instâncias 1-9 e 1-900 do Conjunto 1 e 2-9 e 2-4500 do Conjunto 2 que apresentaram alguma alocação incorreta de demanda e quais são as diferenças de custo de transporte decorrentes deste erro. Como complemento a essas e as demais instâncias nas quais os modelos agregados apresentam valor da FO diferente do modelo desagregado e que não foram apresentadas no Capítulo 5, neste Apêndice são mostradas as alocações da demanda para todos os pontos de demanda e candidatos abertos.

Nas figuras que mostram as alocações de demanda, os locais candidatos abertos são representados por triângulos laranja, os pontos de demanda por pontos azuis e os fluxos de origem-destino entre candidato aberto e ponto de demanda são indicados por linhas pretas.

A Tabela B.1 auxilia a navegação das figuras de alocação da demanda para as instâncias e modelos apresentados. Para o caso de instâncias nas quais as soluções dos modelos desagregado e agregados for a mesma, apenas uma figura é indicada para evitar que uma mesma figura seja repetida duas vezes em sequência.

Tabela B.1 – Figuras que mostram a alocação da demanda

| | Instância | Desagregado | Modelo Ma1 | Modelo Ma2 |
|------------|------------|-------------|-------------|-------------|
| Conjunto 1 | 1-9 | Figura B.1 | Figura B.2 | Figura B.3 |
| | 1-90 | Figura B.4 | Figura B.5 | Figura B.6 |
| | 1-900 | Figura B.7 | Figura B.8 | Figura B.9 |
| | 1-1800 | Figura B.10 | Figura B.11 | Figura B.10 |
| | 1-4500 | | Figura B.12 | |
| | 1-9000 | | Figura B.13 | |
| | 1-18000 | | Figura B.14 | |
| | 1-45000 | | Figura B.15 | |
| | 1-90000 | | Figura B.16 | |
| | Conjunto 2 | 2-9 | Figura B.17 | Figura B.18 |
| 2-90 | | Figura B.20 | Figura B.21 | Figura B.20 |
| 2-900 | | | Figura B.22 | |
| 2-1800 | | | Figura B.23 | |
| 2-4500 | | Figura B.24 | Figura B.25 | Figura B.26 |
| 2-9000 | | Figura B.27 | Figura B.28 | Figura B.27 |
| 2-18000 | | | Figura B.29 | |
| 2-45000 | | | Figura B.30 | |
| 2-90000 | | | Figura B.31 | |

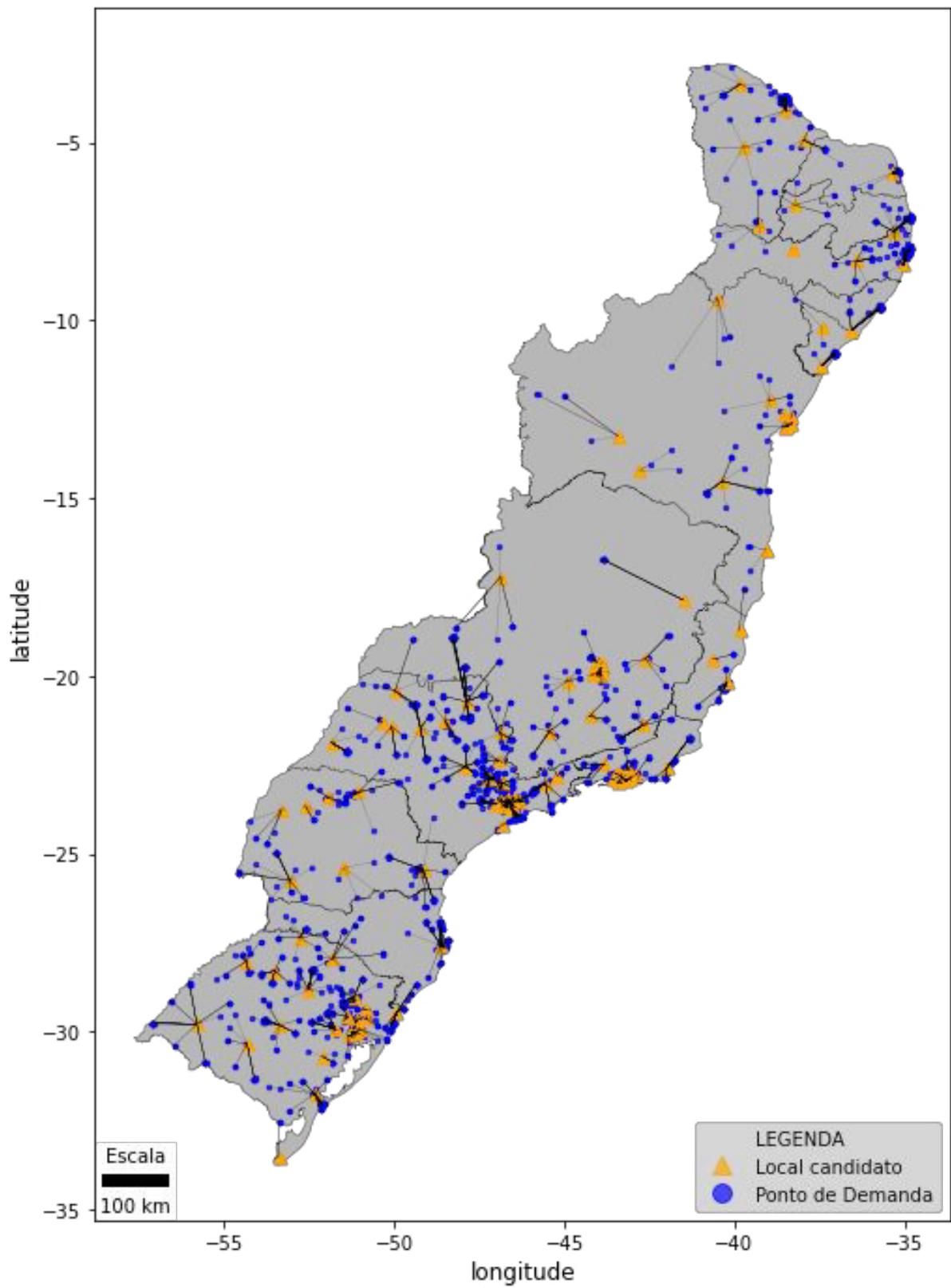


Figura B.1 – Alocação da demanda para o modelo desagregado da instância 1-9

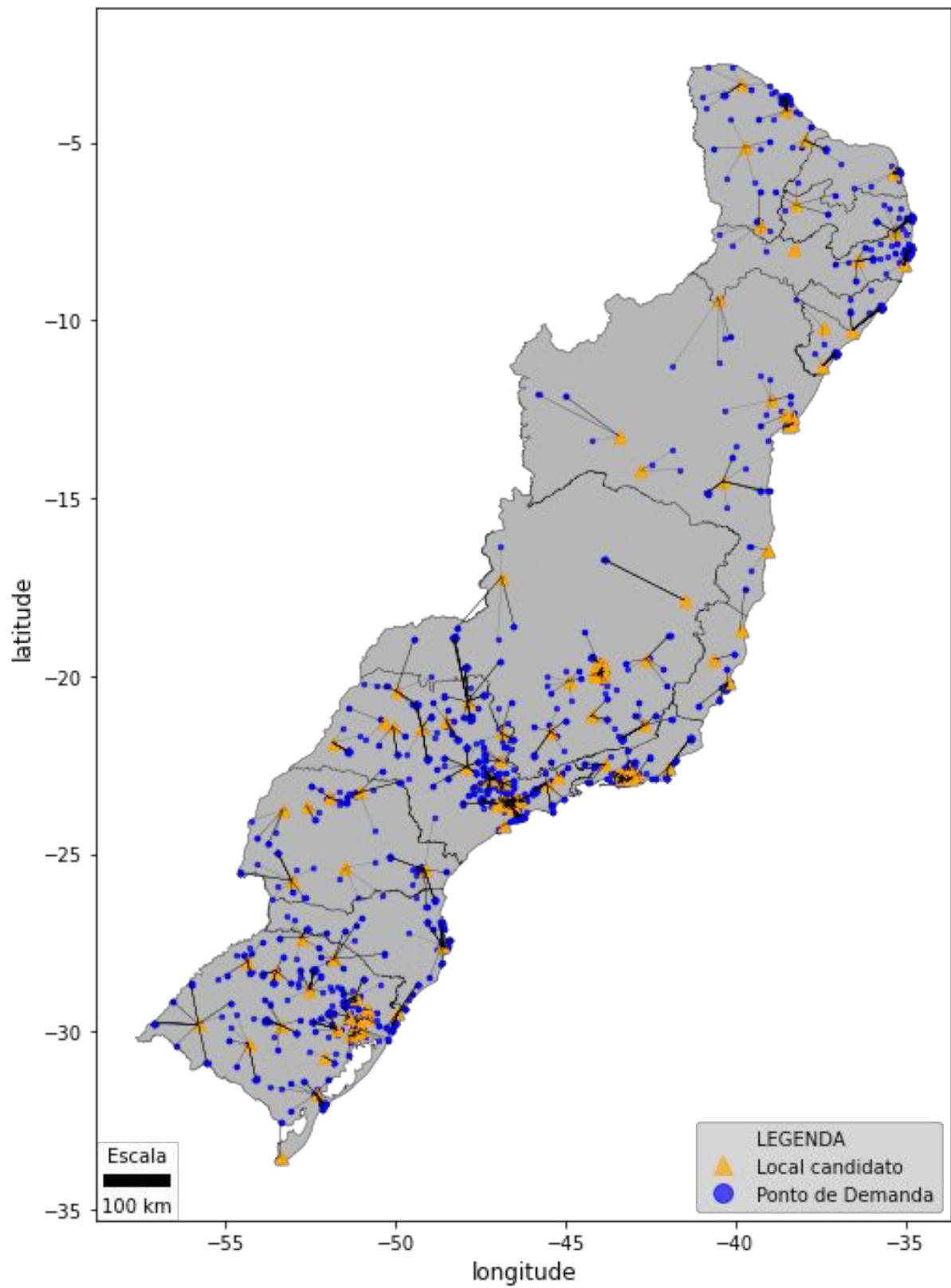


Figura B.2 – Alocação da demanda para o modelo Ma1 da instância 1-9

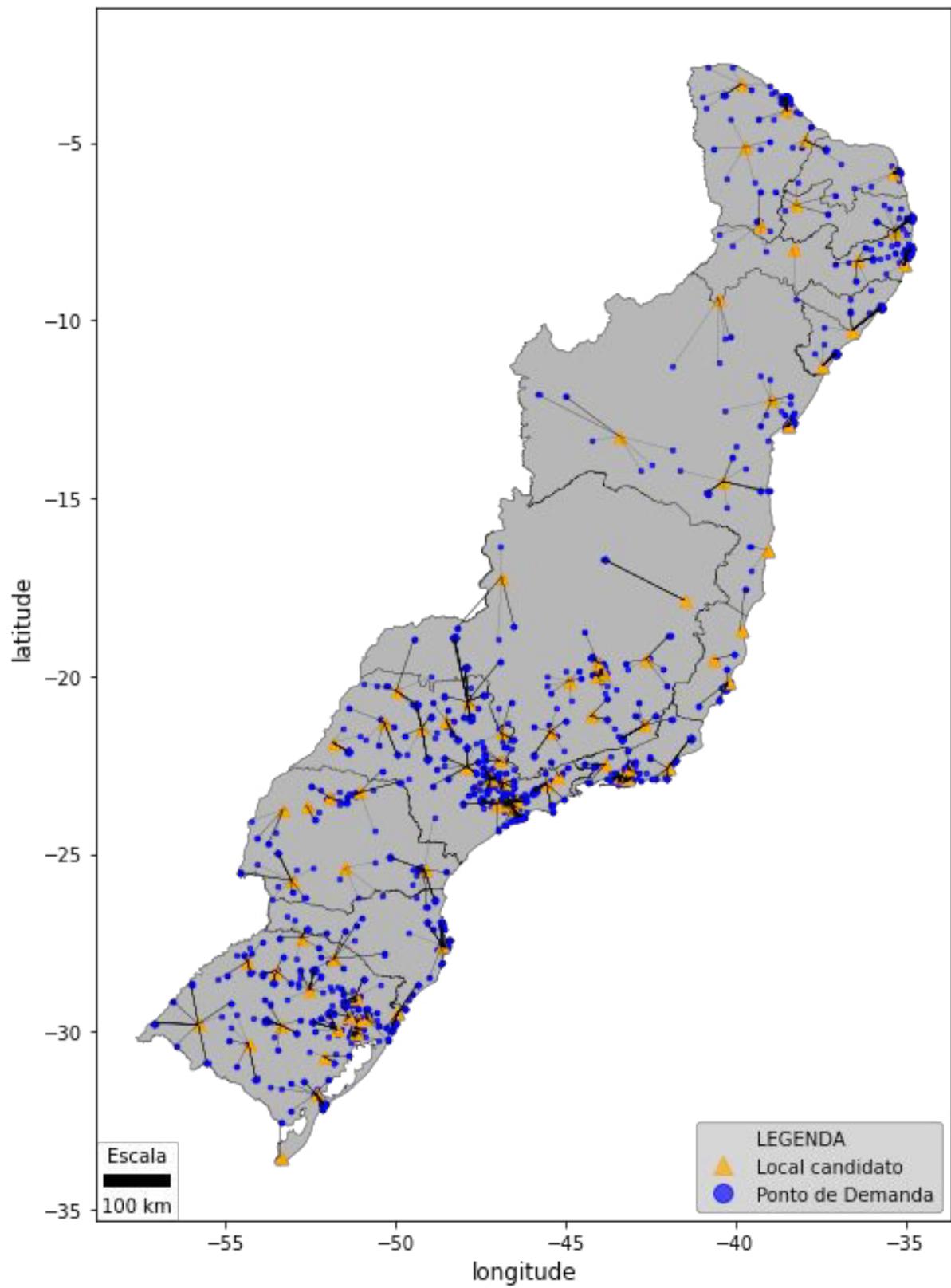


Figura B.3 – Alocação da demanda para o modelo Ma2 da instância 1-9

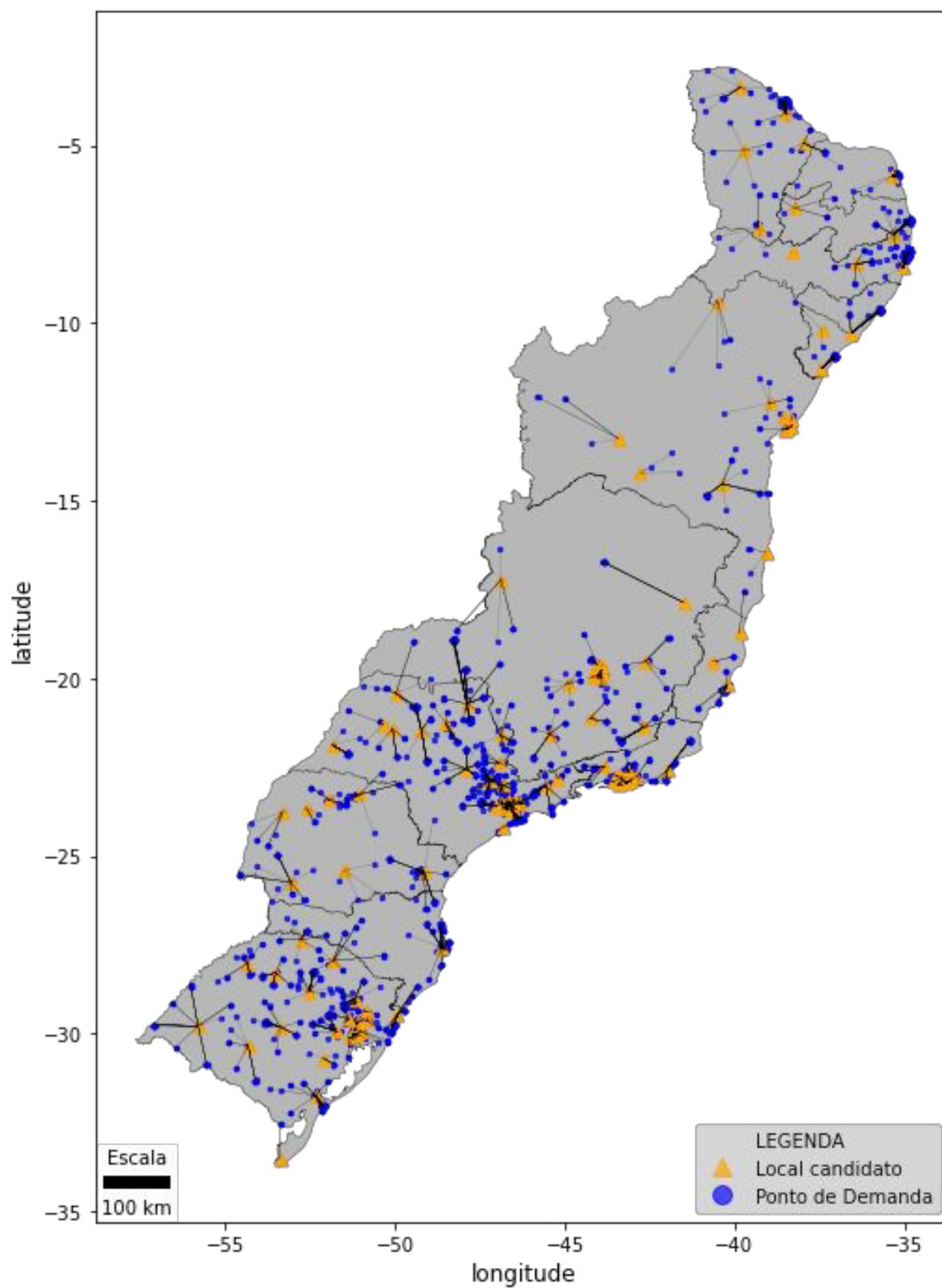


Figura B.4 – Alocação da demanda para o modelo desagregado da instância 1-90

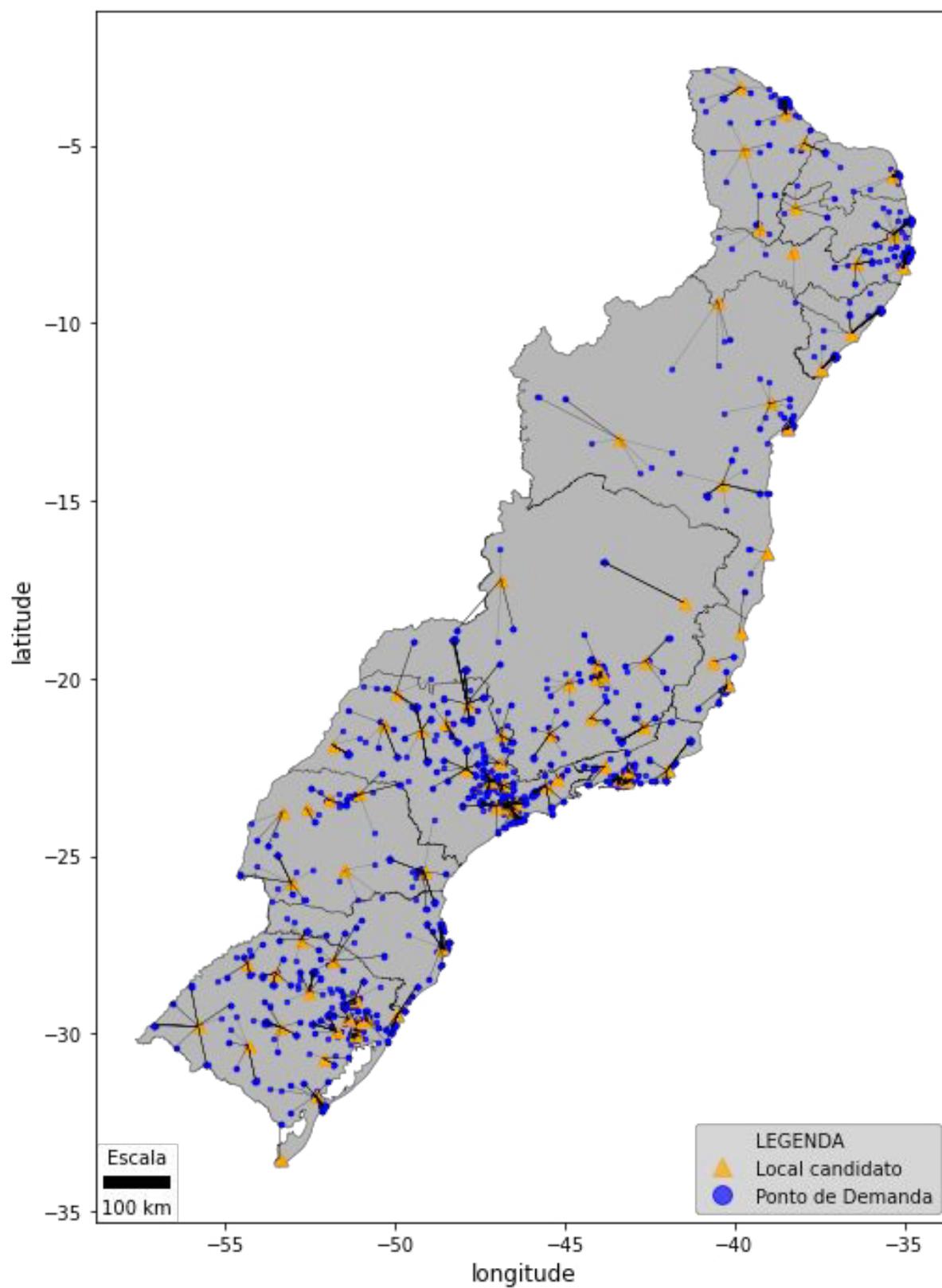


Figura B.5 – Alocação da demanda para o modelo Ma1 da instância 1-90

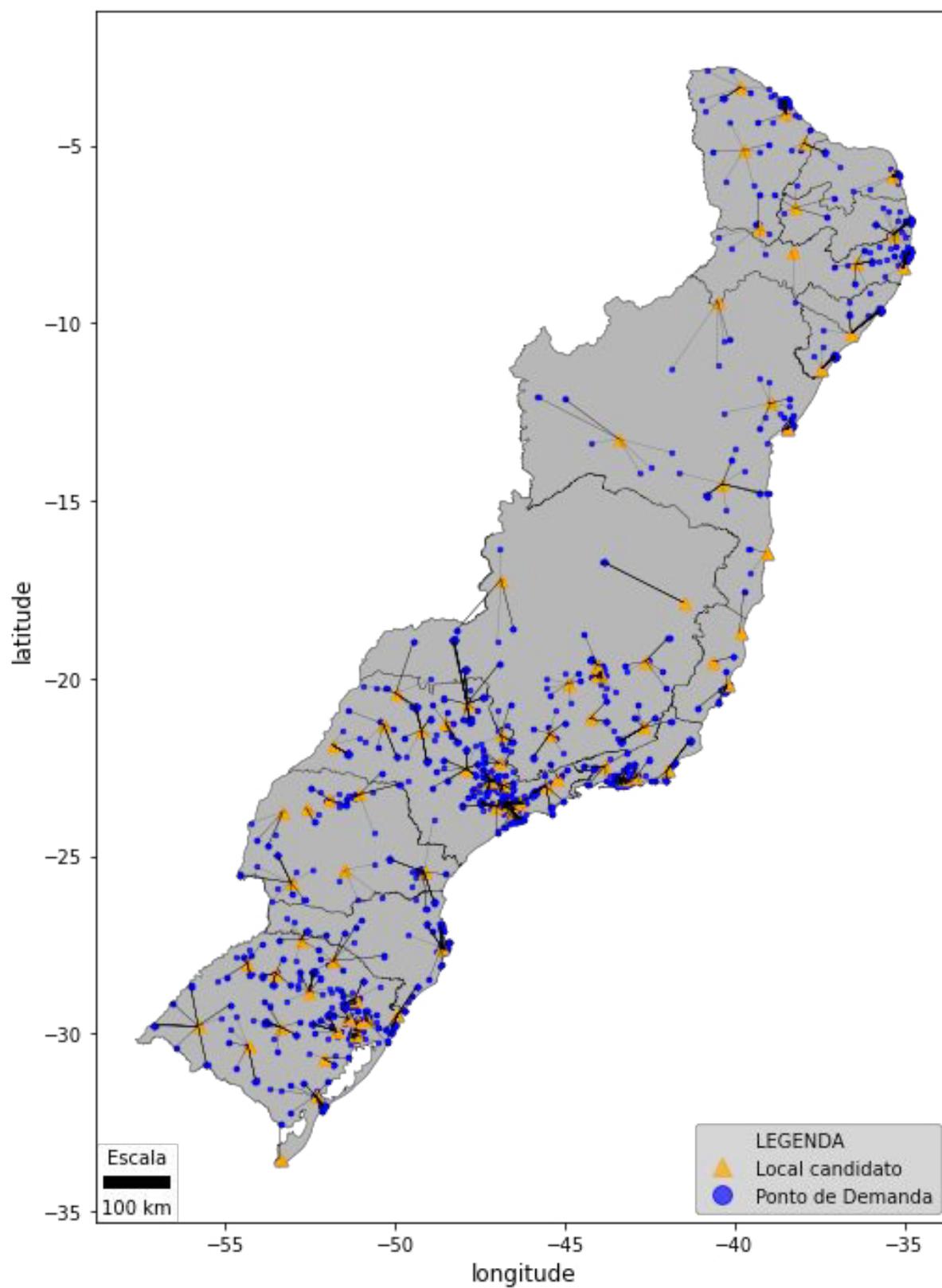


Figura B.6 – Alocação da demanda para o modelo Ma2 da instância 1-90

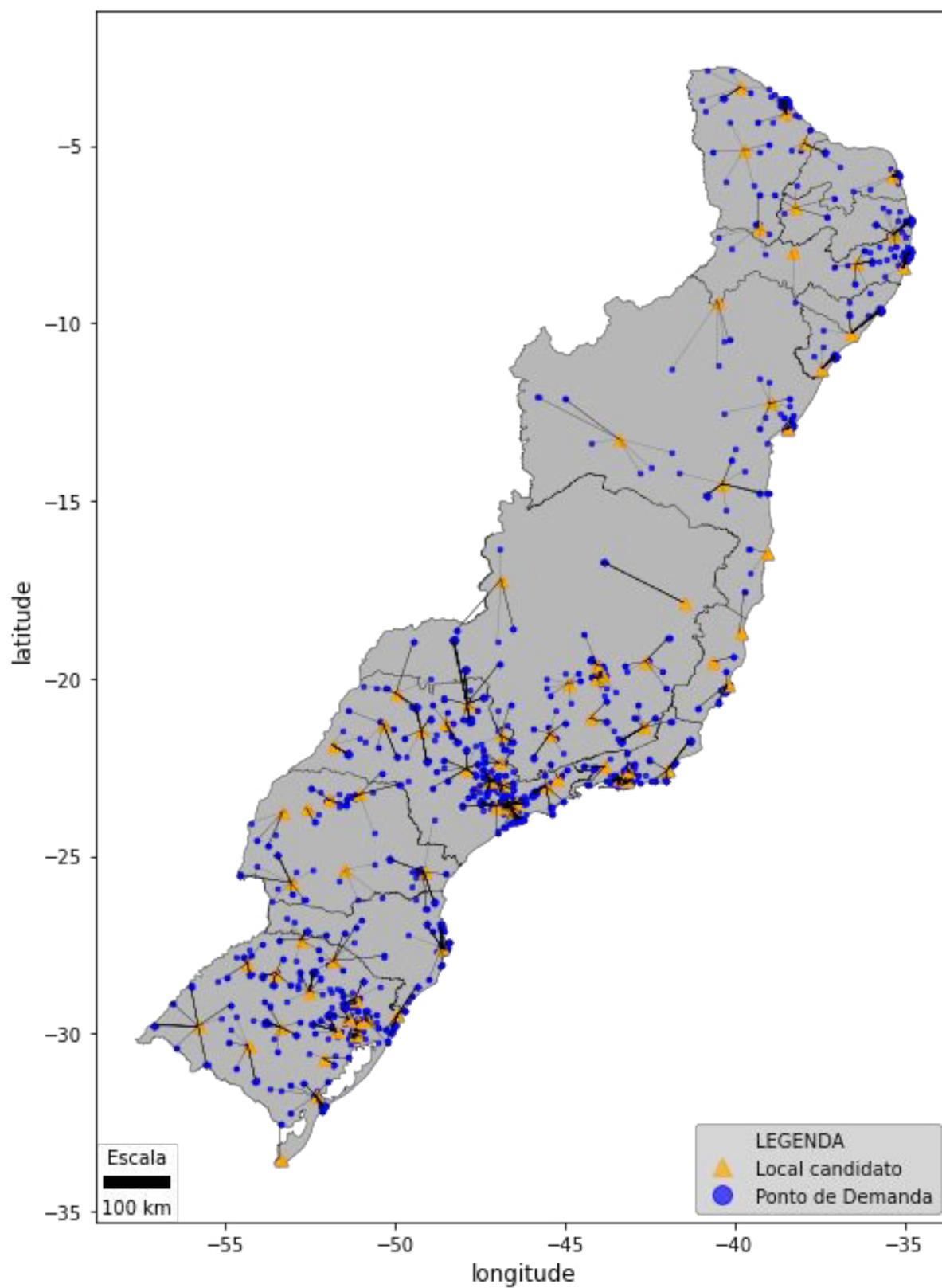


Figura B.7 – Alocação da demanda para o modelo desagregado da instância 1-900

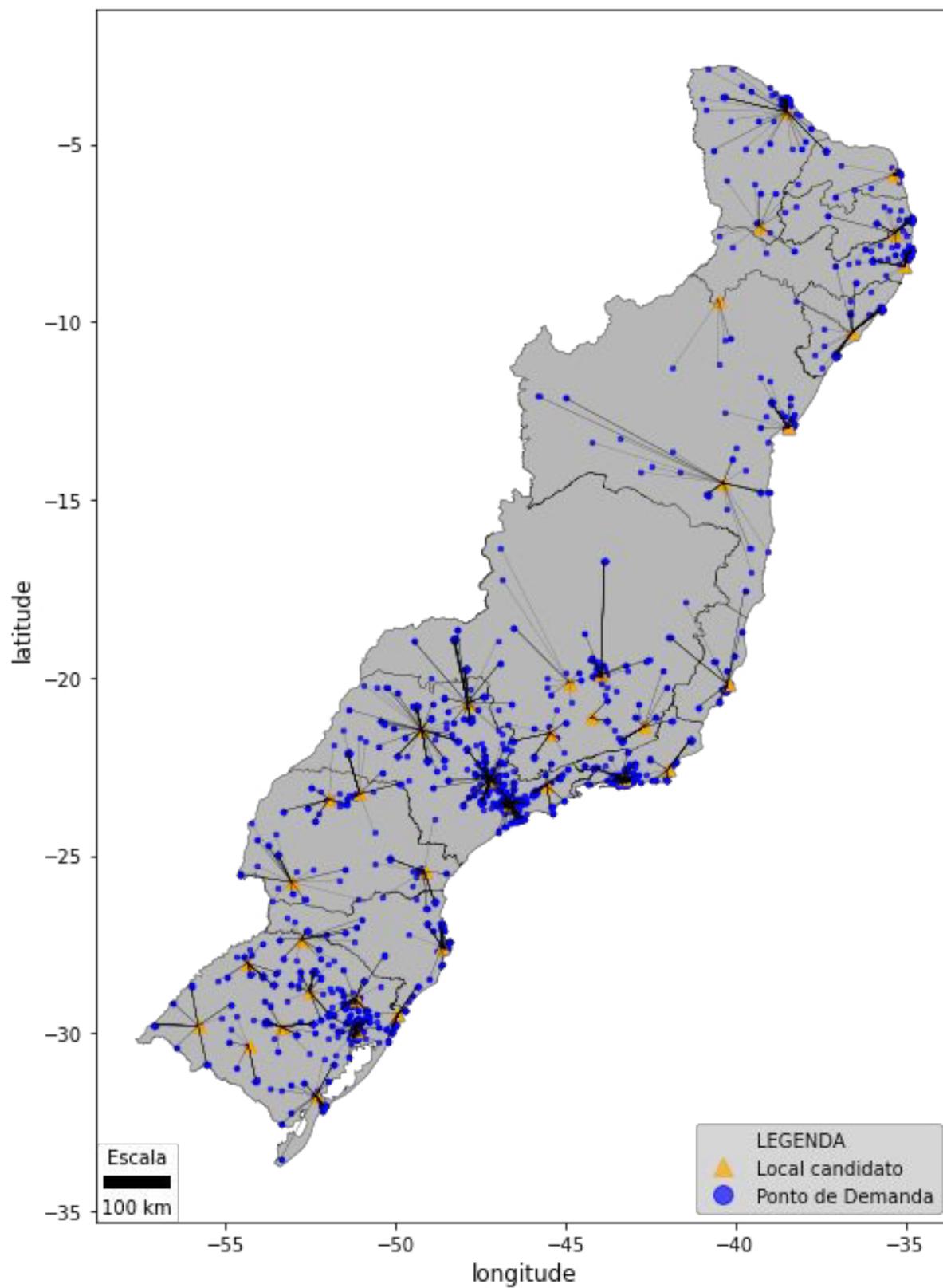


Figura B.8 – Alocação da demanda para o modelo Ma1 da instância 1-900

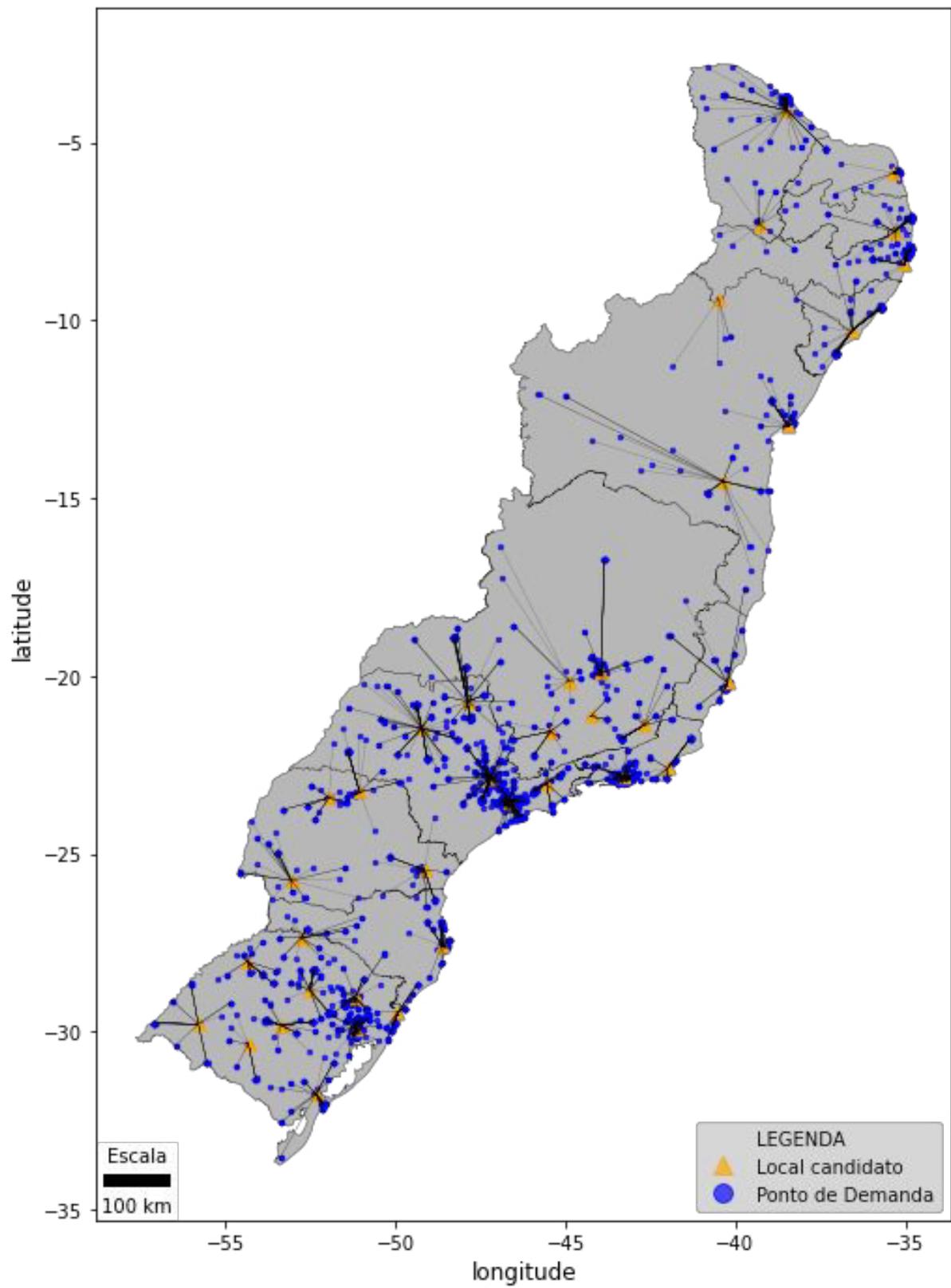


Figura B.9 – Alocação da demanda para o modelo Ma2 da instância 1-900

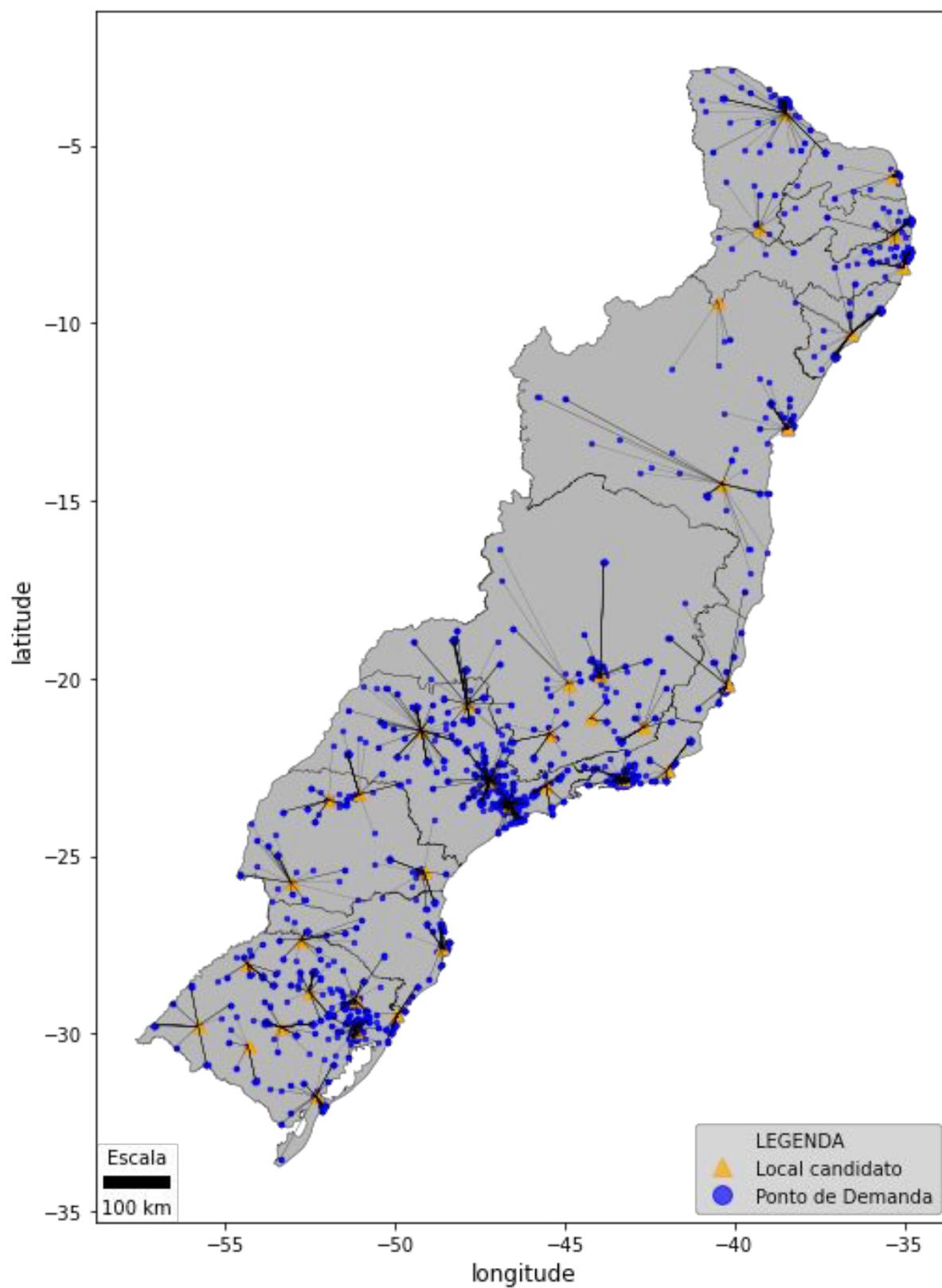


Figura B.10 – Alocação da demanda para os modelos desagregado e Ma2 da instância 1-1800

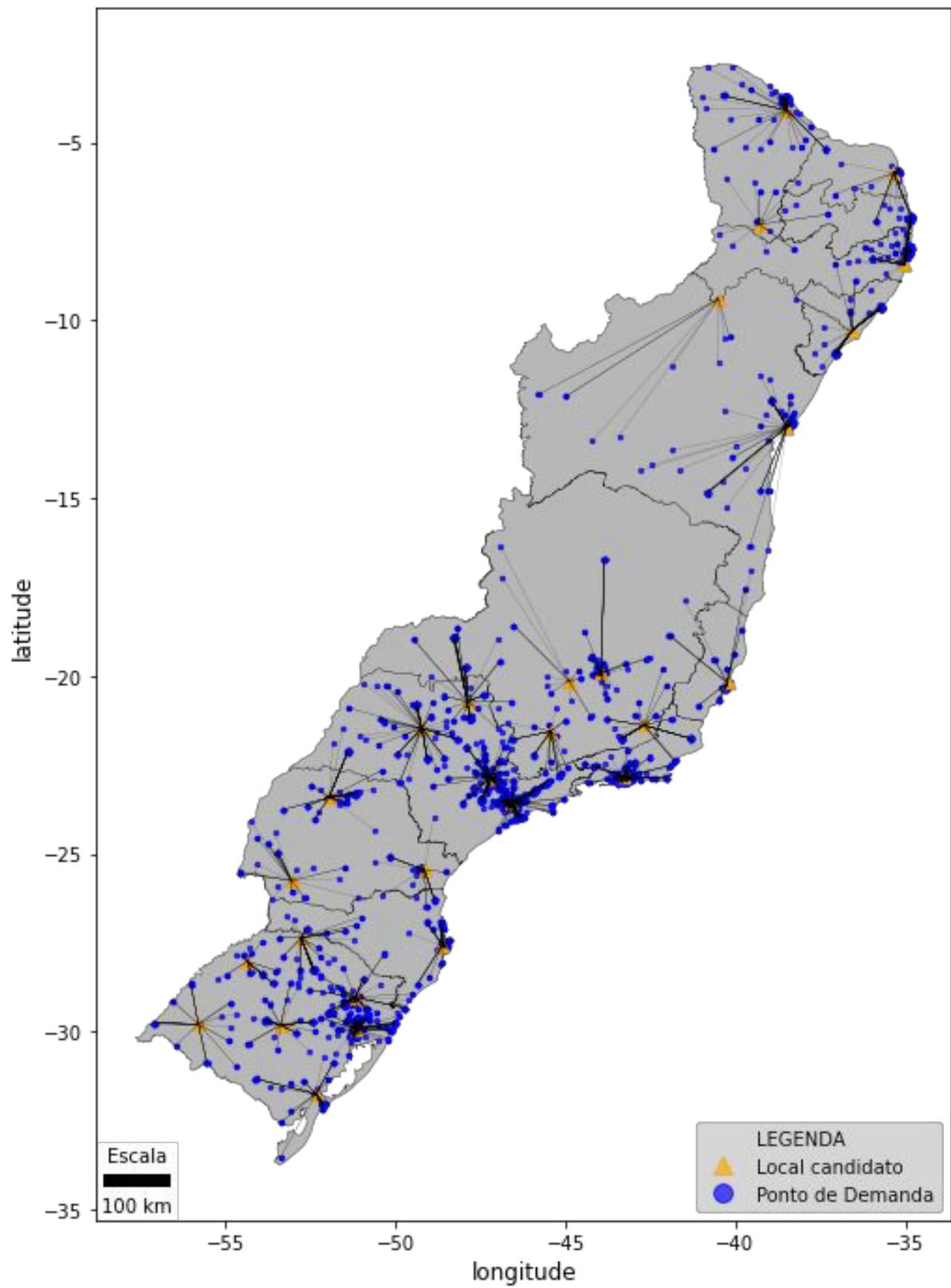


Figura B.11 – Alocação da demanda para o modelo Ma1 da instância 1-1800

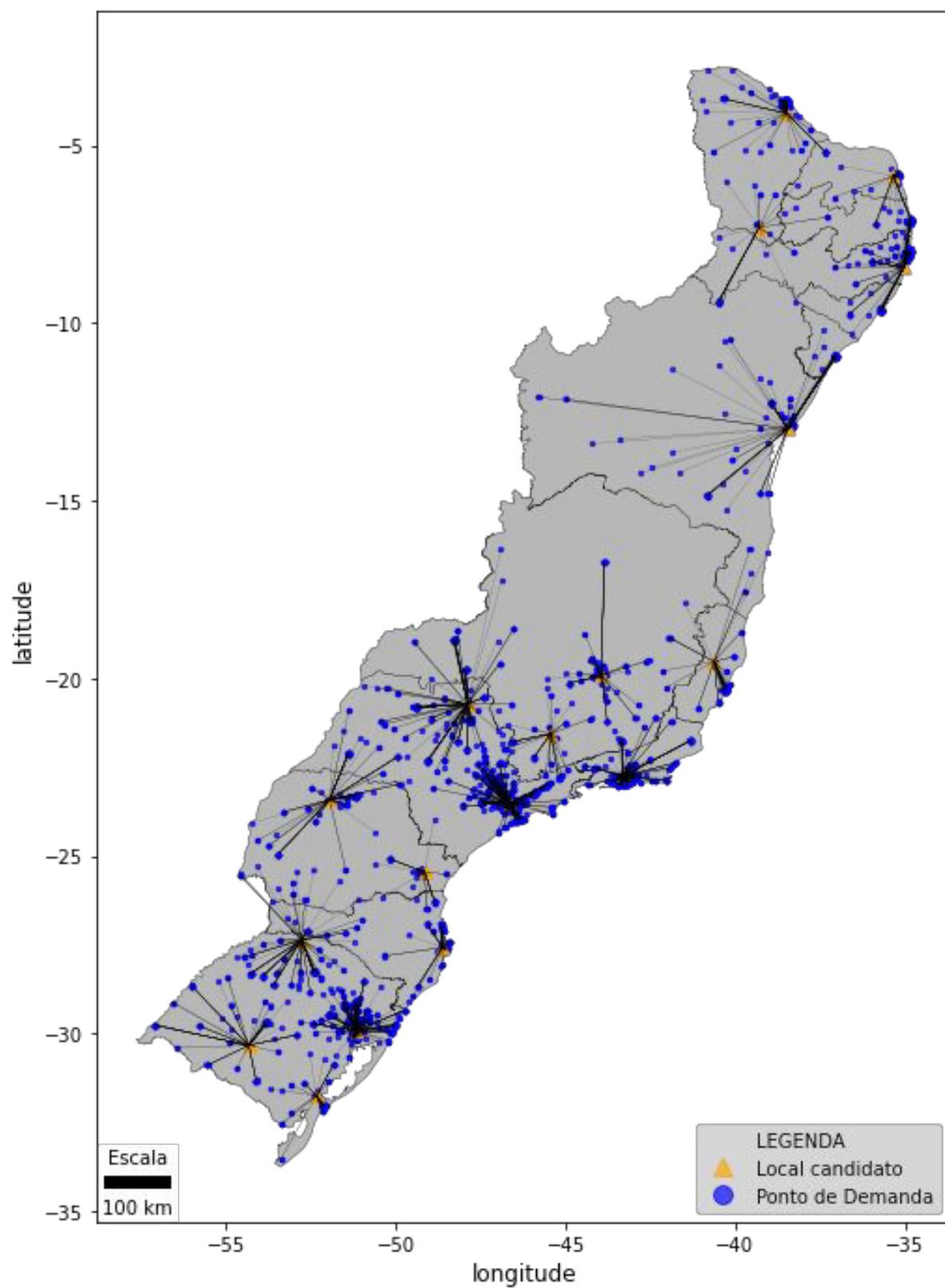


Figura B.12 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-4500

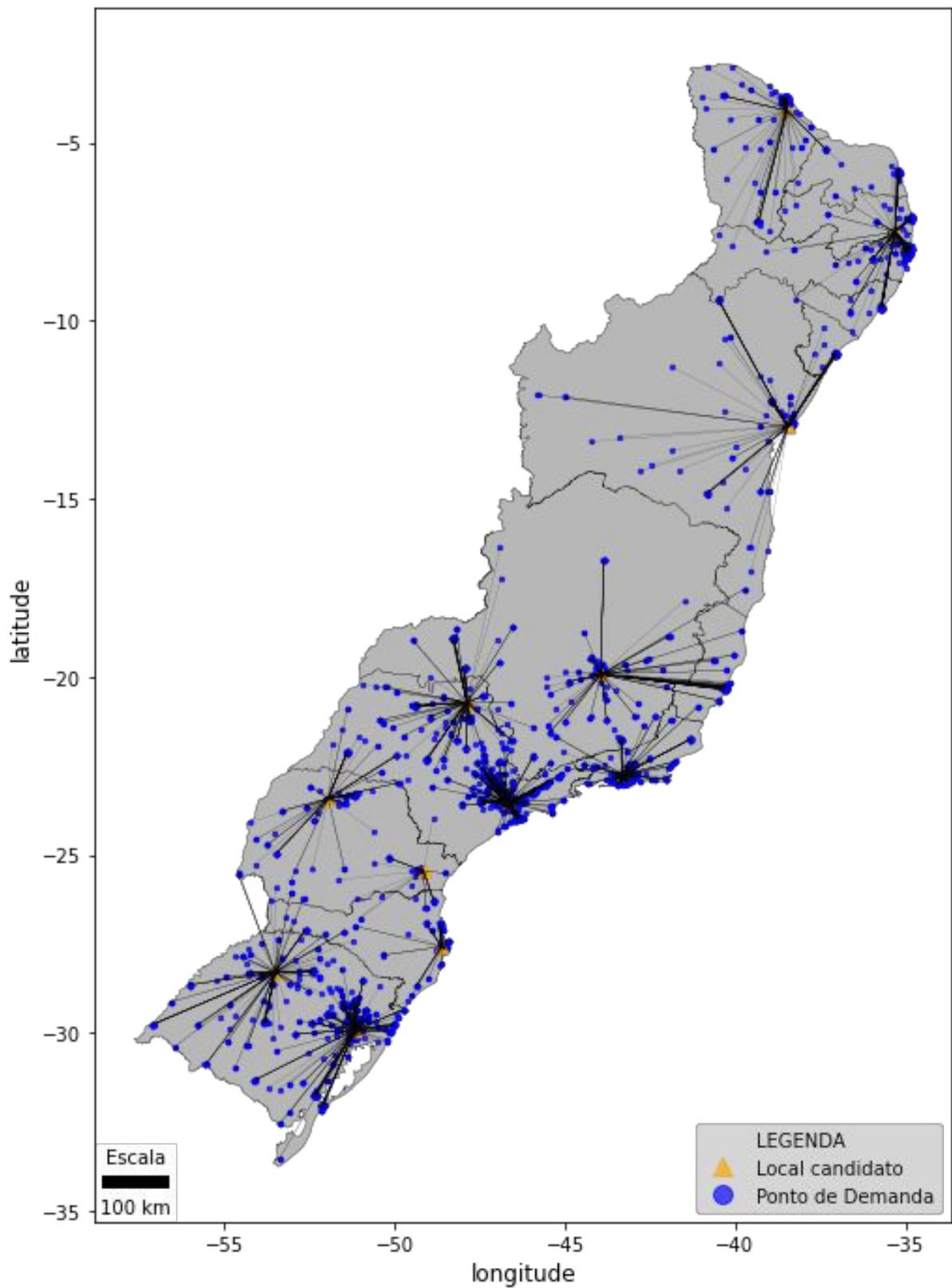


Figura B.13 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-9000

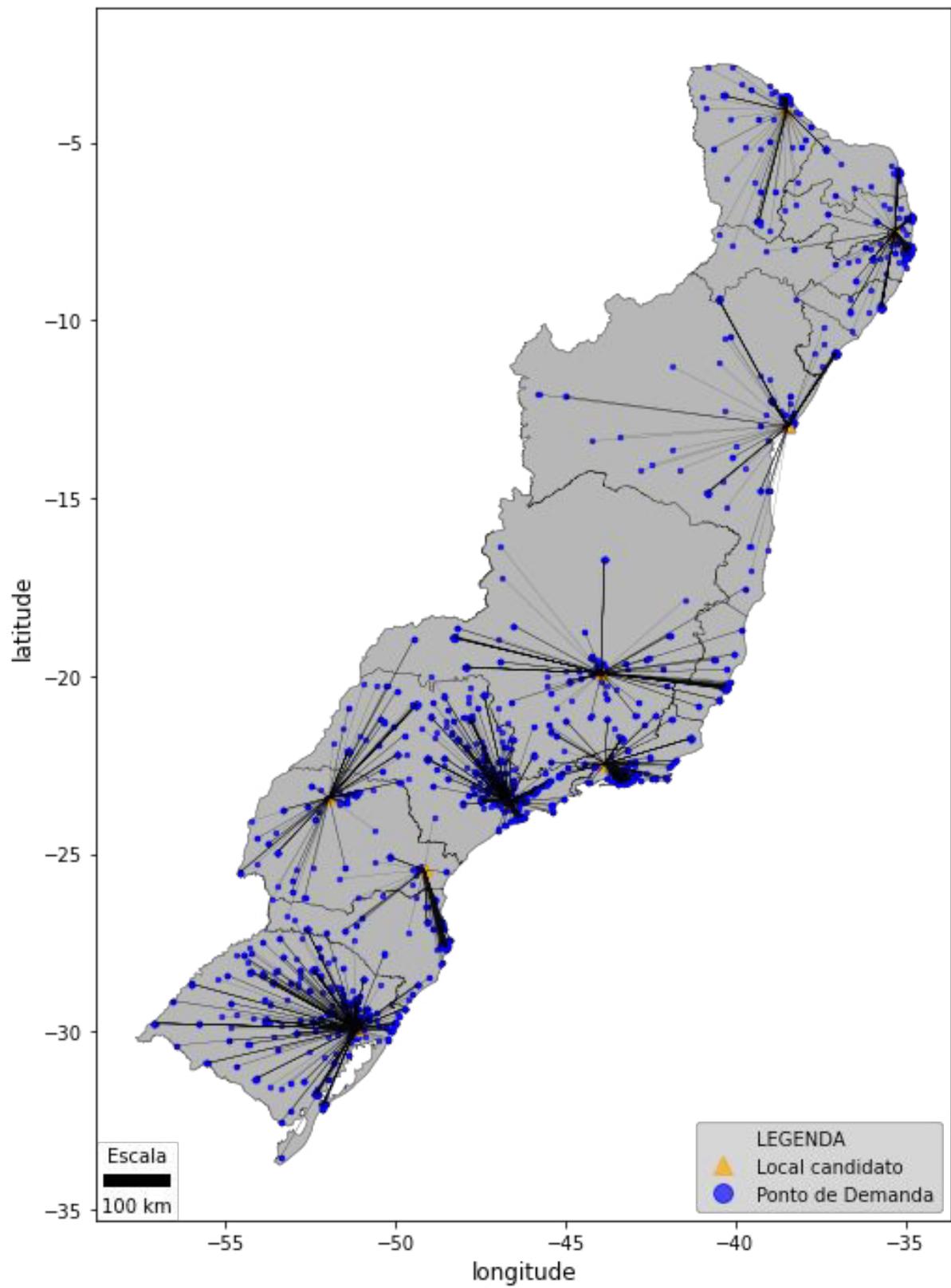


Figura B.14 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-18000

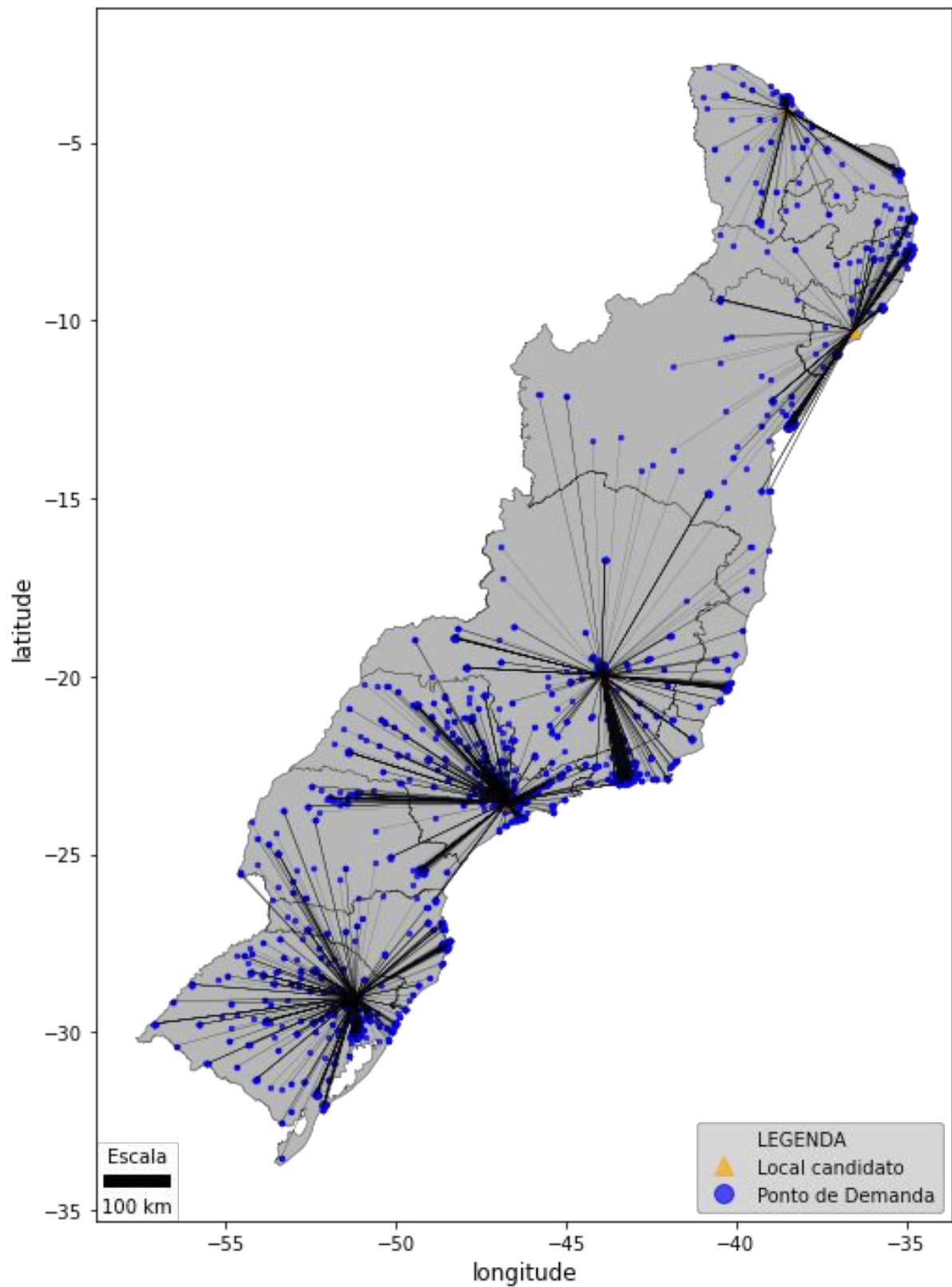


Figura B.15 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-45000

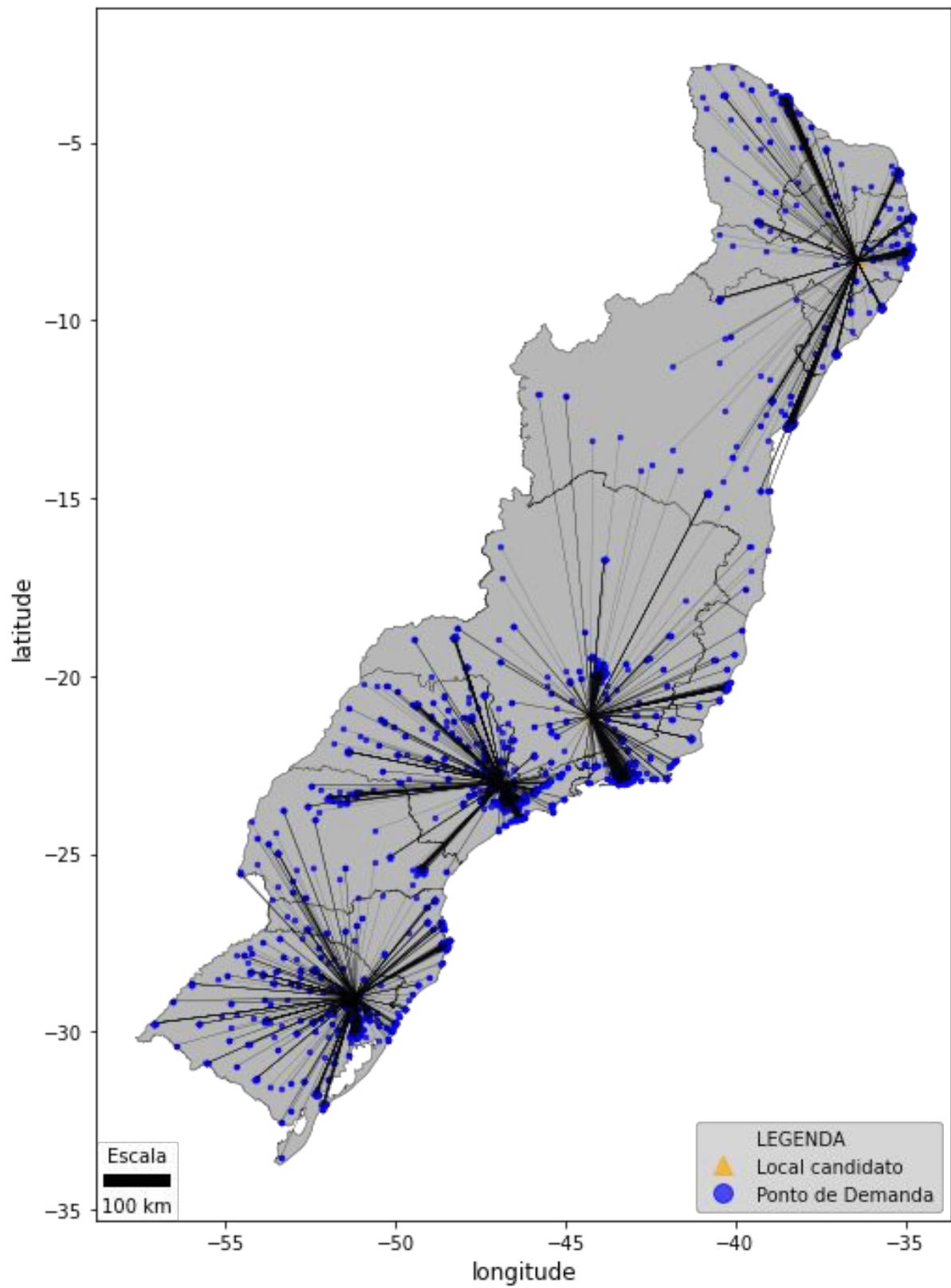


Figura B.16 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 1-90000

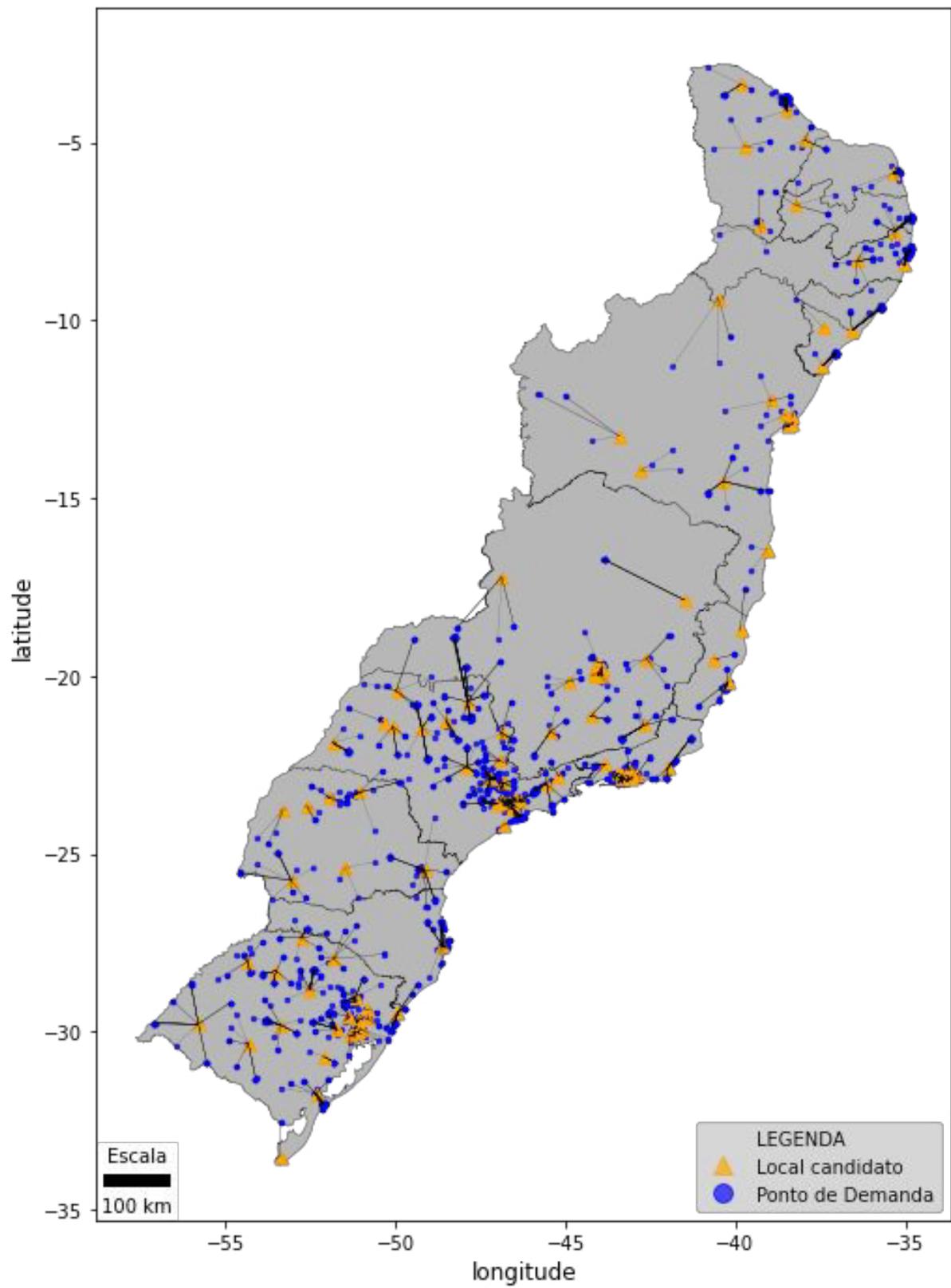


Figura B.17 – Alocação da demanda para o modelo desagregado da instância 2-9

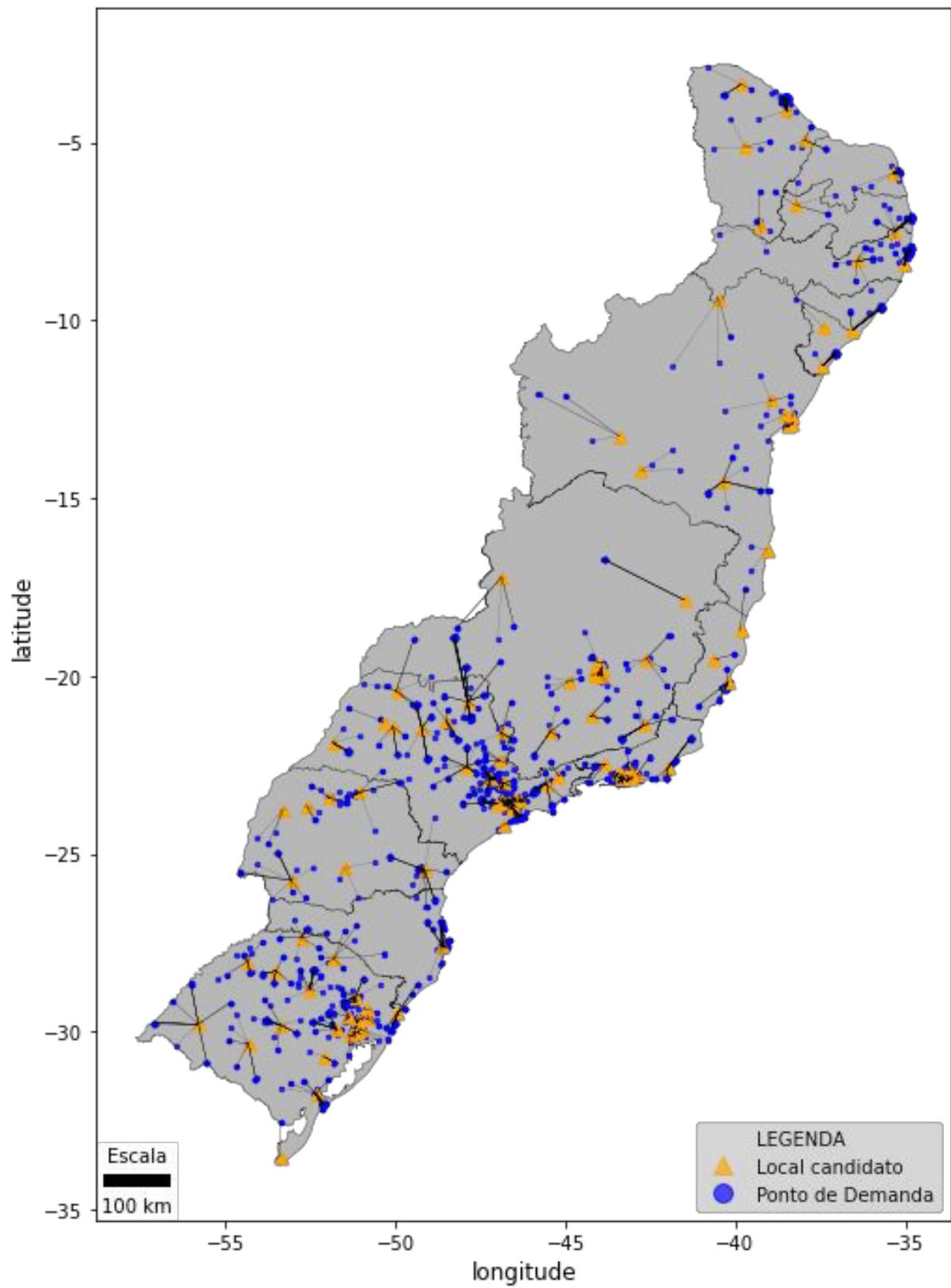


Figura B.18 – Alocação da demanda para o modelo Ma1 da instância 2-9

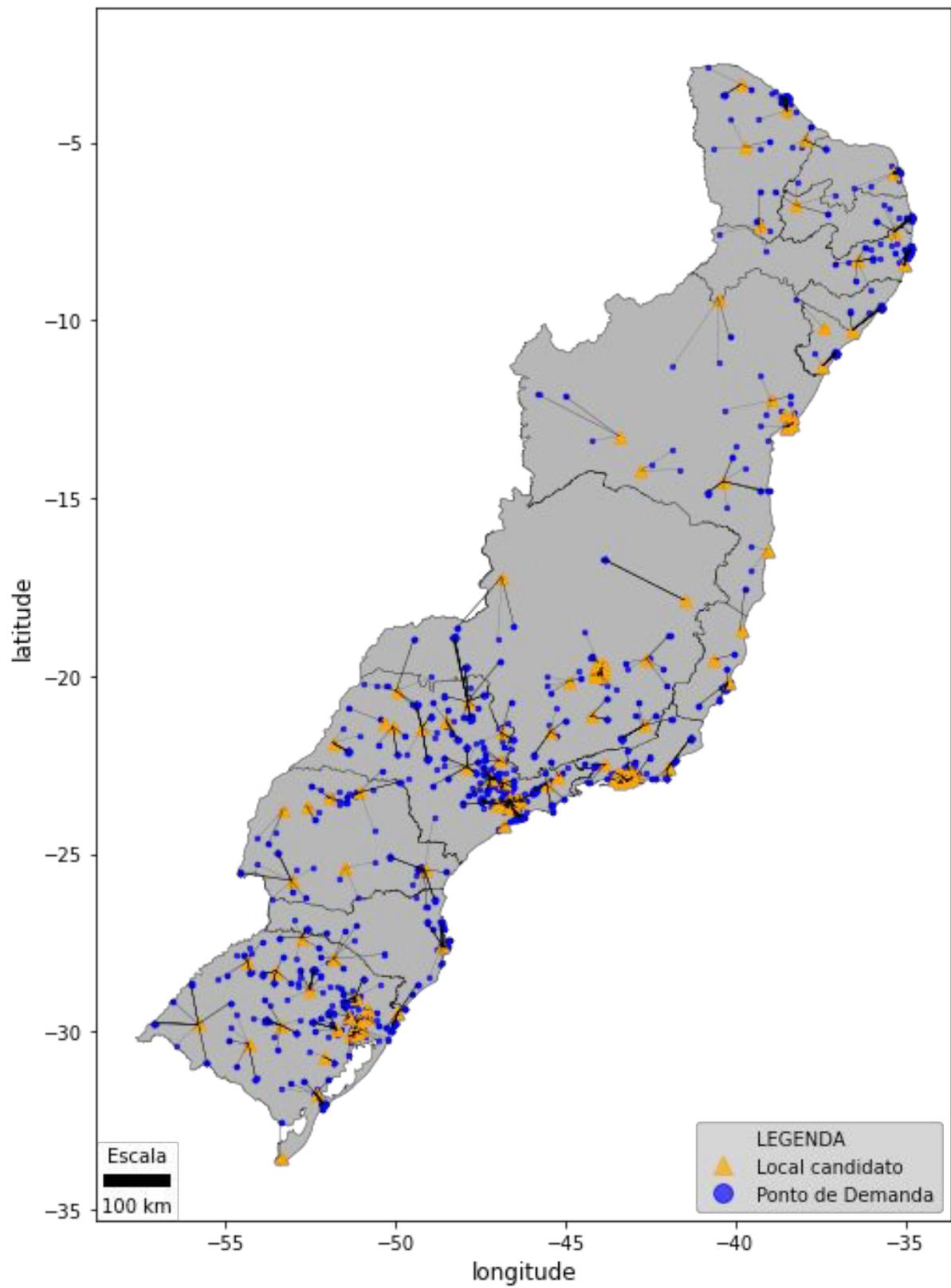


Figura B.19 – Alocação da demanda para o modelo Ma1 da instância 2-9

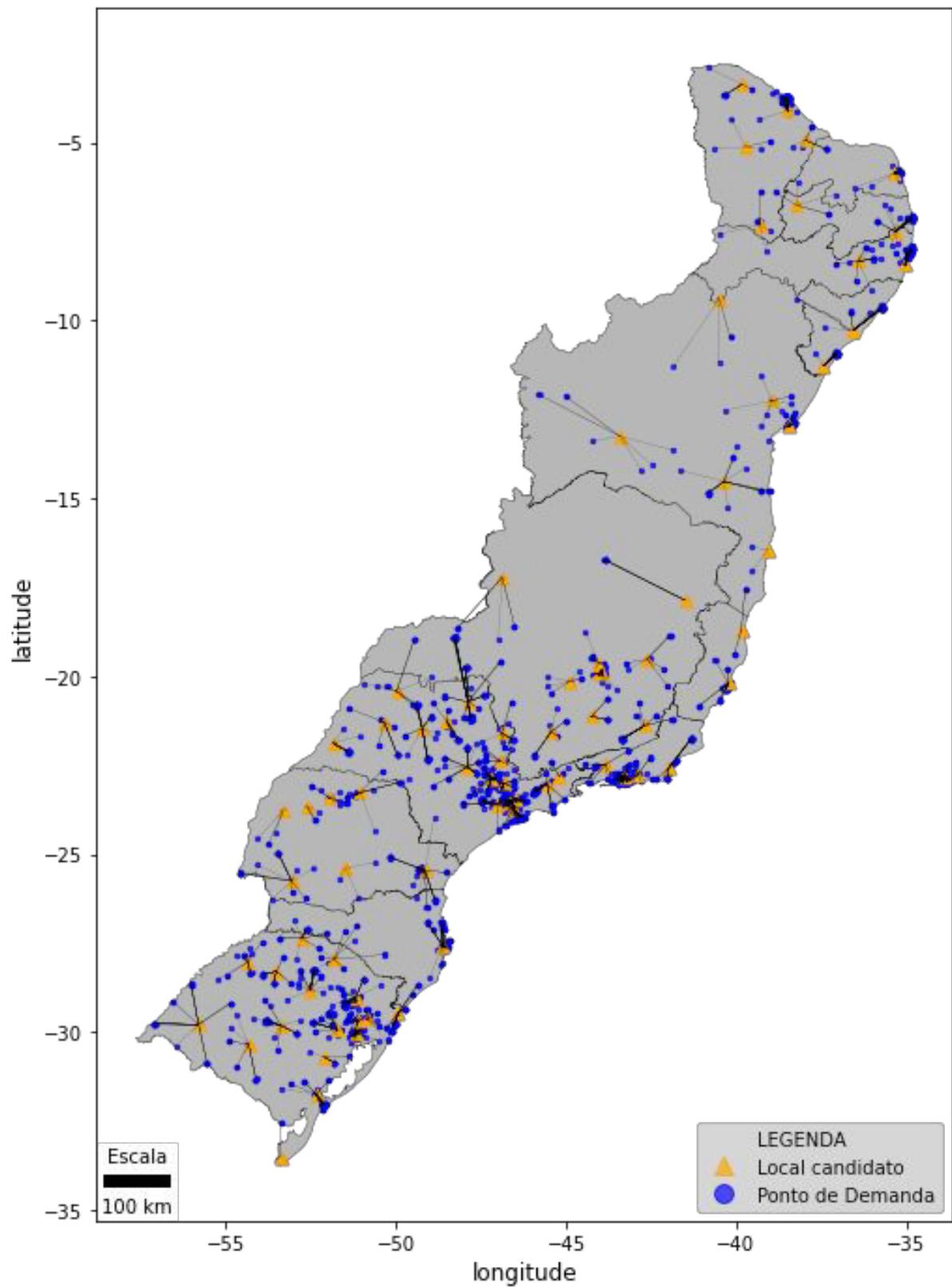


Figura B.20 – Alocação da demanda para os modelos desagregado e Ma2 da instância 2-90

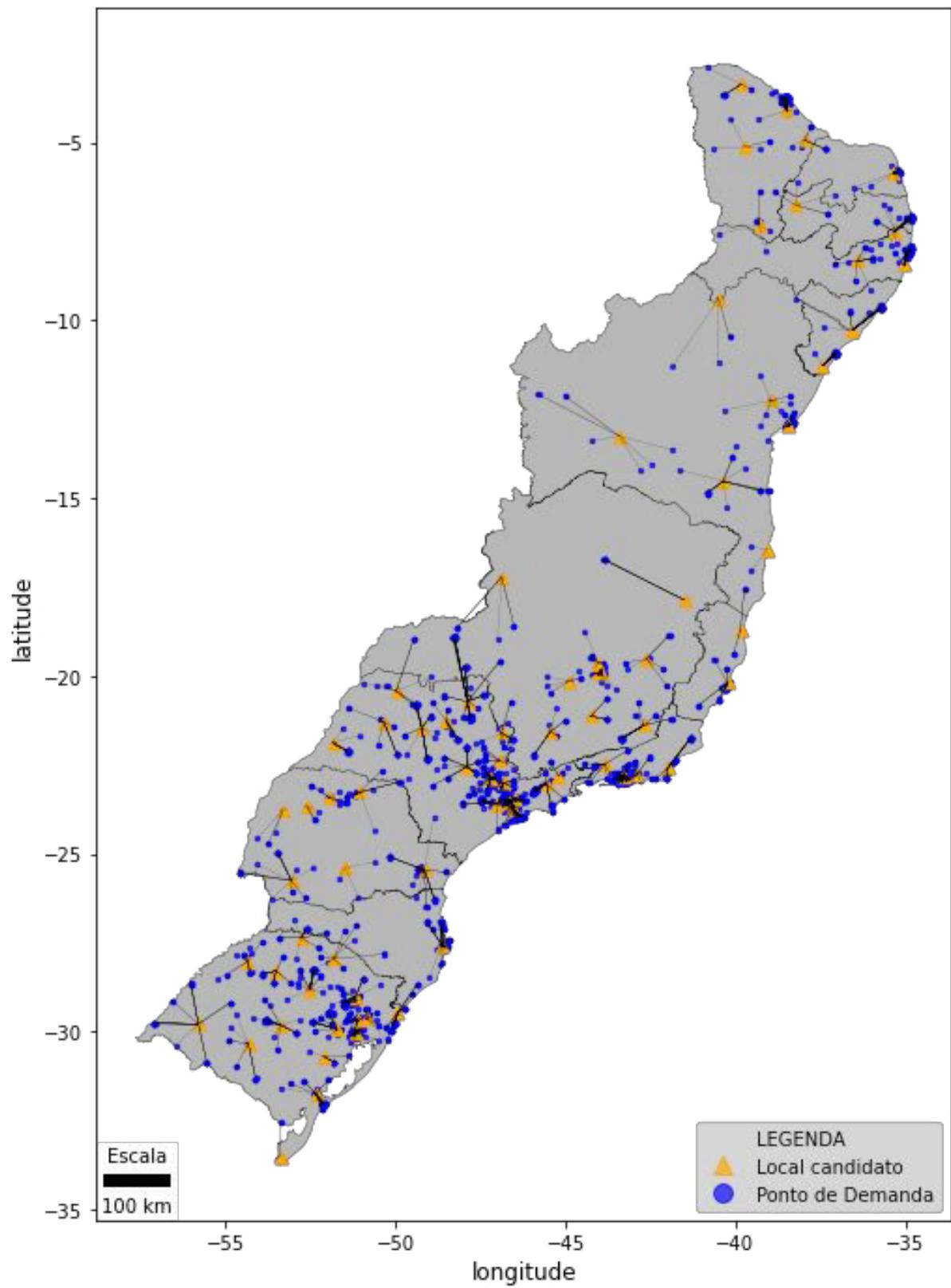


Figura B.21 – Alocação da demanda para o modelo Ma1 da instância 2-90

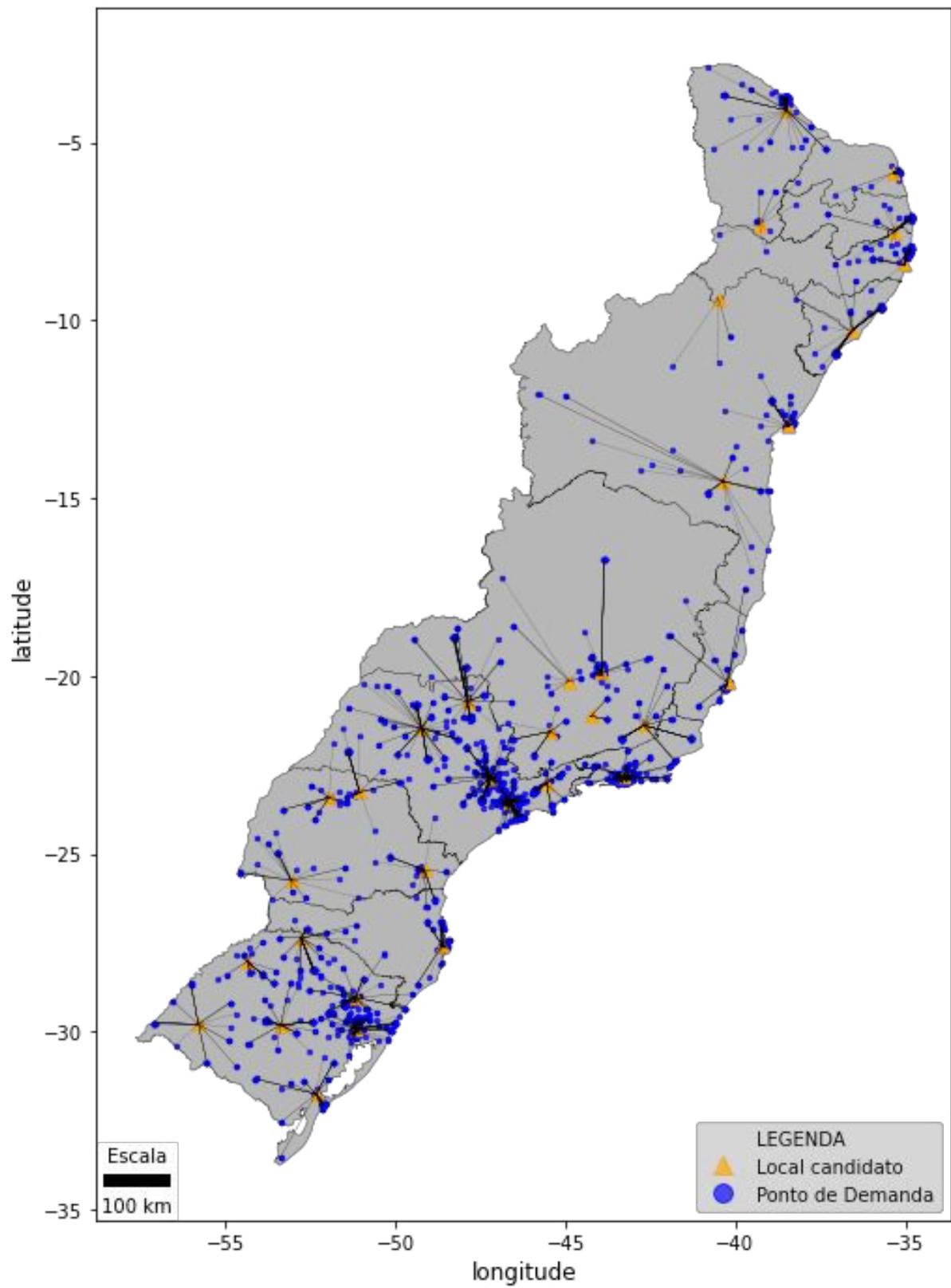


Figura B.22 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-900

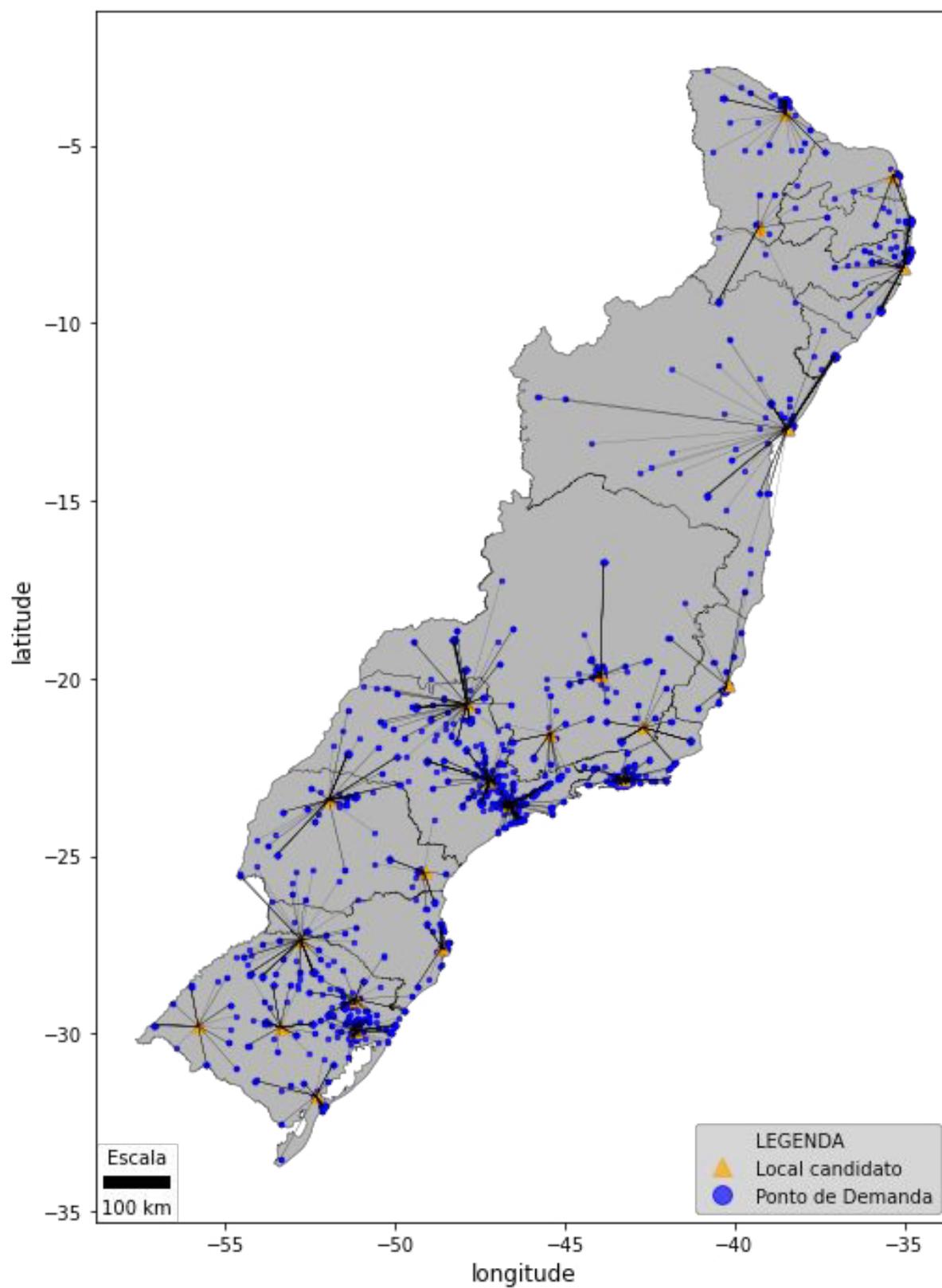


Figura B.23 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-1800

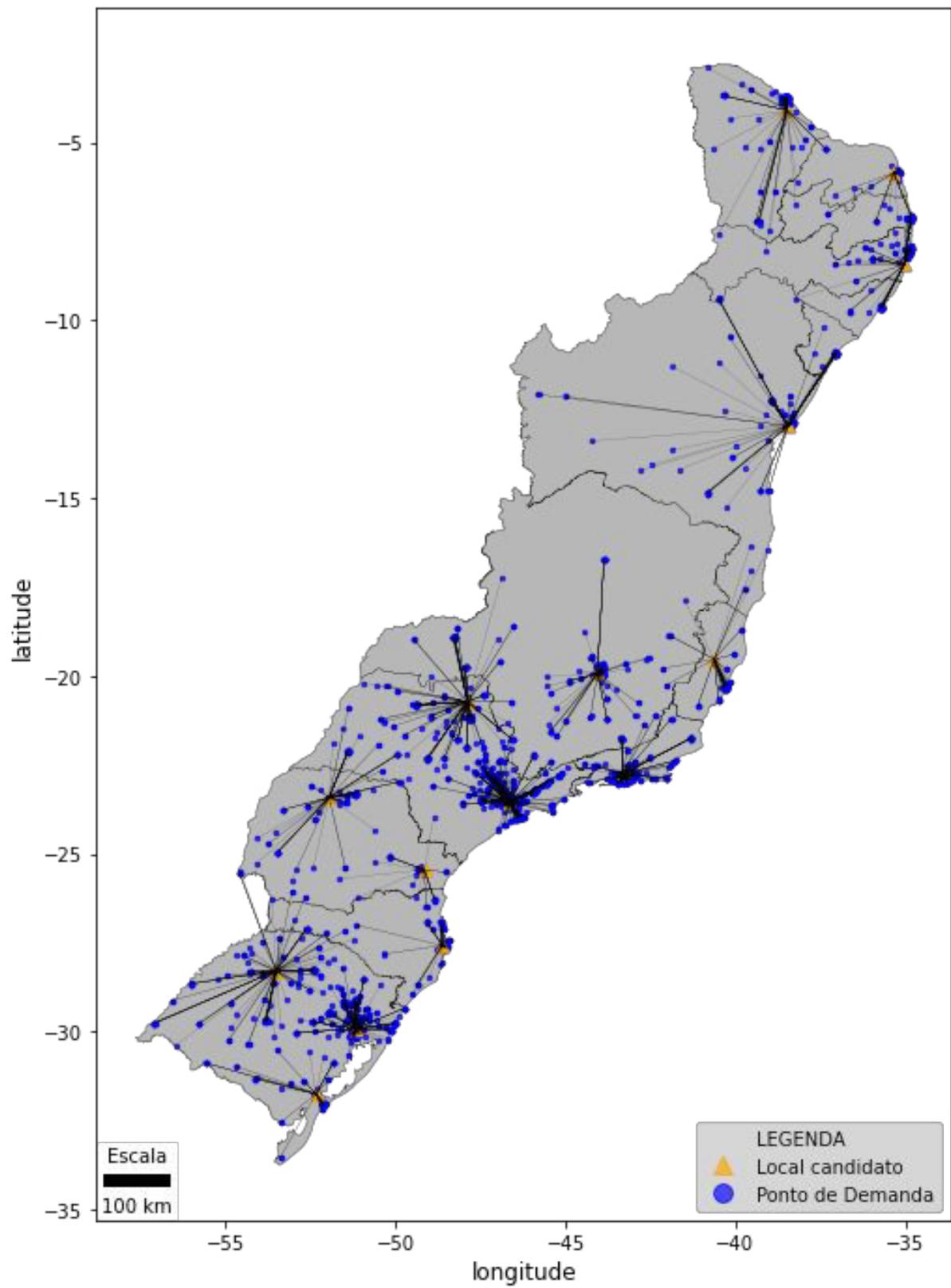


Figura B.24 – Alocação da demanda para o modelo desagregado da instância 2-4500

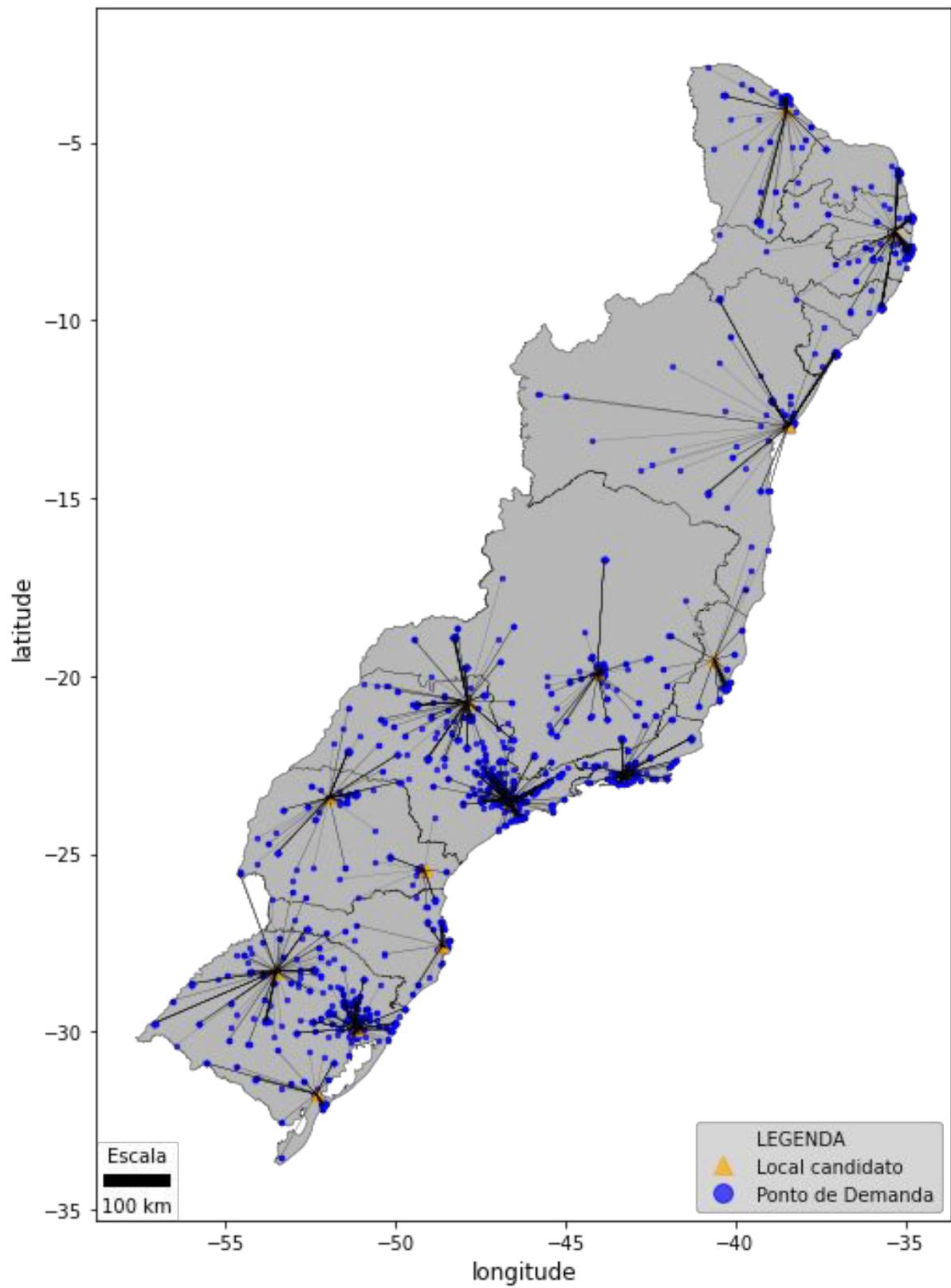


Figura B.25 – Alocação da demanda para o modelo Ma1 da instância 2-4500

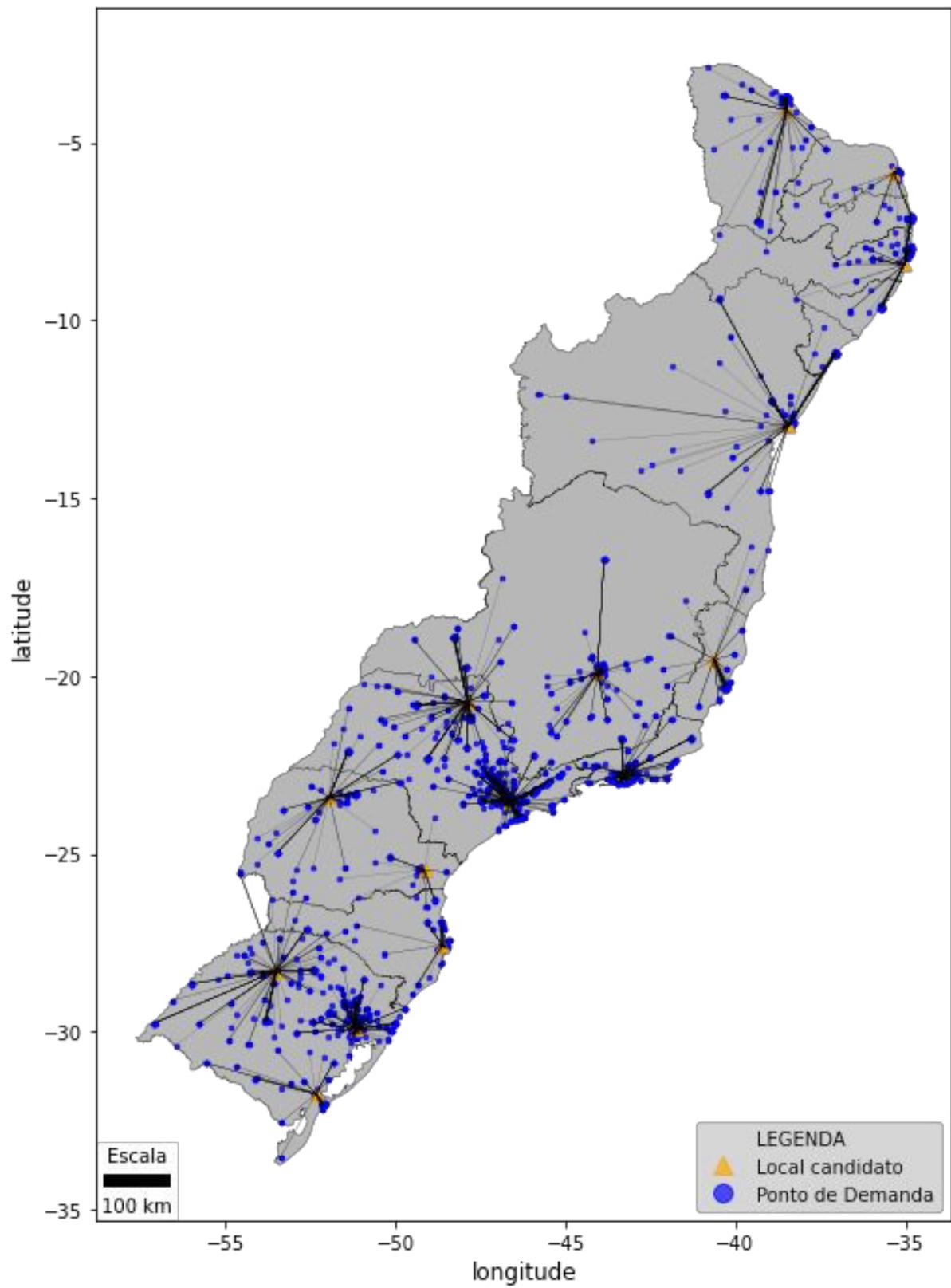


Figura B.26 – Alocação da demanda para o modelo Ma2 da instância 2-4500

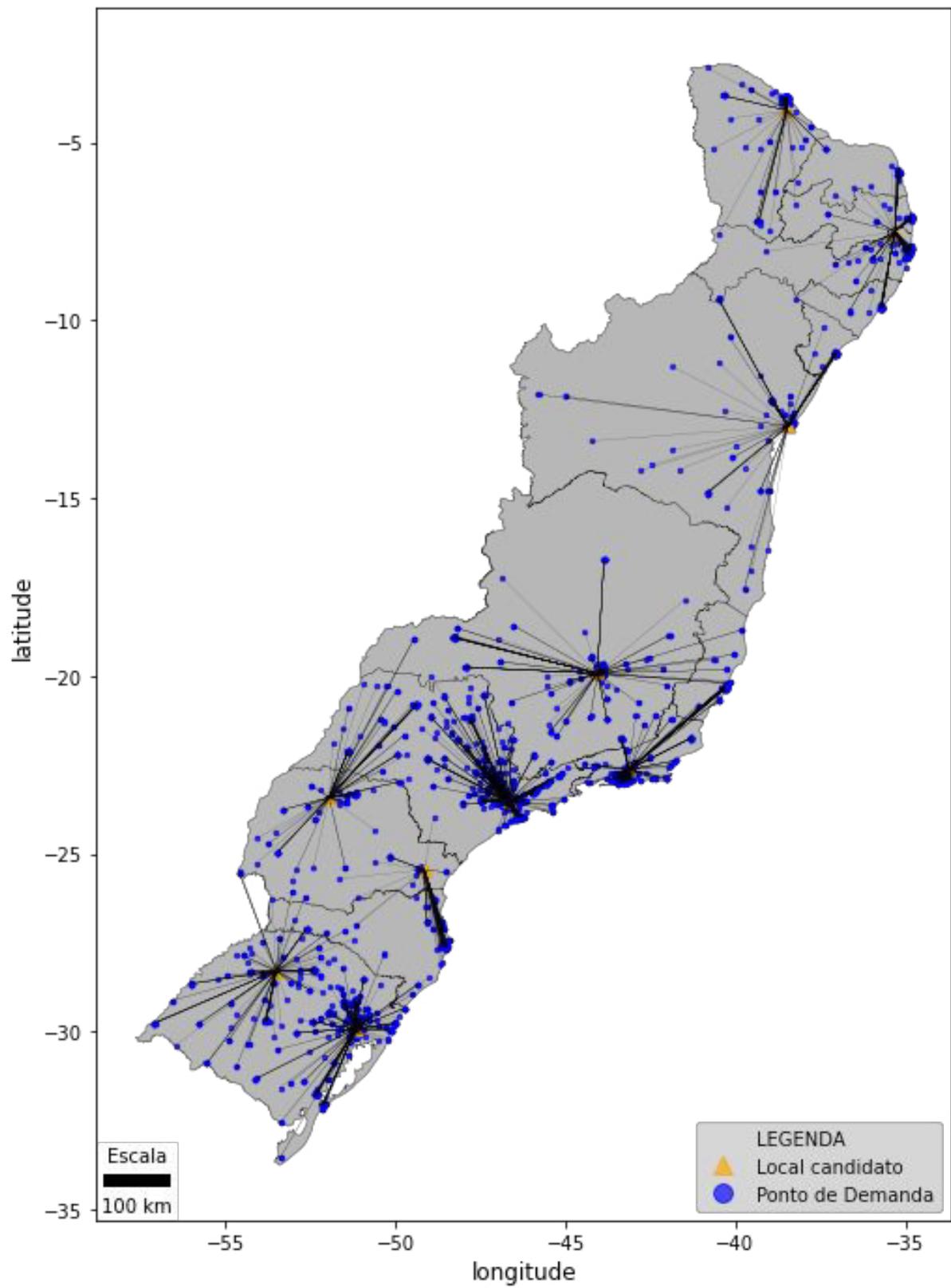


Figura B.27 – Alocação da demanda para os modelos desagregado e Ma2 da instância 2-9000

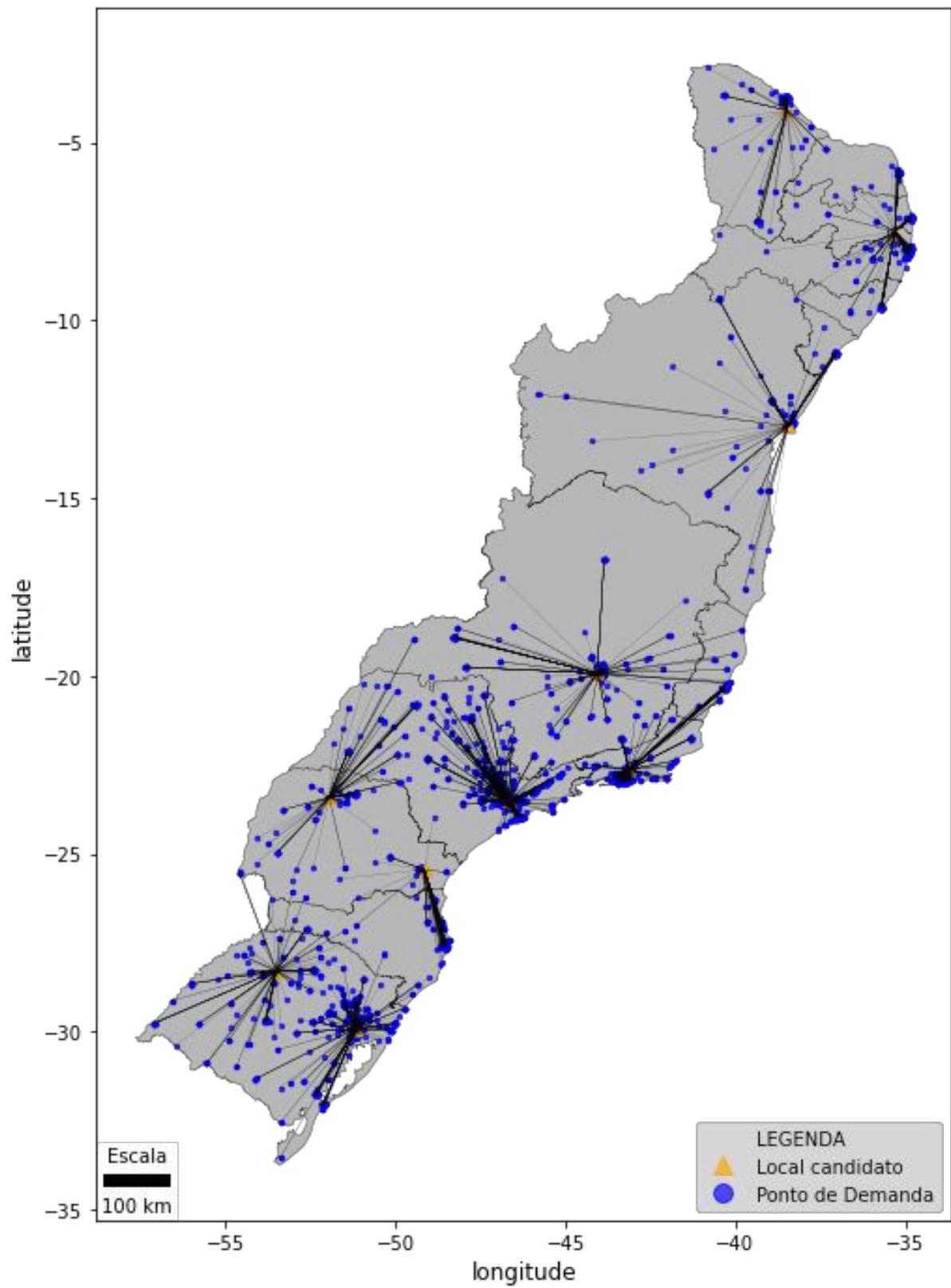


Figura B.28 – Alocação da demanda para o modelo Ma1 da instância 2-9000

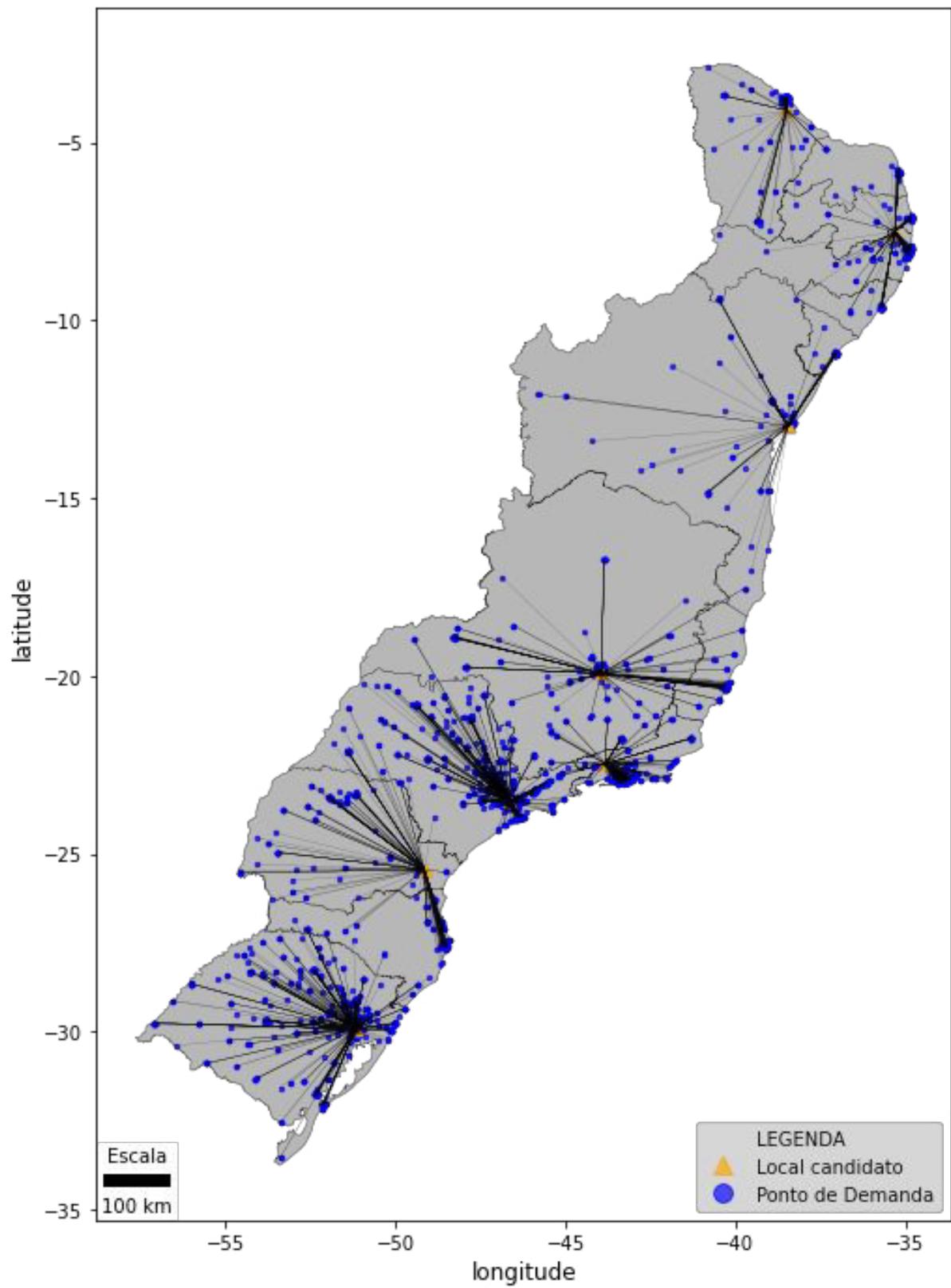


Figura B.29 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-18000

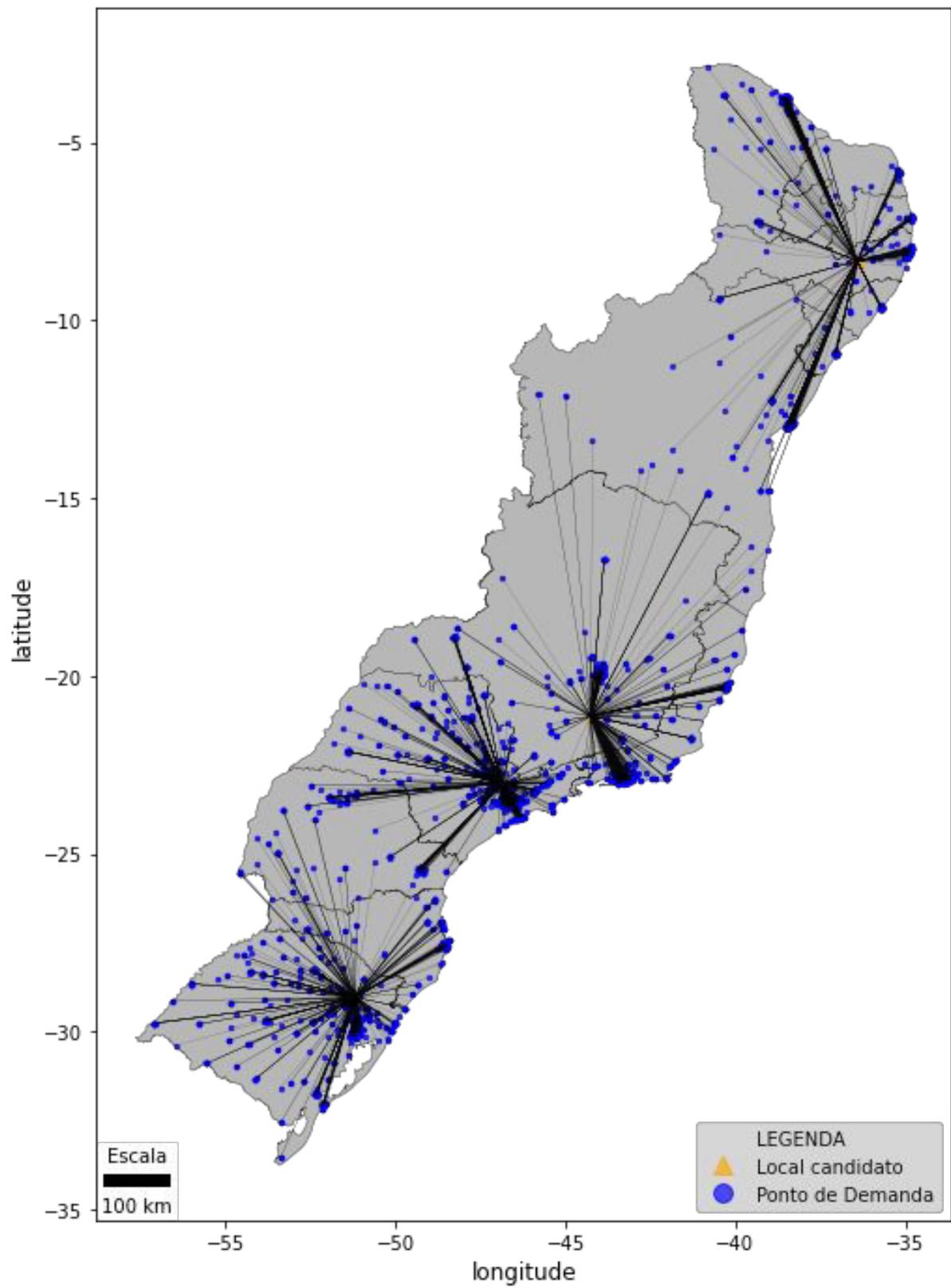


Figura B.30 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-45000

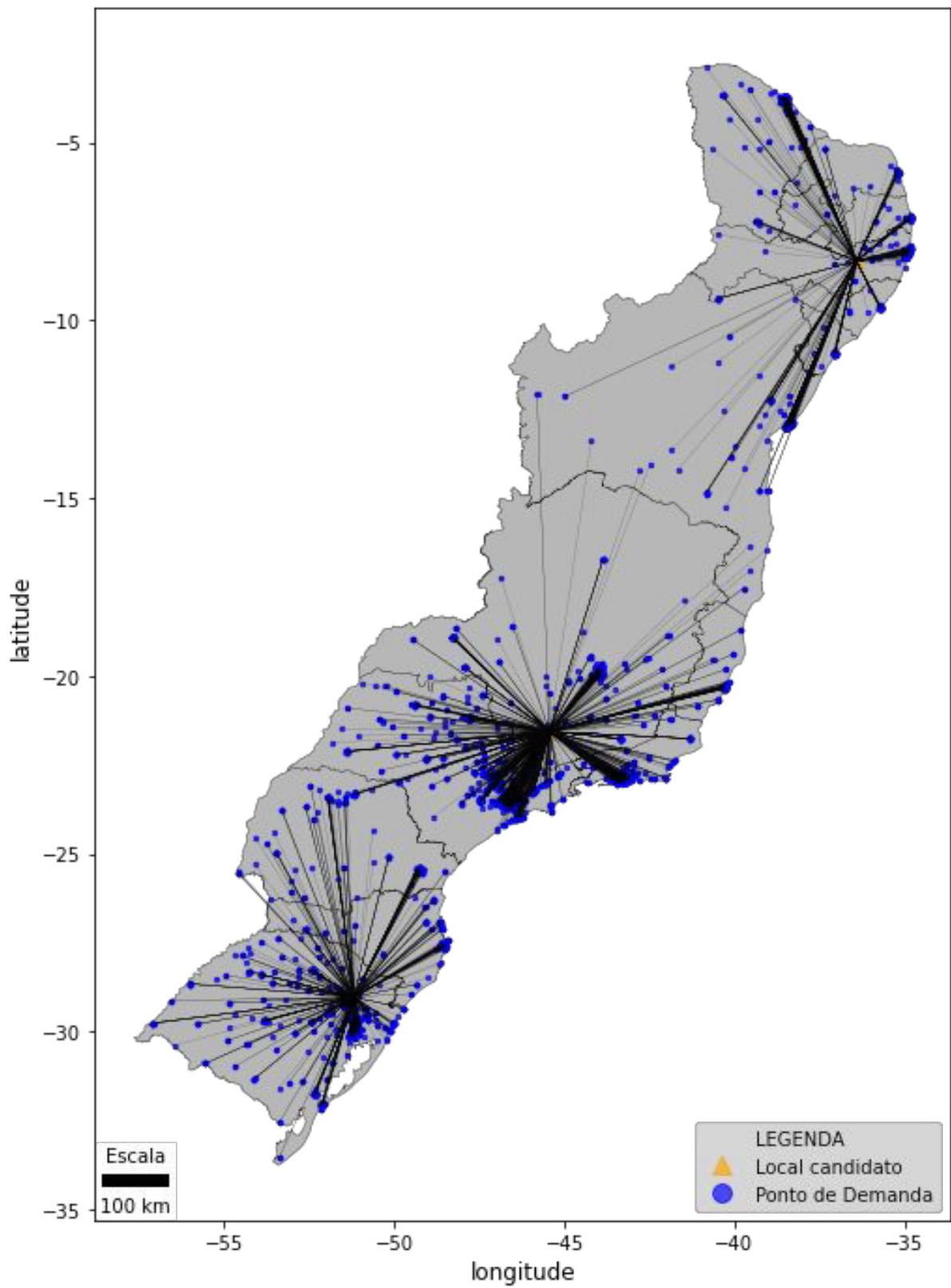


Figura B.31 – Alocação da demanda para os modelos desagregado, Ma1 e Ma2 da instância 2-90000