

# Internet Gratuita para Fins Turísticos

Análise Espacial da Distribuição de Pontos de Acesso

Gratuito à Internet na Região da Lazio, Itália



Análise Espacial de Dados

Sílvia Alexandra Da Silva Mourão, Nº FC57541

Lisboa, 24 de Junho de 2022

## **RESUMO**

*O constante crescimento e evolução das cidades e a sua integração com as novas tecnologias levou ao aparecimento da expressão “cidades inteligentes”, que tem no seu núcleo o conceito da interligação de dispositivos e pessoas, estando o sinal WiFi no centro desta interconectividade. Neste projeto será utilizado o exemplo da rede de WiFi de acesso gratuito em Itália, particularmente na região da Lazio e na cidade de Roma para avaliar a distribuição atual de pontos de acesso numa região com elevada densidade populacional.*

*O objetivo deste projeto é identificar se existe uma existência de um padrão aglomerado da distribuição pontual destes acessos, se existe uma autocorrelação espacial na região, compreender quais as outras características da região que podem explicar a distribuição dos pontos de acesso e pesquisar caminhos ótimos dentro da cidade de Roma para fins turísticos, bem como apresentar lacunas na cobertura. A abordagem utilizada consistiu num ciclo de visualização, análise exploratória e modelação, utilizando técnicas de análise espacial de dados como a análise do padrão pontual, análise de áreas, métodos de regressão lineares e espaciais, interpolação de superfícies e algoritmos de análise de redes. Os resultados obtidos indicam que existem clusters de valores em torno de cidades populosas e a autocorrelação espacial está presente na Lazio. A presença de hotéis, espaços de diversão noturna, pontos de interesse turísticos e comércio parece explicar a distribuição de pontos de acesso WiFi. A cobertura de sinal é insuficiente dentro da cidade de Roma para permitir percursos turísticos cobertos por sinal WiFi.*

## Índice

<b>1. INTRODUÇÃO .....</b>	<b>9</b>
<b>2. ÁREA DE ESTUDO E DADOS.....</b>	<b>10</b>
2.1. Área de Estudo.....	10
2.2. Descrição dos Dados e Metadados usados no projeto.....	11
2.3. Organização e tratamento dos dados.....	13
<b>3. METODOLOGIA.....</b>	<b>14</b>
<b>4. ANÁLISE DA DISTRIBUIÇÃO DO PADRÃO PONTUAL.....</b>	<b>15</b>
4.1 Organização dos Dados Espaciais .....	16
4.2 Visualização dos Dados Espaciais .....	16
4.3 Análise Exploratória de Dados Espaciais - Técnicas de Análise Exploratória de 1 <sup>a</sup> ordem .....	25
4.4 Análise Exploratória de Dados Espaciais - Técnicas de Análise Exploratória de 2 <sup>a</sup> ordem .....	33
4.5 Modelação do Padrão Pontual .....	35
<b>5. ANÁLISE DE ÁREAS .....</b>	<b>37</b>
5.1 - Organização dos dados em unidades de área .....	37
5.2 - Visualização dos Dados Espaciais (Métodos de corte da variável) .....	38
5.3 - Análise Exploratória dos Dados .....	41
5.4 – Modelação dos dados .....	44
<b>6 - MÉTODOS DE REGRESSÃO .....</b>	<b>47</b>
6.1- Análise da correlação entre as variáveis independentes e dependente .....	47
6.2 - Modelo de Regressão Clássico (OLS) - Simples e Múltipla .....	50
6.3 Avaliação do desempenho dos modelos numa área de estudo diferente .....	56
6.4 Avaliação dos resultados e diagnóstico dos resíduos do modelo OLS.....	59
6.5. Regressão Espacial.....	64
6.6 - Comparação dos modelos .....	68
<b>7. INTERPOLAÇÃO ESPACIAL .....</b>	<b>69</b>
7.1 Preparação / Tratamento dos dados .....	69
7.2 Métodos de Interpolação Determinísticos (Locais e Globais) - Sinal WiFi.....	70

Sílvia Mourão, FC 57541

7.3 Métodos de Interpolação Probabilísticos (Kriging) - Sinal WiFi.....	76
7.4 - Validação e análise comparativa dos resultados dos métodos de interpolação.....	79
7.5 - Produção de um Modelo Digital de Elevação a partir de métodos de interpolação .....	79
<b>8. ANÁLISE DE REDES .....</b>	<b>86</b>
8.1 Recolha e Tratamento de Dados .....	86
8.2 Criação de Topologia .....	87
8.3 Optimal Path.....	88
8.4 - Service Area.....	91
8.5 - Closest Facility.....	91
<b>9. CONCLUSÕES .....</b>	<b>94</b>
<b>REFERÊNCIAS .....</b>	<b>95</b>

**Índice de Figuras**

<i>Figura 1 - Enquadramento geográfico da área de estudo .....</i>	11
<i>Figura 2 - Metodologia da análise de dados espaciais .....</i>	14
<i>Figura 3 - Metodologia detalhada .....</i>	15
<i>Figura 4 - Distribuição pontos de acesso em Itália.....</i>	17
<i>Figura 5 - Distribuição pontos de acesso na Lazio.....</i>	18
<i>Figura 6 - Distribuição pontos de acesso em Roma.....</i>	18
<i>Figura 7 - Centro geométrico e mediana espacial da distribuição de pontos de acesso em Itália .....</i>	19
<i>Figura 8 - Centro geométrico e mediana espacial da distribuição de pontos de acesso na Lazio .....</i>	20
<i>Figura 9 - Centro geométrico e mediana espacial da distribuição de pontos de acesso em Roma .....</i>	20
<i>Figura 10 - Dispersão espacial dos pontos de acesso em Itália.....</i>	21
<i>Figura 11 - Elipse padrão dos pontos de acesso em Itália.....</i>	22
<i>Figura 12 - Dispersão espacial dos pontos de acesso na Lazio.....</i>	23
<i>Figura 13 - Elipse padrão dos pontos de acesso na Lazio.....</i>	23
<i>Figura 14 - Dispersão espacial dos pontos de acesso em Roma .....</i>	24
<i>Figura 15 - Elipse padrão dos pontos de acesso em Roma .....</i>	24
<i>Figura 16 - Método da contagem por quadrantes para Itália com grelha de 38 km x 38 km .....</i>	26
<i>Figura 17 - Método da contagem por quadrantes para a Lazio com grelha de 15 km x 15 km .....</i>	27
<i>Figura 18 - Método da contagem por quadrantes para Roma com grelha de 2 km x 2 km .....</i>	28
<i>Figura 19 - Estimador de densidade Kernel para Itália .....</i>	30
<i>Figura 20 - Localização das dez cidades mais populosas de Itália sobrepostas no estimador de densidade .....</i>	31
<i>Figura 21 - Estimador de densidade Kernel para a Lazio .....</i>	32
<i>Figura 22 - Estimador de densidade Kernel para Roma .....</i>	32
<i>Figura 23 - Resultados da Função G para os casos de Itália (esquerda), Lazio (centro), Roma (direita) .....</i>	33
<i>Figura 24 - Resultado Função K-Ripley para Itália .....</i>	34
<i>Figura 25 - Resultado Função K-Ripley para a Lazio.....</i>	34
<i>Figura 26 - Resultado Função K-Ripley para Roma .....</i>	35
<i>Figura 27 - Mapa temático com divisão por quantis para Itália (esquerda).....</i>	38
<i>Figura 28 - Mapa temático com divisão por intervalos iguais para Itália (direita).....</i>	38
<i>Figura 29 - Mapa temático com divisão por desvios padrões para Itália .....</i>	39

<i>Figura 30 - Mapa temático com divisão por quantis para Lazio (esquerda).....</i>	40
<i>Figura 31 - Mapa temático com divisão por intervalos iguais para Lazio (direita).....</i>	40
<i>Figura 32 - Mapa temático com divisão por desvios padrões para Lazio .....</i>	40
<i>Figura 33 - Histograma de vizinhança.....</i>	41
<i>Figura 34 - Scatterplot para o índice de Moran .....</i>	42
<i>Figura 35 - Índice de Moran local para a Lazio .....</i>	43
<i>Figura 36 -Estatística Gi para a Lazio (esquerda).....</i>	44
<i>Figura 37 - Estatística G* para a Lazio (direita) .....</i>	44
<i>Figura 38 - Teste de pseudo-significância ao índice global de Moran .....</i>	45
<i>Figura 39 - Mapa de significância para o índice local de Moran na Lazio .....</i>	46
<i>Figura 40 - Mapa de significância para a estatística Gi na Lazio (esquerda).....</i>	46
<i>Figura 41 - Mapa de significância para a estatística G* na Lazio (direita) .....</i>	46
<i>Figura 42 - Resultados dos testes de correlação entre o número de pontos de acesso e as variáveis independentes .....</i>	49
<i>Figura 43 - Coeficiente de determinação para a regressão com a variável independente hotéis .....</i>	51
<i>Figura 44 - Resíduos do modelo de regressão simples para a variável hotéis .....</i>	51
<i>Figura 45 - Estatística da Regressão Simples com a variável independente hotéis .....</i>	52
<i>Figura 46 - Coeficientes do modelo de regressão simples para a variável hotéis .....</i>	52
<i>Figura 47 - Coeficiente de determinação para a regressão com a variável independente turismo.....</i>	53
<i>Figura 48 - Resíduos do modelo de regressão simples para a variável turismo.....</i>	53
<i>Figura 49 - Estatística da Regressão Simples com a variável independente turismo.....</i>	54
<i>Figura 50 - Coeficientes do modelo de regressão simples para a variável turismo .....</i>	54
<i>Figura 51 - Primeira tentativa de regressão múltipla .....</i>	54
<i>Figura 52 - Resíduos do modelo de regressão múltipla.....</i>	55
<i>Figura 53 - Estatística da regressão múltipla .....</i>	55
<i>Figura 54 - Coeficientes do modelo de regressão múltipla .....</i>	56
<i>Figura 55 - Resíduos do modelo de regressão simples com a variável hotéis aplicada à região da Lombardia .....</i>	57
<i>Figura 56 - Resíduos do modelo de regressão simples com a variável turismo aplicada à região da Lombardia .....</i>	57
<i>Figura 57 - Resíduos do modelo de regressão múltipla aplicado à região da Lombardia .....</i>	58
<i>Figura 58 - Distribuição dos resíduos .....</i>	59
<i>Figura 59 - Histograma dos resíduos.....</i>	60

<i>Figura 60 - QQ plot dos resíduos .....</i>	61
<i>Figura 61 - Índice de Moran Global para a autocorrelação espacial dos resíduos .....</i>	62
<i>Figura 62 - índice de Moran local para os resíduos.....</i>	62
<i>Figura 63 - Mapa de significância para os resíduos .....</i>	63
<i>Figura 64 - Resultados da regressão spatial lag.....</i>	64
<i>Figura 65 - Resultados da regressão spatial error.....</i>	64
<i>Figura 66 - Mapeamento do desvio padrão dos resíduos obtidos pelo GWR .....</i>	65
<i>Figura 67 - Valores reais de pontos de acesso na Lazio (esquerda) .....</i>	66
<i>Figura 68 - Valores estimados de pontos de acesso pelo modelo de regressão geográfica ponderada (direita).....</i>	66
<i>Figura 69 - Variação de <math>R^2</math> calculado pelo modelo GWR.....</i>	66
<i>Figura 70 - Regiões de maior e menor contribuição para cada uma das variáveis independentes .....</i>	67
<i>Figura 71 - Superfície de tendência interpolada para o sinal WiFi.....</i>	71
<i>Figura 72 - Polígonos de Thiessen para o sinal WiFi .....</i>	72
<i>Figura 73 - Detalhe dos polígonos de Thiessen para o sinal WiFi .....</i>	72
<i>Figura 74 - Interpolador local determinístico para o sinal WiFi .....</i>	73
<i>Figura 75 - Interpolação pelo método do inverso da distância ponderada para o sinal WiFi.....</i>	74
<i>Figura 76 - Detalhe da interpolação pelo método do inverso da distância ponderada para o sinal WiFi .....</i>	74
<i>Figura 77 - Interpolação pelo método da média móvel espacial para o sinal WiFi .....</i>	75
<i>Figura 78 – Detalhe da interpolação pelo método da média móvel espacial para o sinal WiFi .....</i>	76
<i>Figura 79 - Semi-variograma e aproximação do modelo .....</i>	77
<i>Figura 80 - Interpolação pelo método Kriging para o sinal WiFi.....</i>	78
<i>Figura 81 - Detalhe da interpolação pelo método Kriging para o sinal WiFi .....</i>	78
<i>Figura 82 - Geração de MDT a partir de uma TIN .....</i>	80
<i>Figura 83 - Interpolação de altimetria resultante da TIN.....</i>	80
<i>Figura 84 - Interpolação da altimetria pelo inverso da distância ponderada .....</i>	81
<i>Figura 85 - Inclinação do terreno .....</i>	82
<i>Figura 86 - Orientação do relevo do terreno .....</i>	82
<i>Figura 87 - Zonas de sombra .....</i>	83
<i>Figura 88 - Estudo da altimetria nas localizações de pontos de interesse turístico .....</i>	84
<i>Figura 89 - Detalhe de altimetria na zona da ilha tiberina .....</i>	85
<i>Figura 90 - Fotografia da localização na realidade .....</i>	85

Sílvia Mourão, FC 57541

<i>Figura 91 - Correção de erros de topologia .....</i>	88
<i>Figura 92 - Rota de pontos turísticos utilizando apenas o comprimento como custo .....</i>	89
<i>Figura 93 - Rota de pontos turísticos utilizando o sinal WiFi como custo, sobreposto na interpolação de sinal WiFi obtida pela média móvel espacial .....</i>	90
<i>Figura 94 - Comparação das duas soluções da rede .....</i>	90
<i>Figura 95 - Área de serviço dos pontos de acesso .....</i>	91
<i>Figura 96 - Pesquisa do ponto de acesso WiFi mais próximo de cada ponto de interesse turístico .....</i>	92
<i>Figura 97 - Detalhe da pesquisa de acesso WiFi mais próximo.....</i>	92
<i>Figura 98 - Distância mais longa a um ponto de acesso .....</i>	93

**Índice de Tabelas**

<i>Tabela 1 - Dados, fontes e metadados utilizados no projeto .....</i>	12
<i>Tabela 2 - Medidas das regiões em estudo .....</i>	16
<i>Tabela 3 - Resultados estatísticos do método de contagem por quadrantes para Itália.....</i>	26
<i>Tabela 4 - Resultados estatísticos do método de contagem por quadrantes para a Lazio.....</i>	27
<i>Tabela 5 - Resultados estatísticos do método de contagem por quadrantes para Roma .....</i>	29
<i>Tabela 6 - Resultados do teste de <math>\chi^2</math> para o método de contagem por quadrantes para a região de Itália .....</i>	36
<i>Tabela 7 - Resultados do teste de <math>\chi^2</math> para o método de contagem por quadrantes para a região da Lazio .....</i>	36
<i>Tabela 8 - Resultados do teste de <math>\chi^2</math> para o método de contagem por quadrantes para a região de Roma .....</i>	36
<i>Tabela 9 - Resultados estatísticos Função G .....</i>	36
<i>Tabela 10 - Resultados dos coeficientes de determinação obtidos para a variável dependente.....</i>	50
<i>Tabela 11 - Análise da performance dos modelos na região da Lombardia .....</i>	58
<i>Tabela 12 - Sumário dos resultados obtidos para as regressões lineares .....</i>	63
<i>Tabela 13 - Comparação dos modelos de regressão.....</i>	68
<i>Tabela 14 - Comparação de resultados obtidos pelos vários métodos de interpolação para o sinal WiFi .....</i>	79
<i>Tabela 15 - Comparação de resultados obtidos pelos vários métodos de interpolação para a altimetria .....</i>	81

## 1. Introdução

As cidades modernas estão cada vez mais viradas para tecnologias especializadas para dar resposta a questões relacionadas com sustentabilidade e sociedade. O conceito de cidades inteligentes já não é apenas uma visão futura do planeamento urbano, mas sim um processo em implementação contínua (Suchocka et al., 2019). O aumento do interesse em cidades inteligentes traz consigo um aumento de objetos inteligentes em localizações públicas, o que requer conectividade permanente de forma que estes objetos possam comunicar os seus dados e receber configurações. Os pontos de acesso WiFi gratuitos são alternativas a redes privadas, que têm um custo de implementação mais elevado e que não servem à comunidade em geral (Valenzano et al., 2016). No entanto, existem ainda limitações tanto na extensão destas redes WiFi como nos próprios protocolos utilizados para transmissão de dados, que podem não estar preparados para a enorme quantidade de dispositivos que estão interligados entre si e com um ponto de acesso (Yaqoob et al., 2017). Com o crescimento continuo da proporção de população que habita em cidades e com a tendência de integração de dispositivos em cidades inteligentes, a transferência de informação através da internet é um dos fatores mais críticos que podem condicionar o seu desenvolvimento, sendo necessário que a qualidade e quantidade de sinal WiFi disponível acompanhe as necessidades dos utilizadores (Zhang et al., 2017)(Li et al., 2019). A análise espacial de dados é um processo de interpretação, exploração e modelação de dados de um sistema de informação geográfica. A informação obtida é processada com recurso a software GIS que permite realizar várias medições sobre os dados, como a medição de distâncias, o estabelecimento de correlações entre eventos ou a interpolação de superfícies a partir de dados pontuais (*Spatial Analysis Of Geospatial Data: Processing And Use Cases*, n.d.).

As ferramentas de análise espacial são então críticas para compreender as oportunidades apresentadas pelo estabelecimento de redes WiFi públicas em cidades inteligentes (Evennou & Marx, 2006; Gao & Prasad, 2016) mas também para compreender as limitações que existem ainda hoje em dia no que toca não só à disponibilidade e qualidade do sinal (Grubesic & Murray, 2004), mas também à densidade de sinal necessária para as aplicações de uma cidade moderna.

Neste trabalho, a análise espacial de dados vai ser utilizada no âmbito da compreensão da distribuição espacial de pontos de acesso WiFi gratuitos no território italiano, com particular foco na região da Lazio.

O objetivo deste estudo é compreender como é que estes pontos estão distribuídos, quais os fatores que influenciam a sua distribuição, a existência de lacunas na rede e que tipo de cobertura pode um

Sílvia Mourão, FC 57541

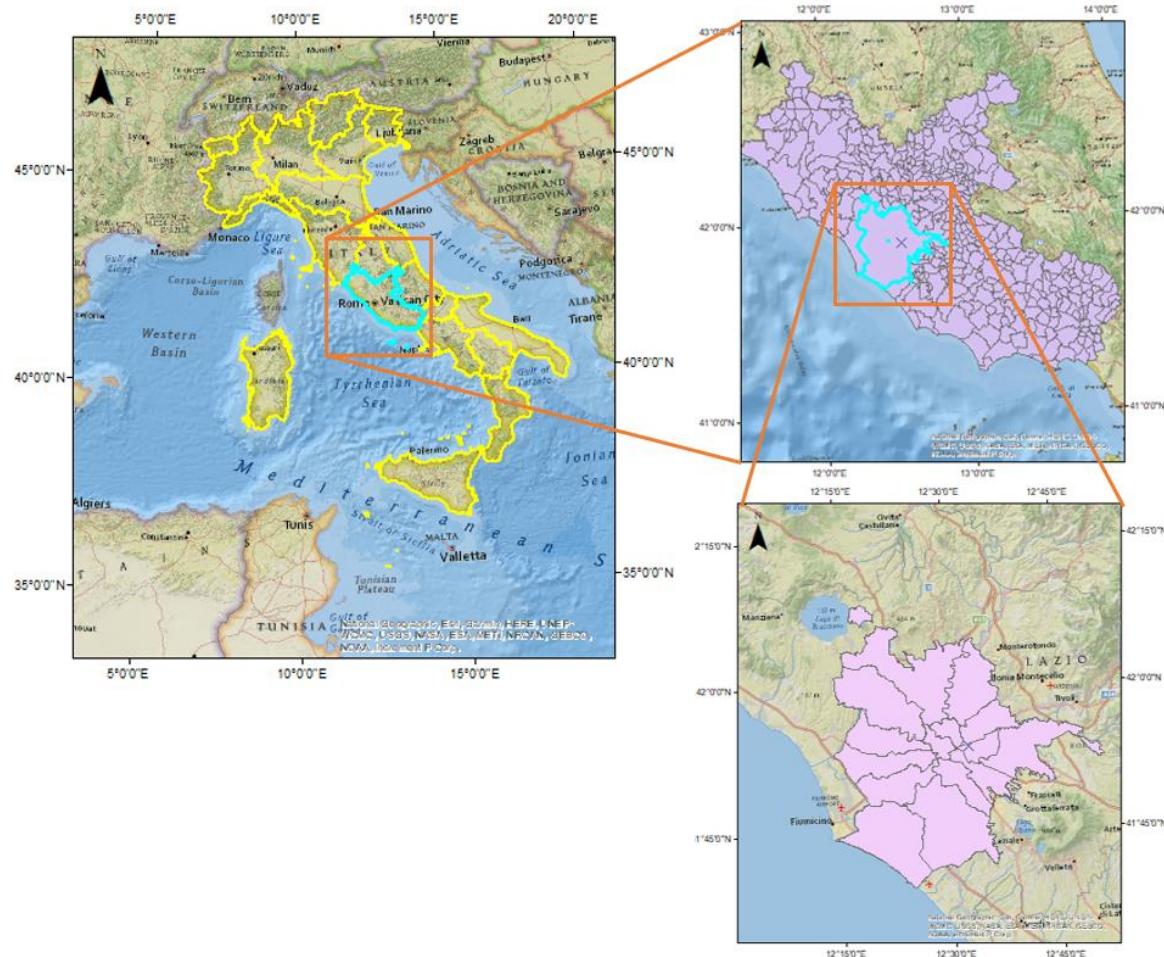
turista esperar quando visita os monumentos mais populares na cidade de Roma. Para isso, o trabalho vai ser dividido em cinco etapas: análise do padrão pontual, análise espacial de áreas, métodos de regressão, análise de superfícies e análise de redes. Para cada uma destas etapas foi adotada uma metodologia em três partes, que consiste numa primeira parte de visualização dos dados espaciais, seguida por uma etapa de análise exploratória, que vai permitir a classificação do tipo de dados que está a ser observado, e finalmente a etapa de modelação de dados, onde serão utilizados métodos estatísticos para validar os resultados obtidos.

## 2. Área de Estudo e Dados

### 2.1. Área de Estudo

Este projeto pretende estudar a distribuição espacial de pontos de acesso WiFi em Itália.

As áreas consideradas durante o processo são três, e que podem ser vistas na figura 1. Inicialmente utilizou-se toda a área de Itália para o estudo, no entanto devido à extensão deste país foi necessário reduzir para uma área mais pequena. Itália encontra-se dividida em vinte regiões, sendo as mais populosas a região da Lombardia, a norte, que inclui a cidade de Milão, e a região da Lazio, na zona central, que inclui a cidade de Roma, a capital do país (*Regioni Italiane*, n.d.). Entre estas regiões, foi escolhida aquela que apresentava maior densidade de pontos de acesso, a Lazio para a fase seguinte do projeto. Havendo ainda necessidade numa fase posterior de reduzir novamente a área de estudo, foi utilizada a área metropolitana da cidade de Roma, a região com maior densidade de eventos no país inteiro.



*Figura 1 - Enquadramento geográfico da área de estudo*

## 2.2. Descrição dos Dados e Metadados usados no projeto

Com o objetivo de analisar a distribuição da cobertura WiFi em Itália foram procurados dados relativamente a esses pontos de acesso gratuito, mas também shapefiles do território em estudo, assim como outros eventos que poderão vir a ser necessários durante o projeto como localizações de pontos turísticos, vias rodoviárias, dados populacionais, entre outros. Todos estes dados foram adquiridos a partir de repositórios de dados abertos, como o Open Street Map ou Dati.it, site de dados abertos italiano. Uma descrição mais exaustiva das fontes de dados encontra-se na tabela 1:

*Tabela 1 - Dados, fontes e metadados utilizados no projeto*

Dados	Fonte	Metadados
<b>Representação Geográfica Itália</b>	Geodati - <a href="https://geodati.gov.it/geoportale">https://geodati.gov.it/geoportale</a>	Shapefile (shp); Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Lista de Hotspots WiFi</b>	Free Itália WiFi <a href="https://opendata.cittametropolitanaroma.it/">https://opendata.cittametropolitanaroma.it/</a>	Ficheiros de texto (csv), Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Lista de Hotspots WiFi</b>	Bologna WiFi <a href="https://opendata.comune.bologna.it/">https://opendata.comune.bologna.it/</a>	Ficheiros de texto (csv), Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Lista de Hotspots WiFi</b>	Workshop <a href="https://opendata.comune.bologna.it/">https://opendata.comune.bologna.it/</a>	Ficheiros de texto (csv), Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Lista de Hotspots WiFi</b>	WiFi Itália <a href="https://www.wifi.italia.it/it/">https://www.wifi.italia.it/it/</a>	Ficheiros de texto (csv), Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Dados Turismo por Área; População residente</b>	ISTAT <a href="https://www.istat.it/en/analysis-and-products/datasets">https://www.istat.it/en/analysis-and-products/datasets</a>	Ficheiros de texto (csv)
<b>Rede rodoviária</b>	Open Street Map <a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a>	Shapefile (shp); Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Curvas de nível</b>	Open Street Map via Geofabrik <a href="https://www.geofabrik.de/">https://www.geofabrik.de/</a>	Shapefile (shp); Sistema de referenciamento geográfico WGS84, Datum WGS84, 1:10000 m
<b>Pontos de interesse vários na região da Lazio</b>	Open Street Map via Geofabrik <a href="https://www.geofabrik.de/">https://www.geofabrik.de/</a>	Shapefile (shp); Sistema de referenciamento geográfico WGS84, Datum WGS84
<b>Lista de atrações mais populares em Roma</b>	Touropia <a href="https://www.touropia.com/tourist-attractions-in-rome/">https://www.touropia.com/tourist-attractions-in-rome/</a>	Ficheiros de texto (csv)
<b>Shapefile Municípios Roma, Milão</b>	Dati Open <a href="http://www.datopen.it/">http://www.datopen.it/</a>	Shapefile (shp);

		Sistema de referenciação geográfica WGS84, Datum WGS84
<b>Localização hotéis, hostels e airbnbs</b>	Dati Open <a href="http://www.datiopen.it/">http://www.datiopen.it/</a>	Ficheiros de texto (csv), Sistema de referenciação geográfica WGS84, Datum WGS84

### 2.3. Organização e tratamento dos dados

A principal tarefa da etapa de tratamento dos dados foi a normalização destes entre as várias fontes de aquisição.

Todos os dados obtidos no sistema de referência geográfico WGS84 foram projetados para um sistema de referência cartesiano, tendo sido para este trabalho escolhida o sistema de referência italiano, denominado *Rete Dinamica Nazionale (RDN2008 / Italy Zone (E-N) - EPSG:7794, n.d.)* devido ao facto de o país estar dividido em duas zonas UTM diversas.

Uma outra etapa foi a transformação dos dados obtidos em formato CSV para shapefiles e a sua posterior conversão também para o mesmo sistema de referência. Devido ao facto de que os dados foram obtidos de várias fontes diversas, foi necessário verificar a existência de repetições de dados – por exemplo existe uma rede de pontos WiFi em hospitais onde existe um ponto de acesso WiFi em cada piso, sendo que estes foram reduzidos a apenas um ponto por coordenadas pois a análise realizada no projeto tem apenas componente planimétrica.

Na fase da regressão existiu a necessidade de procurar mais informação para tentar encontrar fatores que pudessem explicar o fenómeno em estudo. Foram encontradas shapefiles com muitos dados de variados pontos de interesse nas fontes já descritas na tabela 1, no entanto foi necessário realizar alguma filtragem destes dados para que pudessem ser utilizados apenas aqueles relevantes. Por exemplo, a shapefile dos pontos de interesse foi separada em pontos turísticos, pontos de comércio, pontos de diversão noturna, tendo ainda outros dados sido apagados por não serem relevantes para o projeto.

As tabelas de população e indicadores turísticos não continham informação espacial, mas apenas o nome da zona a que se referiam. Estas foram ligadas com os outros dados existentes através de uma operação de juncção no ArcMap.

Finalmente, uma outra etapa que necessitou de grande edição de dados foi a etapa das redes, cujo tratamento de dados será descrito em maior detalhe no capítulo X.

### 3. Metodologia

O projeto dividiu-se em cinco etapas, que seguiram aproximadamente a metodologia de visualização, análise exploratória e modelação dos dados (figura 2). As operações realizadas dentro de cada um destes capítulos divergem dependendo da etapa considerada.



Figura 2 - Metodologia da análise de dados espaciais

Os métodos de análise considerados para cada fase do projeto estão descritos na figura 3:

### Análise da distribuição do padrão pontual

- Visualização dos dados
  - Medidas de Centrografia
  - Análise exploratória dos dados
    - Método da contagem por quadrantes
    - Estimador de densidade
    - Função G
    - Função K
  - Modelação dos dados

### Análise espacial de áreas

- Visualização dos dados
  - Mapas temáticos - métodos de corte da variável
- Análise exploratória dos dados
  - Índices de autocorrelação global
  - Índices de autocorrelação local
- Modelação dos dados

### Métodos de Regressão

- Visualização dos dados
- Identificação de variáveis independentes
- Análise exploratória dos dados
  - Regressão clássica
- Modelação dos dados
  - Avaliação de resultados e resíduos
  - Modelos de regressão espacial
  - Comparação dos modelos

### Análise de superfícies

- Métodos de interpolação determinísticos
- Métodos de interpolação probabilísticos
- Validação e análise comparativa dos modelos

### Análise de redes

- Tratamento de dados
- Criação de topologia
- Algoritmos de redes
  - Best route
  - Service area
  - Closest Facility

*Figura 3 - Metodologia detalhada*

## 4. Análise da distribuição do Padrão Pontual

A análise da distribuição do padrão pontual consiste no estudo da distribuição espacial dos eventos e a sua classificação num dos três tipos de padrão sistemático: aleatório, aglomerado (cluster) ou uniforme. Os principais objetivos nesta fase do projeto são a quantificação da dispersão de eventos na área geográfica considerada, a estimativa da variação da densidade dos eventos na região de estudo

Sílvia Mourão, FC 57541

e a procura de modelos teóricos que expliquem esta distribuição, sendo estes posteriormente validados através de métodos estatísticos (Catita, 2022).

Com o objetivo de fornecer uma visão global da variável, mas também de aprofundar o estudo a áreas que irão posteriormente ser consideradas para outras fases do projeto, esta análise foi dividida em três áreas, consistindo a primeira na área total do país (Itália), a segunda na região que contém o maior número de pontos de acesso, a Lazio, e um foco depois na Área Municipal da Cidade de Roma, a área de menores dimensões.

Relativamente aos métodos utilizados nesta fase do projeto, consistiram numa abordagem que seguiu os passos da visualização dos dados espaciais, análise exploratória dos dados espaciais e modelação dos dados espaciais.

#### 4.1 Organização dos Dados Espaciais

Os dados utilizados nesta primeira etapa de análise do padrão pontual correspondem aos 15735 pontos de acesso gratuito à internet, analisados em três etapas – em primeiro lugar utilizando a área total de Itália, depois reduzindo a área de estudo para a região da Lazio, que contém 2185 pontos de acesso e finalmente a Área Metropolitana da Cidade de Roma, que contém 1405 pontos de acesso. Estas áreas foram utilizadas nesta fase por duas razões diversas: a área de Itália permite observar uma visão global do evento em estudo e as duas áreas mais pequenas vão ser utilizadas para estudos mais profundos em capítulos posteriores deste trabalho.

#### 4.2 Visualização dos Dados Espaciais

A etapa da visualização dos dados espaciais consiste na utilização de descritores que permitem uma percepção rápida de como é que estes eventos se distribuem no espaço. Esta descrição é feita utilizando medidas numéricas como a frequência e a densidade, mas também medidas com componente de visualização espacial como as medidas de centro geométrico ou de dispersão espacial.

No que toca à área de estudo escolhida, foi numa primeira fase calculada a densidade para a região e sub-regiões, sendo os dados obtidos visíveis na tabela 2:

Tabela 2 - Medidas das regiões em estudo

Área de Estudo	Área Km <sup>2</sup>	Num Pontos	Densidade (Pontos/Km <sup>2</sup> )
Itália	302073	15735	0.0521
Lazio	17242	2185	0.1267
AM Roma	1285	1405	1.0933

Analisando esta tabela é possível perceber que as zonas escolhidas para uma análise mais aprofundada apresentam densidades pontuais progressivamente maiores, ou seja, existe uma maior concentração de pontos nestas regiões.

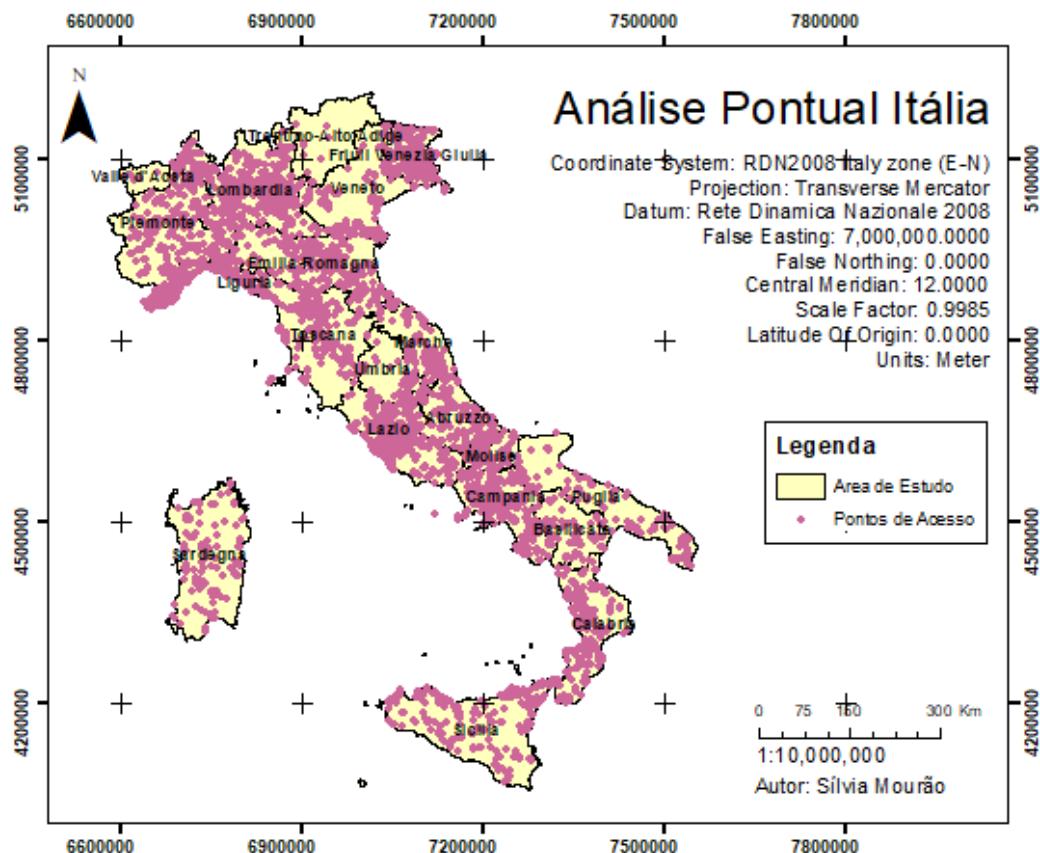


Figura 4 - Distribuição pontos de acesso em Itália

Observando a figura 4 verifica-se que existe uma elevada quantidade de pontos de acesso distribuídos ao longo do território, no entanto existem ainda várias regiões que têm pouca cobertura, como é o caso para Puglia, Trento-Alto Adige e Sardenha.

A nossa região de interesse foi então refinada e o resultado obtido é apresentado na figura 5:

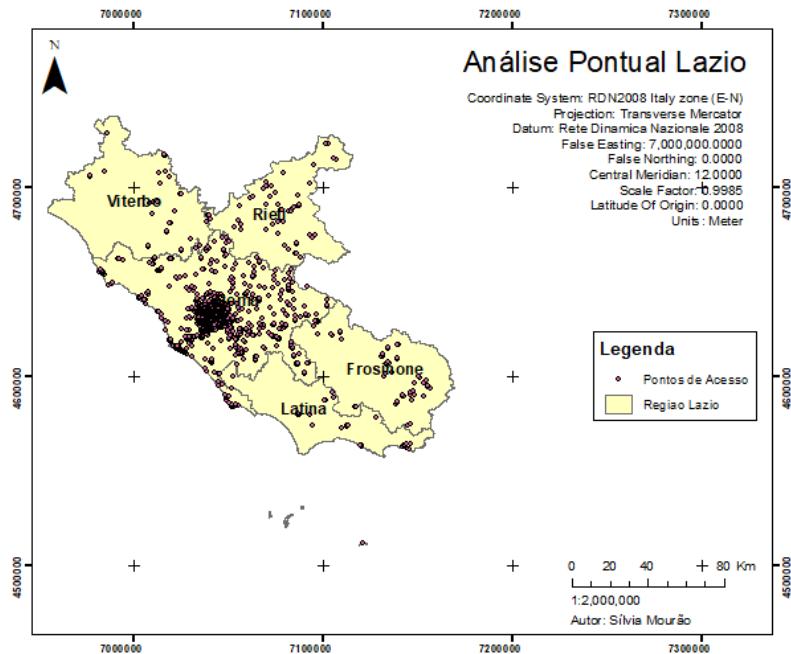


Figura 5 - Distribuição pontos de acesso na Lazio

Para este caso podemos observar que a área apresenta uma distribuição muito menos regular, existindo muito poucos pontos principalmente nas províncias de Latina a sul e Viterbo a norte e estando a grande maioria dos pontos de acesso localizados na zona central da província de Roma, que corresponde à área metropolitana da cidade de Roma. Assim sendo, foi em seguida feita uma redução da área de estudo apenas para esta área, de forma a perceber como é que os pontos se encontravam distribuídos dentro desta, o que pode ser visto na figura 6:

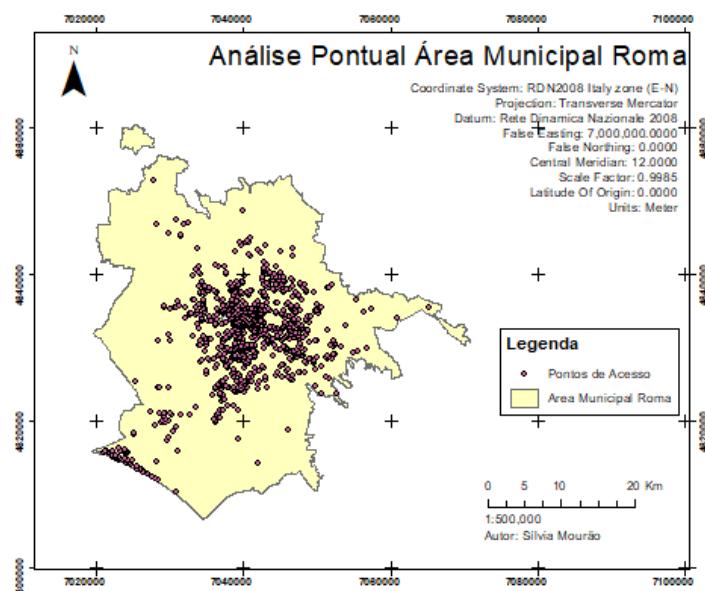
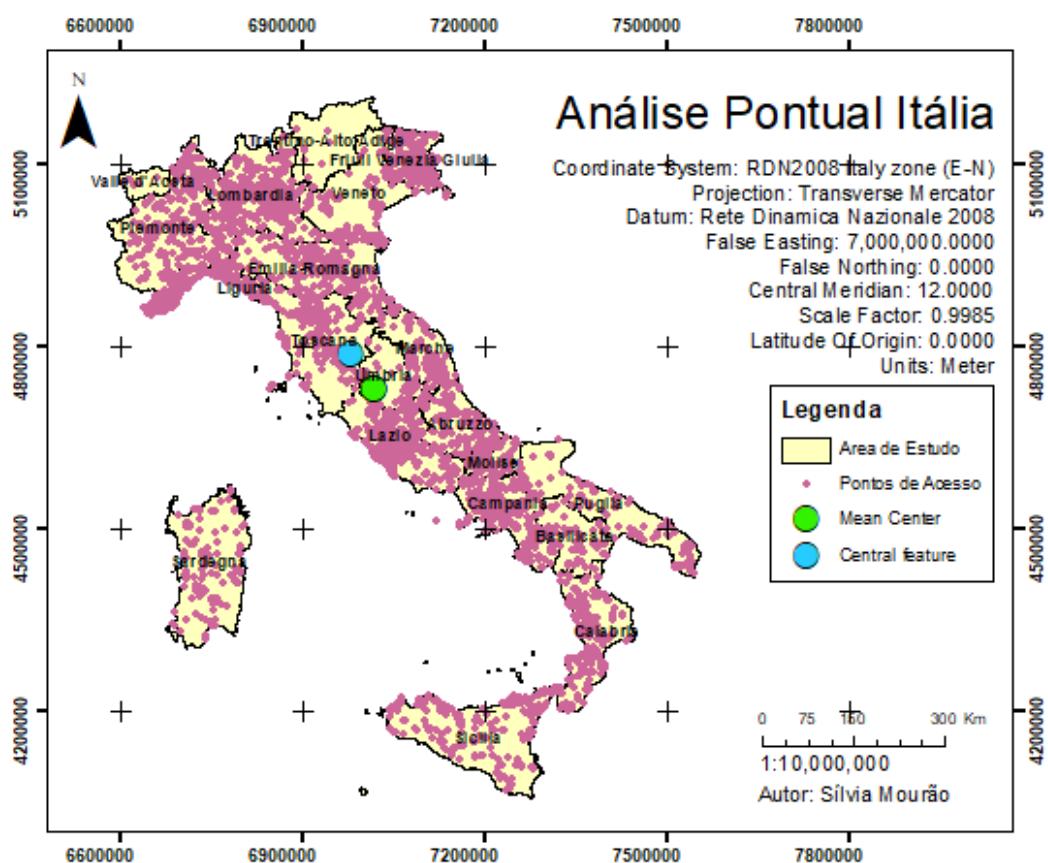


Figura 6 - Distribuição pontos de acesso em Roma

Analisando esta distribuição podemos constatar que existem de facto grandes concentrações de pontos de acesso na zona costeira e na zona central do município, onde se situa a zona histórica da cidade de Roma.

Em seguida, foram realizadas as medidas de centro geométrico e de mediana espacial, que correspondem, respetivamente, às coordenadas médias e medianas em X e Y. O resultado obtido para a área de Itália encontra-se na figura 7:



*Figura 7 - Centro geométrico e mediana espacial da distribuição de pontos de acesso em Itália*

O centro geométrico dos pontos de acesso para a área de Itália situa-se na região da Umbria e a mediana espacial na região da Toscana, ligeiramente a norte do centro do país, devido a uma elevada quantidade de pontos no norte e centro do país e uma menor quantidade de pontos na parte sul do país.

Relativamente às outras áreas de estudo, com resultados visíveis nas figuras 8 e 9, o resultado obtido foi muito semelhante para todas elas, sendo que a amostra de pontos de acesso está a ser fortemente influenciada pela densidade de pontos existente na zona central de Roma.

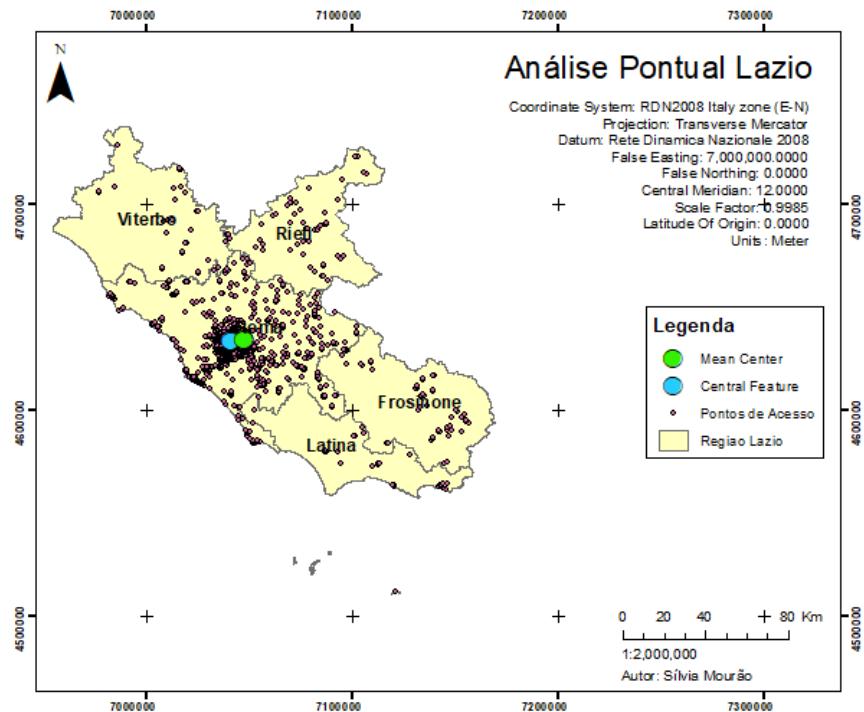


Figura 8 - Centro geométrico e mediana espacial da distribuição de pontos de acesso na Lazio

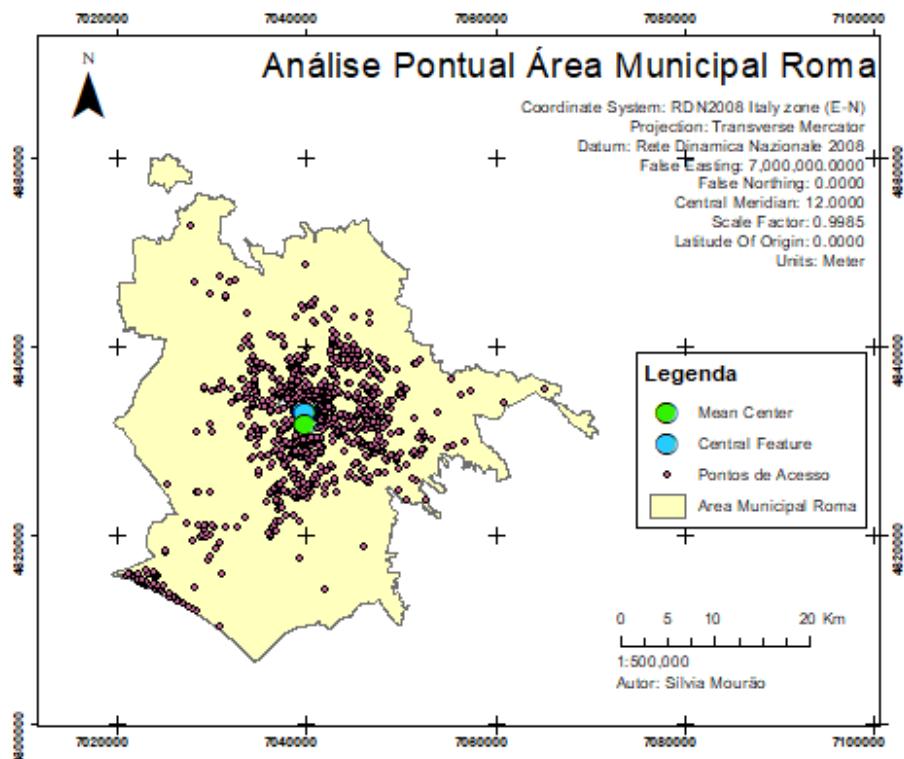


Figura 9 - Centro geométrico e mediana espacial da distribuição de pontos de acesso em Roma

Utilizando o centro geométrico foram também feitas medidas de elipses padra e elipses de dispersão. É relevante referir que foi necessário reduzir a escala de forma a ser possível visualizar os resultados. Na figura 10 temos o resultado obtido para a dispersão espacial enquanto que na figura 11 o resultado obtido para a elipse padrão. Os círculos ou elipses representados a verde têm um raio correspondente a um desvio padrão contêm 68% dos pontos da região enquanto que os círculos a vermelho têm um raio correspondente a dois desvios padrões e contêm 95% dos pontos.

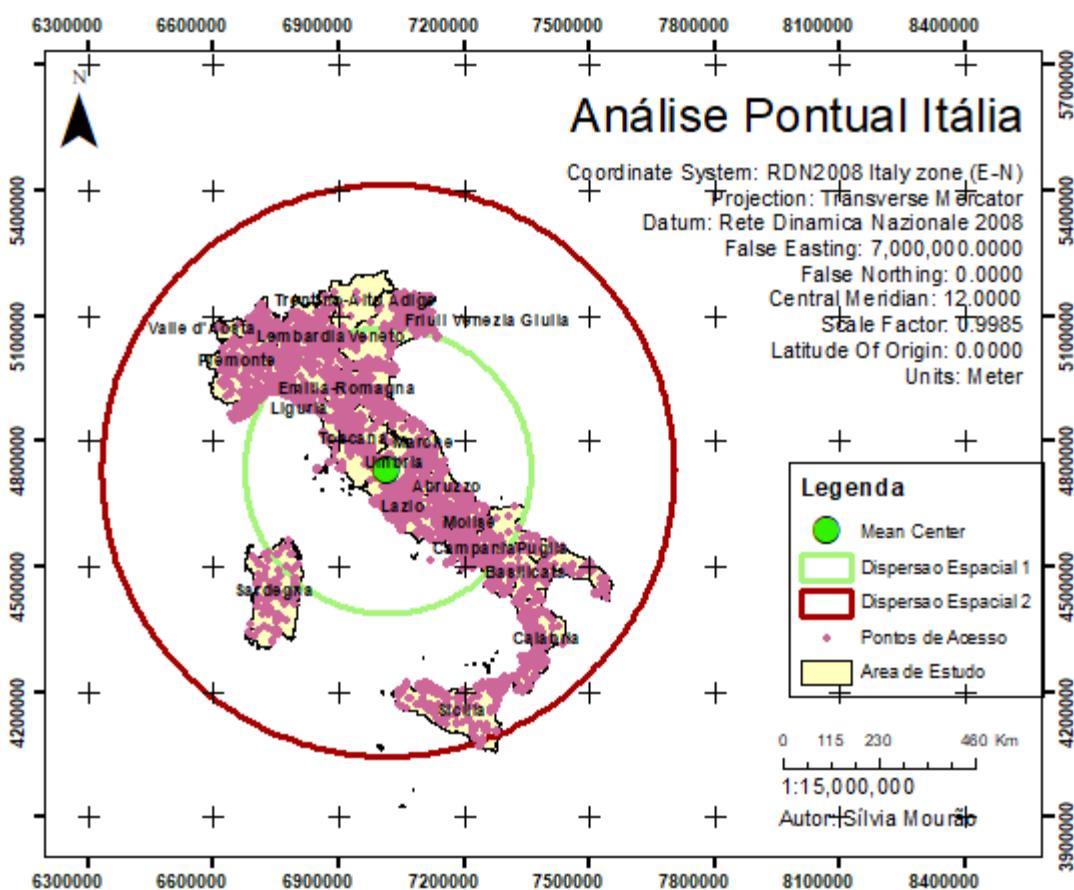


Figura 10 - Dispersão espacial dos pontos de acesso em Itália

Para o caso da elipse padrão, o desvio padrão é calculado em X e em Y, sendo esses valores correspondentes aos eixos maiores e menores da elipse. Sendo Itália um país com uma distribuição espacial com maior variação na componente Y e menor variação na componente X das coordenadas, são obtidas elipses bastante alongadas, que seguem também a ligeira inclinação do país de noroeste para sudeste.

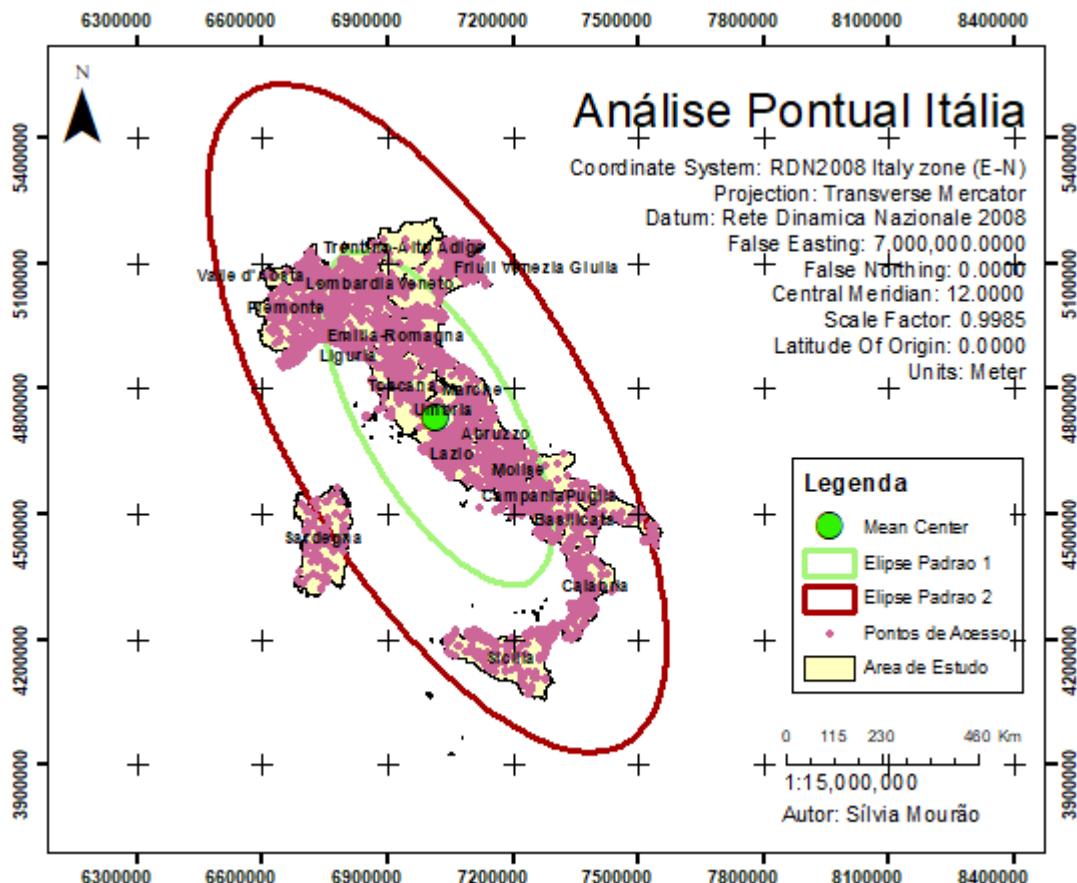


Figura 11 - Elipse padrão dos pontos de acesso em Itália

As seguintes figuras (12-15) mostram os resultados destas medidas para as outras áreas consideradas. Em todos os casos as elipses padrão seguem de forma bastante perceptível as orientações da região, no entanto é de principal relevância o resultado obtido para a região da Lazio, onde as elipses e círculos que correspondem a 95% dos valores abrangem uma menor percentagem da área de estudo quando comparadas com os outros resultados, devido à grande concentração de pontos na zona central e à muito baixa prevalência de pontos de acesso nas regiões exteriores.

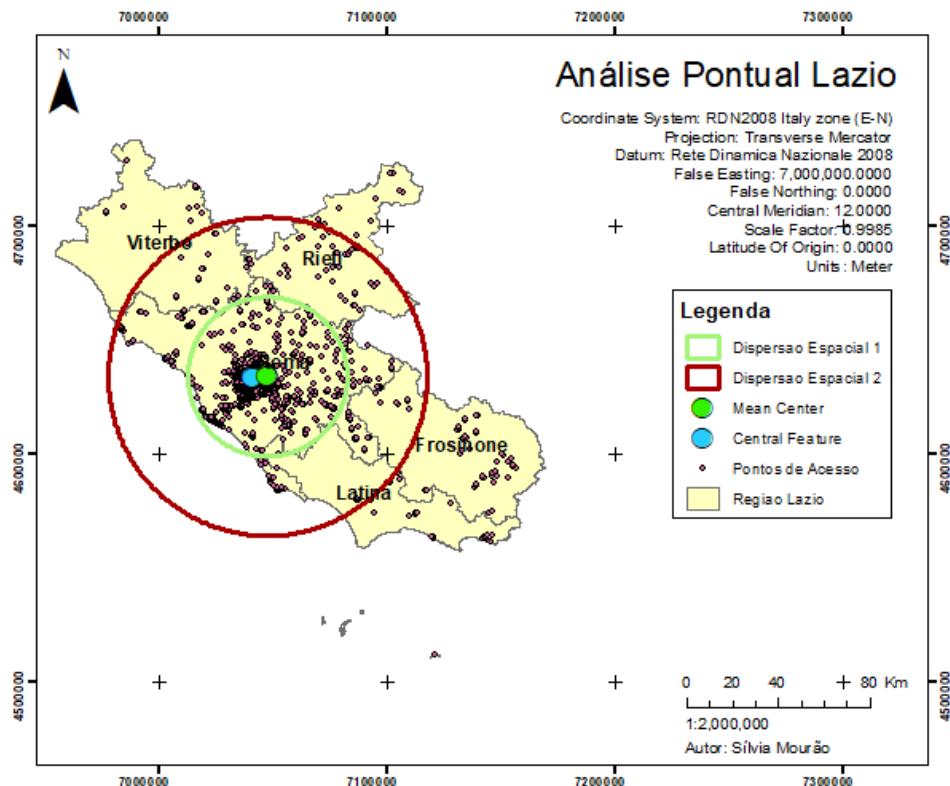


Figura 12 - Dispersão espacial dos pontos de acesso na Lazio

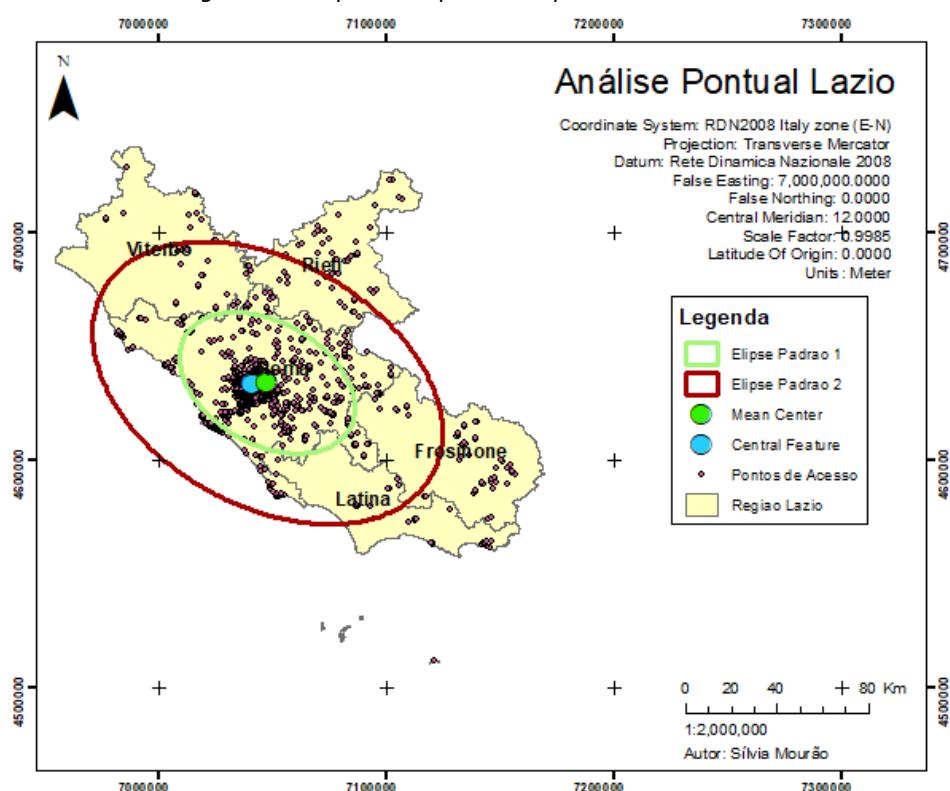


Figura 13 - Elipse padrão dos pontos de acesso na Lazio

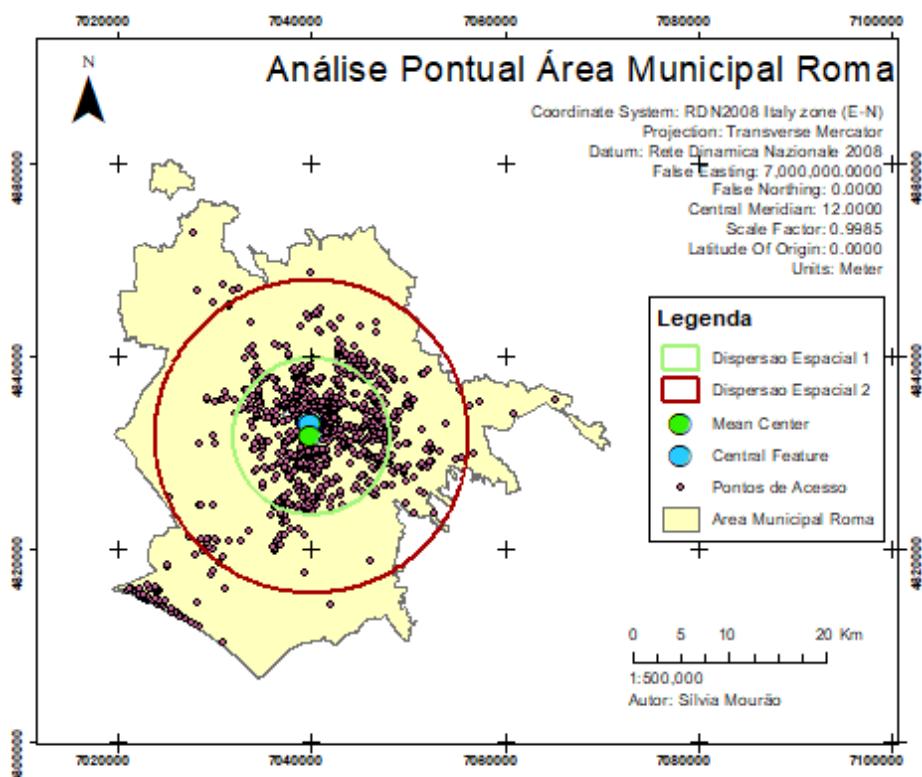


Figura 14 - Dispersão espacial dos pontos de acesso em Roma

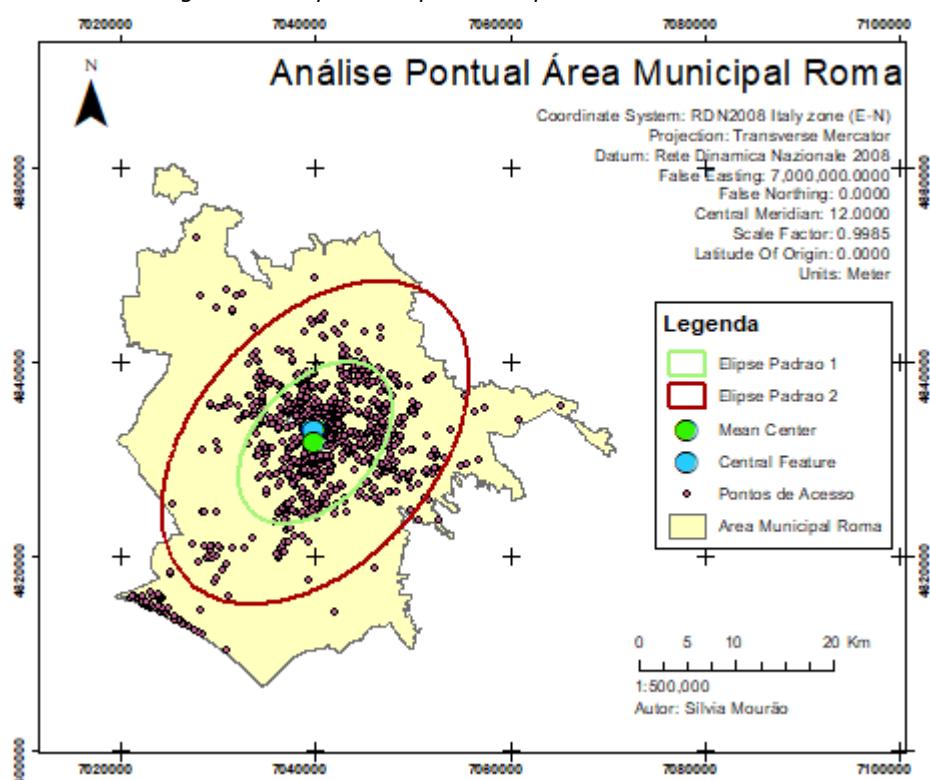


Figura 15 - Elipse padrão dos pontos de acesso em Roma

Sílvia Mourão, FC 57541

Na figura 15 podemos ver que existe uma quantidade relevante de pontos de acesso próximos a zonas costeiras, pelo que a elipse padrão sofre uma ligeira deformação quando comparada com o caso anterior. Embora a segunda elipse padrão não cubra os valores na zona costeira, esses refletem-se na orientação espacial das elipses.

#### 4.3 Análise Exploratória de Dados Espaciais - Técnicas de Análise Exploratória de 1<sup>a</sup> ordem

Após visualizar a distribuição espacial dos dados estes parecem conter zonas com elevada aglomeração de eventos e zonas com muito baixa densidade de eventos. Para avaliar o tipo de padrão que está a ser observado nesta distribuição é então necessário passar à fase de análise exploratória de dados, onde em primeiro lugar vão ser consideradas técnicas de primeira ordem para avaliar efeitos globais que se traduzam em variações no valor médio da densidade do processo, sendo numa fase posterior consideradas as técnicas de segunda ordem para determinar efeitos locais que representam o grau de associação espacial dos eventos.

Começando pelo método da contagem de quadrantes, foi necessário definir um espaçamento de grelha para efetuar a divisão dos dados. Devido à forma como a disposição das grelhas pode influenciar o resultado, foram criadas várias grelhas para cada um dos casos. Foi calculado o valor ideal da área pelo método de censos (indicado nas tabelas pela letra (c)) e foram também considerados valores acima e abaixo desse cálculos. A análise estatística dos resultados destas foi feita para todas as grelhas, estando em seguida representado o exemplo gráfico que corresponde ao caso do valor médio.

Inicialmente foi feita a contagem de quadrantes para Itália inteira, sendo 38 km x 38 km o valor de grelha calculada pelo método de censos. Além desta divisão foram também calculados resultados para grelhas de 20 km x 20 km, 30 km x 30 km e 50 km x 50 km. O resultado da primeira divisão encontra-se na figura 16:

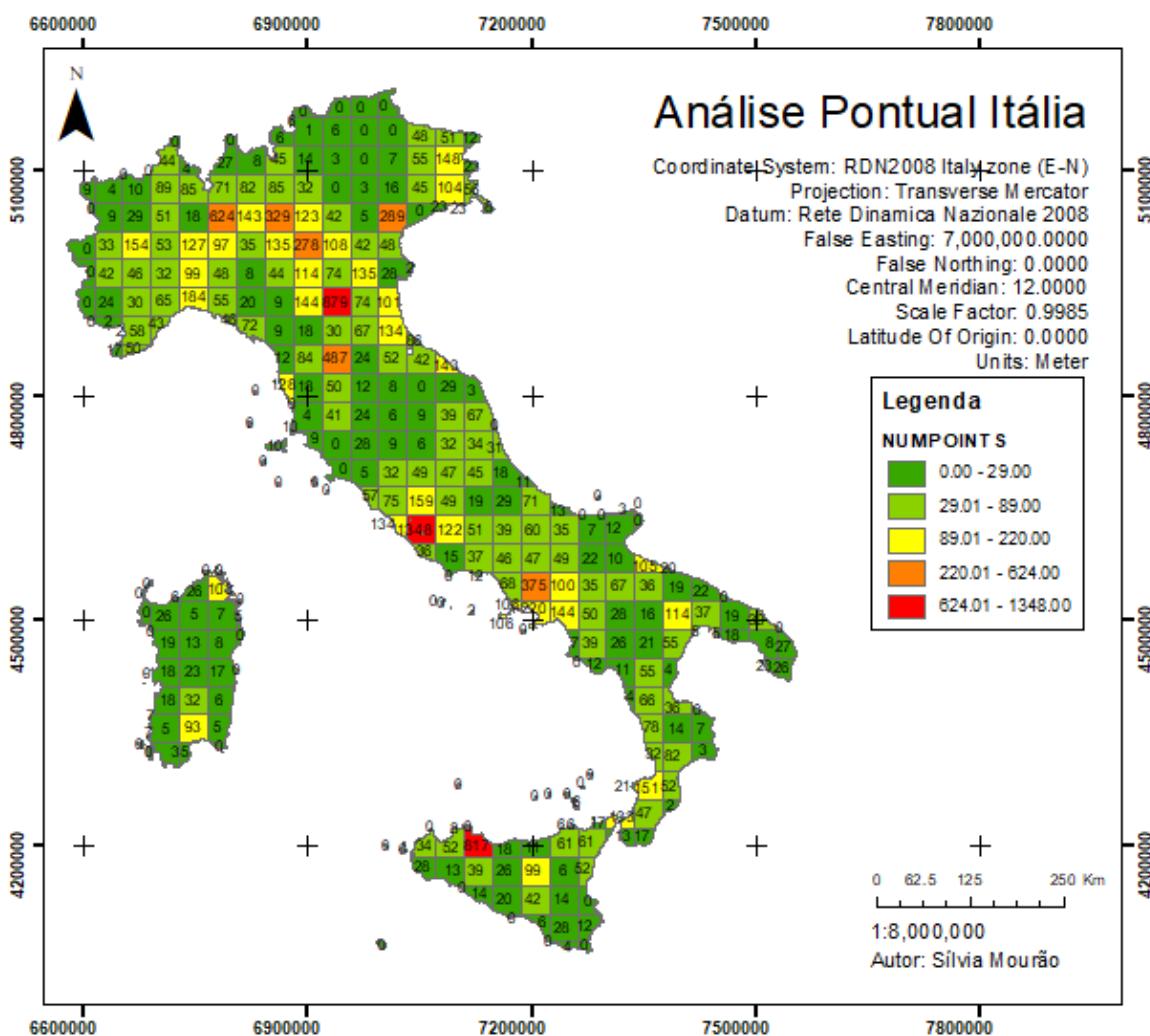


Figura 16 - Método da contagem por quadrantes para Itália com grelha de 38 km x 38 km

Na tabela 3 temos os resultados obtidos para a média, variância e VMTR (variance to mean ratio – rácio entre a variância e a média), sendo que todos os valores de VMTR são bastante superiores a 1 temos uma indicação de que estamos na presença de um padrão aglomerado.

Tabela 3 - Resultados estatísticos do método de contagem por quadrantes para Itália

Variáveis	Grelha			
	20 km x 20 km	30 km x 30 Km	38 km x 38 km (c)	50 km x 50 km
N Quadrantes	987	486	326	211
Pontos	15737	15737	15737	15737
Média	15.94	32.38	48.27	74.58
Variância	2643.88	8205.62	13401.49	24173.83
VMR	165.82	253.41	277.70	324.12

Analisando a imagem da contagem por quadrantes podemos ver que as zonas de maior densidade se encontram na zona de Roma, numa pequena região na zona norte da ilha da Sicília e depois uma maior zona de concentração na zona centro norte, que corresponde às zonas das cidades de Milão e Bologna.

Seguindo a mesma abordagem que na visualização de dados, foi em seguida realizada a análise detalhada da região da Lazio, sendo a grelha calculada pelo método do censos para esta área de 15 km x 15 km (visível na figura 17). Foram também consideradas grelhas de dimensão 20 km x 20 km, 10 km x 10 km e 5 km x 5 km.

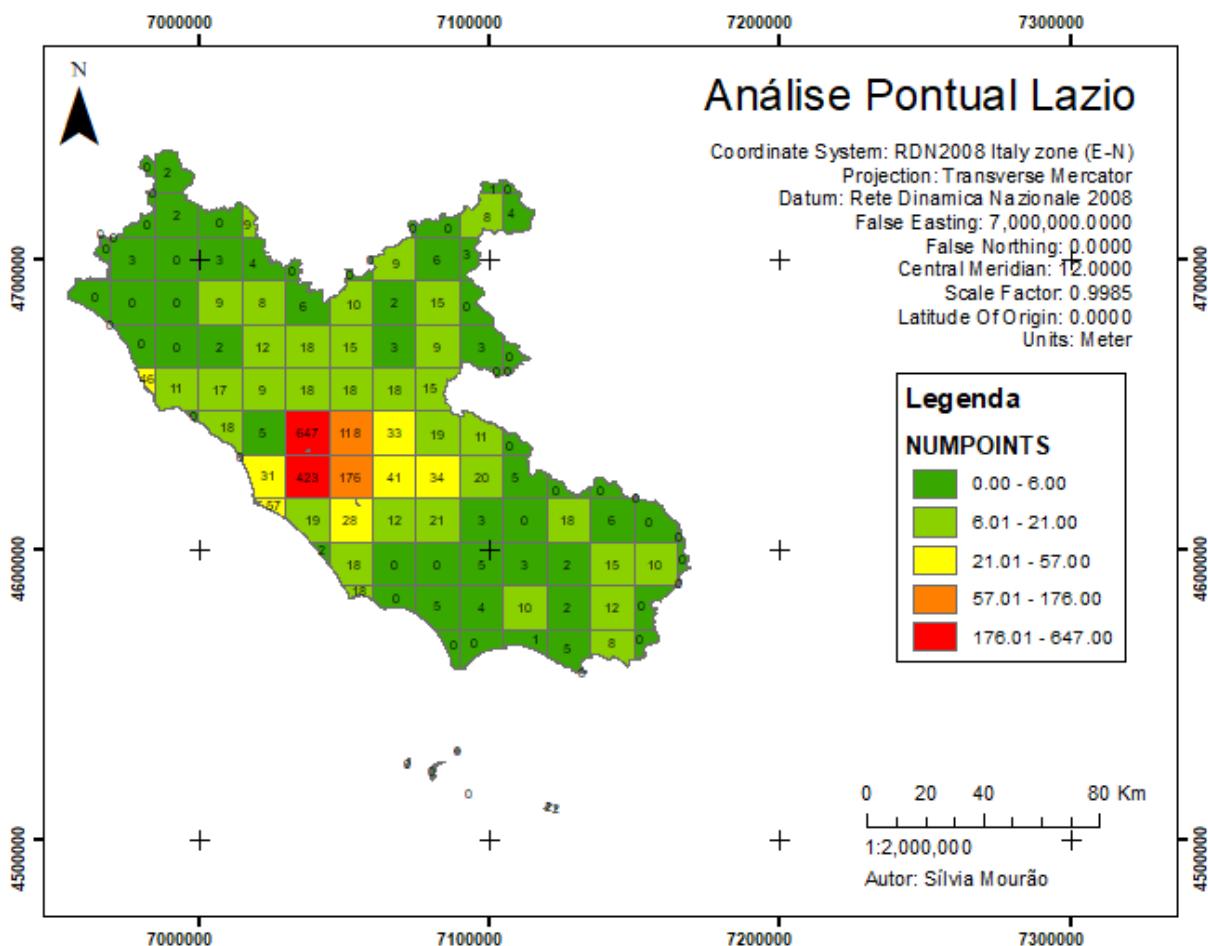


Figura 17 - Método da contagem por quadrantes para a Lazio com grelha de 15 km x 15 km

Na tabela x é possível observar a estatística correspondente a esta região:

Tabela 4 - Resultados estatísticos do método de contagem por quadrantes para a Lazio

Lazio	Grelha			
Variáveis	5 km x 5 km	10 km x 10 km	15 km x 15 km (c)	20 km x 20 km
N Quadrantes	810	228	119	70

<b>Pontos</b>	2185	2185	2185	2185
<b>Média</b>	2.70	9.58	18.36	31.21
<b>Variância</b>	258.37	2697.38	5215.52	17445.25
<b>VMR</b>	95.78	281.47	284.07	558.89

Mais uma vez, todos os valores obtidos são muito superiores a 1, o que indica a presença de um padrão aglomerado. Comparando o resultado obtido com as grelhas de 20 km x 20 km, que são comuns aos dois casos, esta área apresenta maior aglomeração do que a área total de Itália.

Finalmente, foi aplicada a contagem por quadrantes na AM Roma com uma grelha de 2 km x 2 km, o que resultou na seguinte contagem:

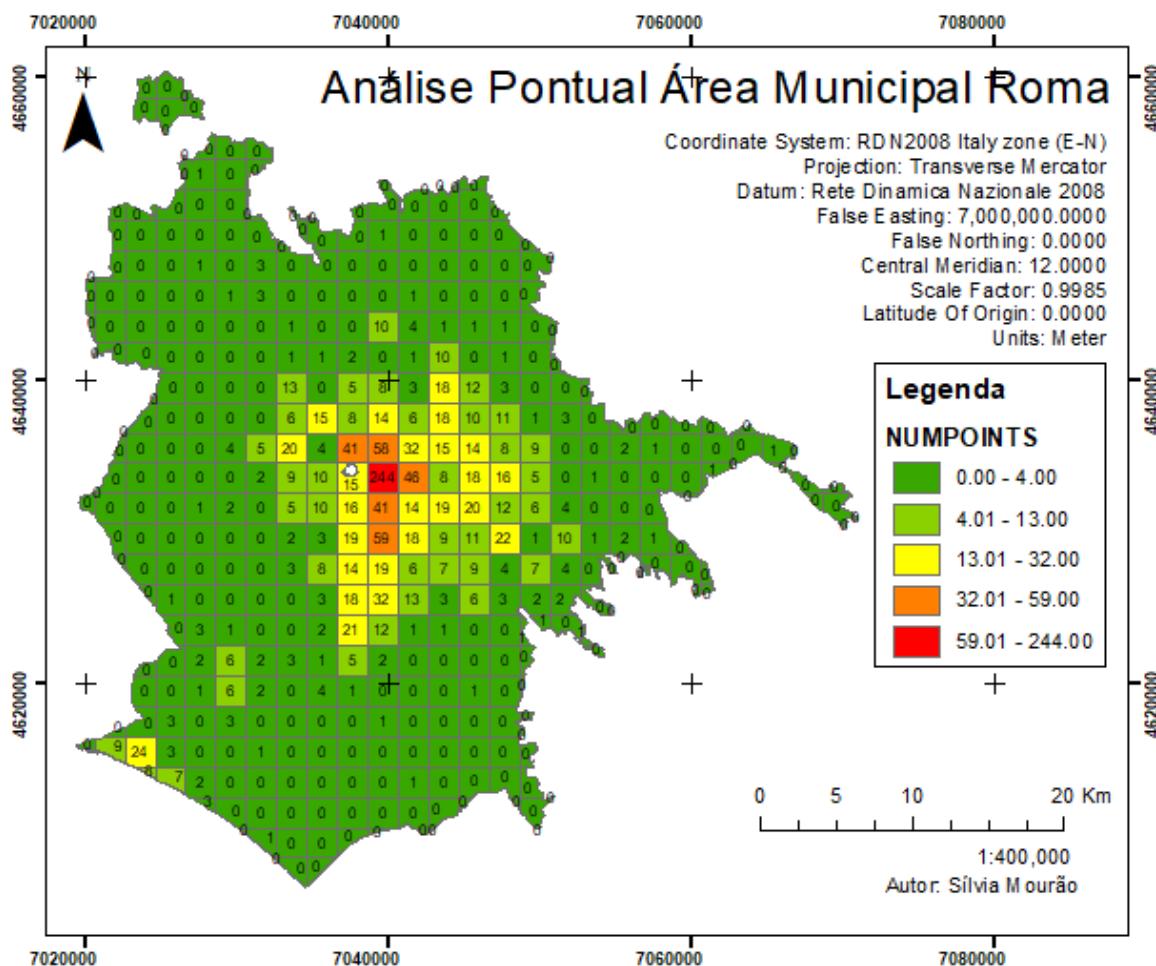


Figura 18 - Método da contagem por quadrantes para Roma com grelha de 2 km x 2 km

Na figura 18 é visível a elevada concentração de pontos na zona central de Roma, sendo a única exceção a zona costeira a Sudoeste.

Os resultados obtidos por esta grelha e outras são visíveis na tabela 5:

*Tabela 5 - Resultados estatísticos do método de contagem por quadrantes para Roma*

AM Roma	Grelha		
Variáveis	1x1 Km	2x2 Km (c)	5x5 Km
N Quadrantes	80	402	1460
Pontos	1405	1405	1405
Média	17.56	3.5	0.96
Variância	2675.25	197.13	21.69
VMR	152.33	56.32	22.53

Mais uma vez, em todos estes casos o valor de VMR indica que estamos na presença de um padrão aglomerado.

Em seguida, foi aplicado o estimador de densidade Kernel para as três áreas de estudo. Este método de análise exploratória é particularmente indicado para detetar hot-spots nos dados e verificar como é que a densidade evolui ao longo da área de estudo e consiste em converter o conjunto dos eventos numa superfície representativa da sua densidade.

A figura 19 mostra o resultado obtido para a área total de Itália.

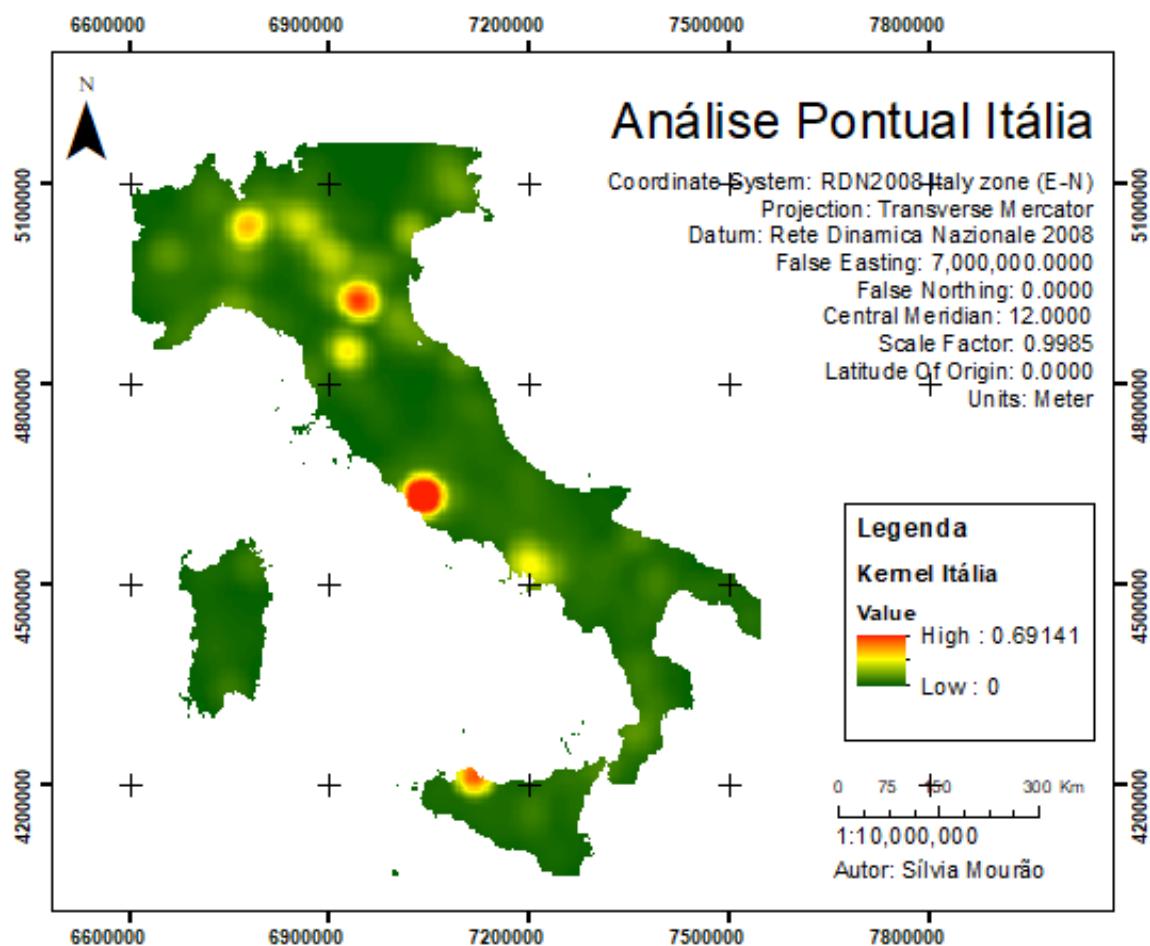


Figura 19 - Estimador de densidade Kernel para Itália

Na imagem é possível ver que existe uma aglomeração de pontos de acesso em certas zonas do país, pelo que pode ser relevante observar se existe alguma relação entre estes pontos e as cidades com maior população no país

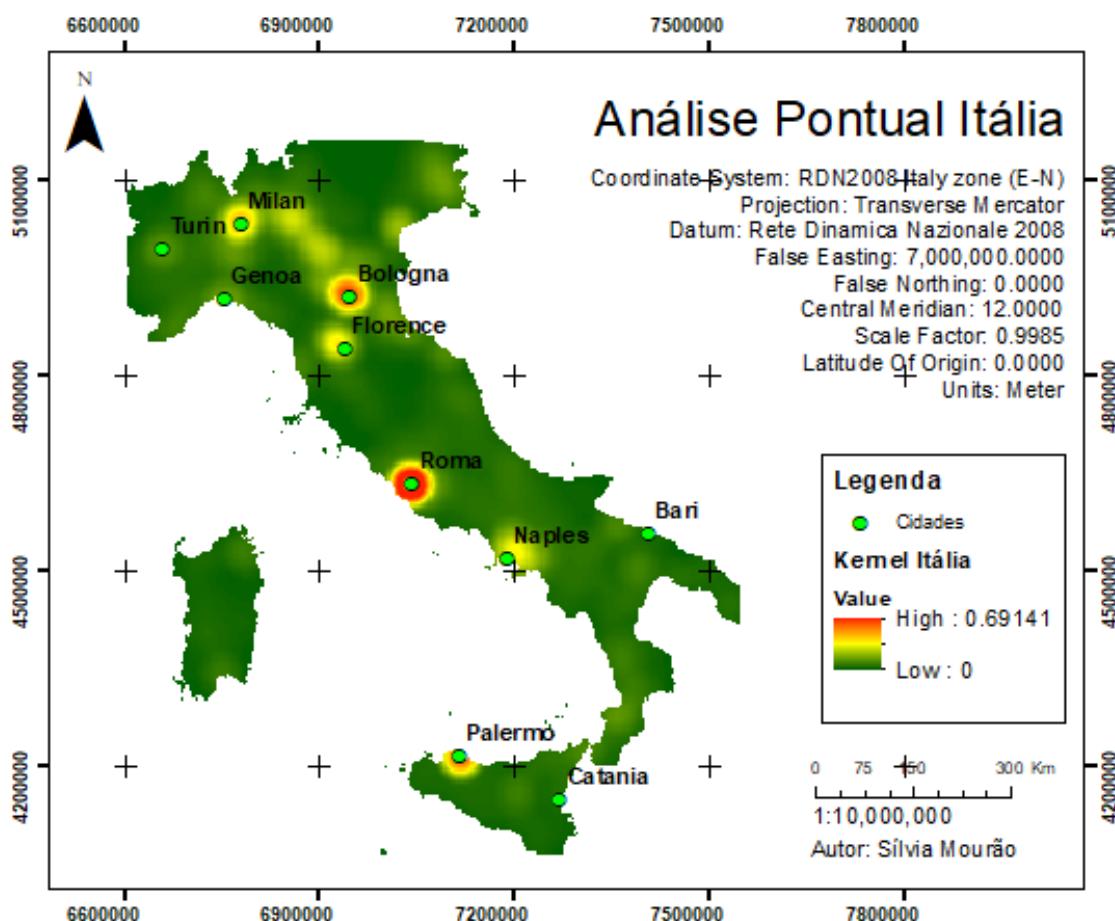


Figura 20 - Localização das dez cidades mais populosas de Itália sobrepostas no estimador de densidade

Na figura 20 estão sobrepostas as dez cidades mais populosas de Itália com o resultado obtido pelo estimador de densidade Kernel. É possível verificar que em muitos destes casos as zonas de maior densidade correspondem a cidades com grande população, mas existem ainda outras zonas com alguma densidade que não correspondem a nenhuma destas cidades.

O mesmo método foi depois aplicado nas outras regiões de estudo, sendo os resultados visíveis na figura 21 e 22.

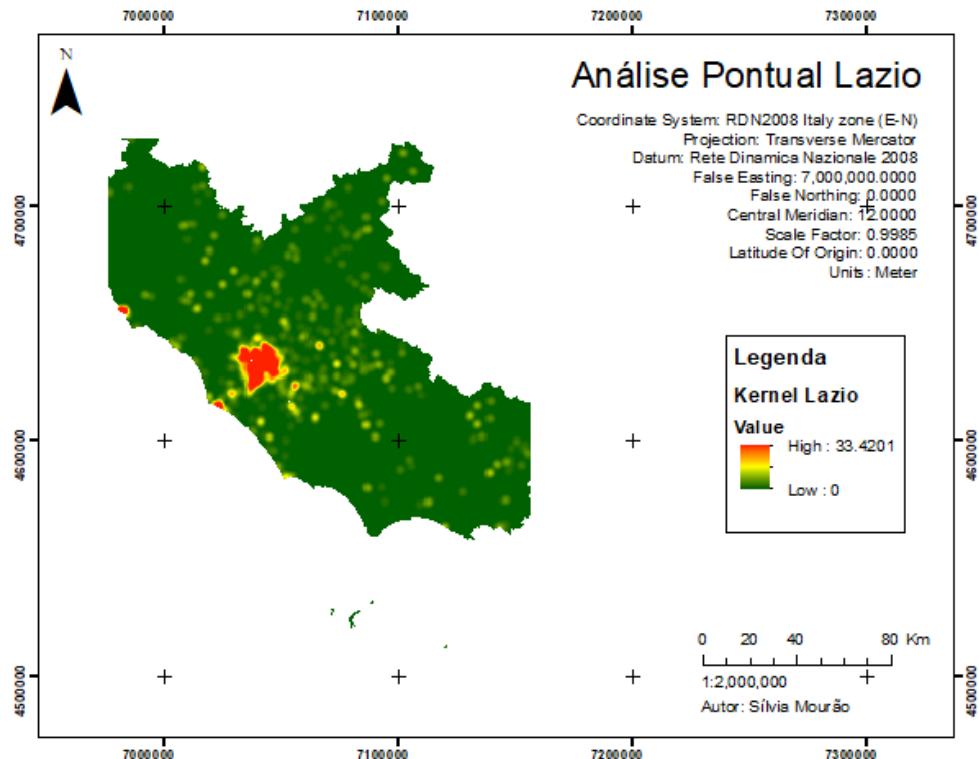


Figura 21 - Estimador de densidade Kernel para a Lazio

Na Região da Lazio, as áreas de maior densidade são dominadas pela zona central de Roma, com algumas zonas mais pequenas junto à costa.

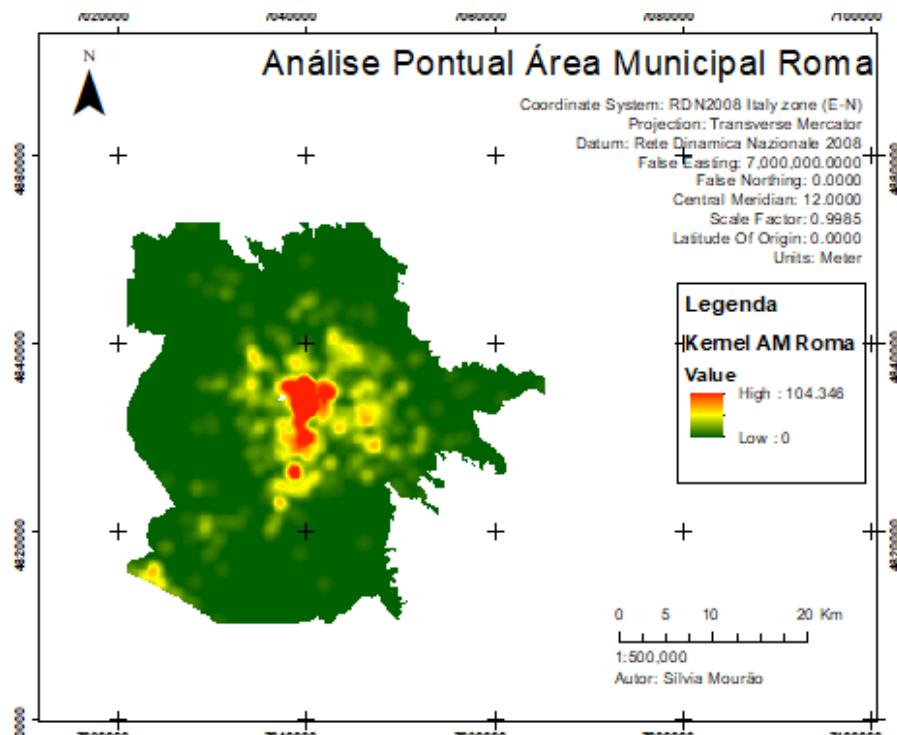


Figura 22 - Estimador de densidade Kernel para Roma

Nesta última figura é visível uma zona de grande densidade na região central, com zonas de valores muito baixos na periferia da área de estudo, com exceção de um pequeno aglomerado na zona costeira mais a oeste.

#### 4.4 Análise Exploratória de Dados Espaciais - Técnicas de Análise Exploratória de 2ª ordem

Depois de visualizar os efeitos de primeira ordem foi também feita uma abordagem para os efeitos de segunda ordem. Estes efeitos são denominados de pequena escala e representam o grau de associação espacial dos eventos. Para este efeito foram utilizados métodos que são baseados na medição de distâncias entre pontos, nomeadamente a função G com o método do vizinho mais próximo e a função de K-Ripley.

O método do vizinho mais próximo procura as distâncias entre cada um dos eventos do padrão de pontos que está a ser analisado, utilizando essas distâncias para estimar a função da distribuição cumulativa. Este teste foi efetuado com recurso ao ArcMap, que imediatamente apresenta os resultados na forma de comparação do resultado obtido com uma distribuição normal dos eventos e efetua a sua classificação relativamente ao tipo de padrão observado. Estes resultados apresentam também a validação estatística do modelo, que será discutida no capítulo x.

Para as áreas de estudo de Itália, Lazio e Roma (figura 23) a Função G apresenta um resultado de padrão aglomerado

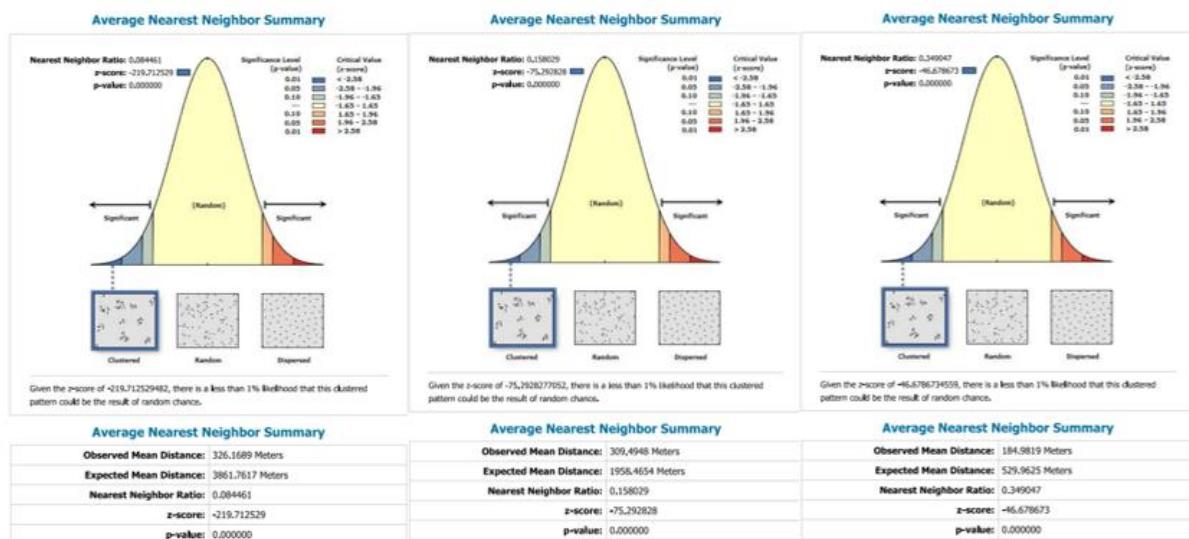


Figura 23 - Resultados da Função G para os casos de Itália (esquerda), Lazio (centro), Roma (direita)

No caso da Função de K-Ripley, esta não se limita a utilizar a distância mínima entre dois eventos como a função G (o que pode ser uma desvantagem para padrões muito aglomerados) mas sim a distância entre todos os eventos. As funções de K-Ripley foram calculadas para cada um dos casos em primeiro lugar com 0 permutações de forma a ser detetado o intervalo de distâncias ideal para depois poder dar significância estatística aos resultados através de permutações, que são computacionalmente mais exigentes. No entanto, devido à elevada quantidade de pontos na primeira amostra não foi possível calcular esta função com permutações para esse caso, correspondendo o gráfico obtido ao caso sem permutações. Para todos os conjuntos de pontos verifica-se a existência de um padrão de clusters até às distâncias de aproximadamente 380 000 m (Itália, figura 24), 100 000 m (Lazio, figura 25) e 22 000 m (Roma, figura 26).

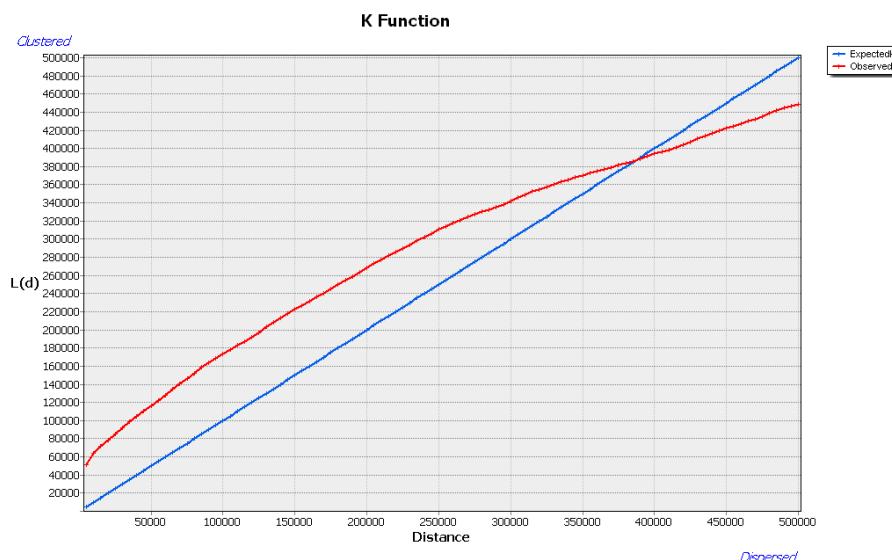


Figura 24 - Resultado Função K-Ripley para Itália

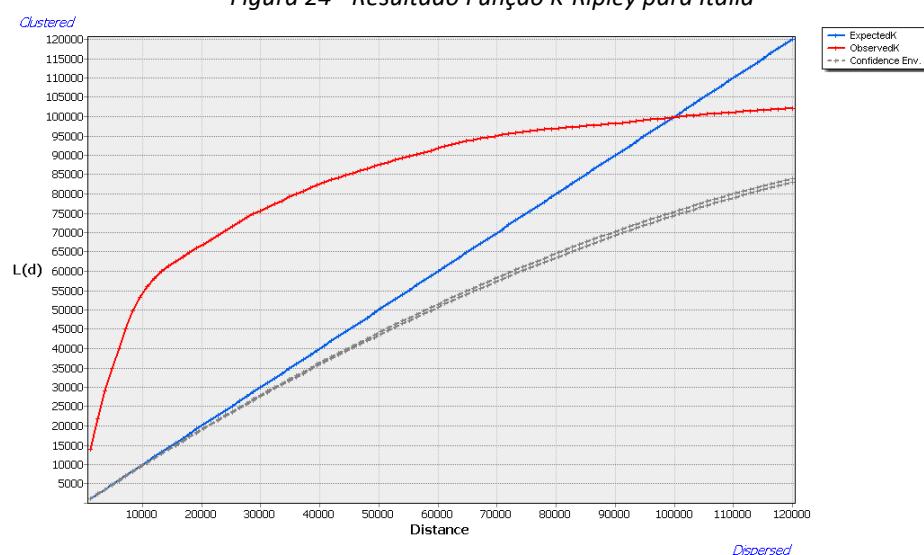


Figura 25 - Resultado Função K-Ripley para a Lazio

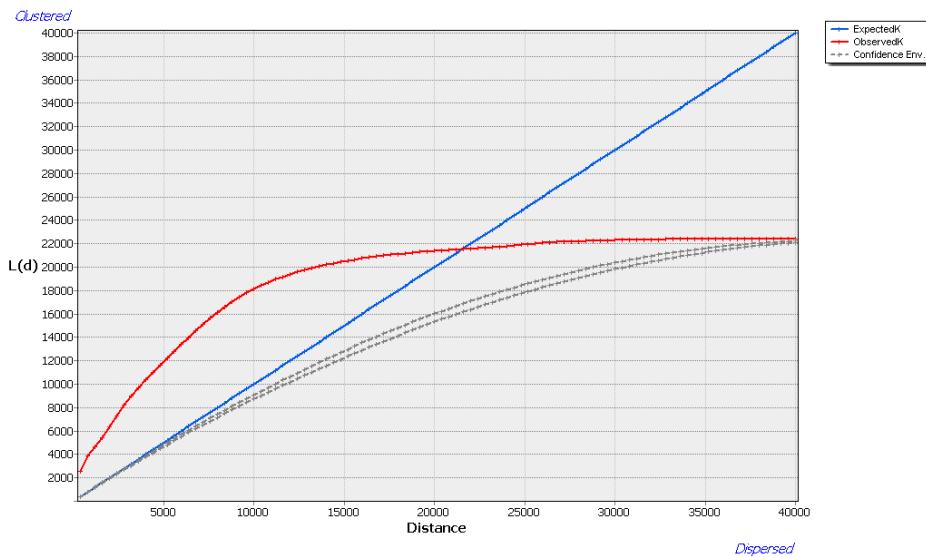


Figura 26 - Resultado Função K-Ripley para Roma

#### 4.5 Modelação do Padrão Pontual

Após realizar a análise exploratória dos dados e verificar o tipo de padrão presente nos dados em estudo é necessário atribuir significância estatística a estes resultados. Isto é feito através da definição de um modelo teórico de referência – CSR (complete spatial randomness – aleatoriedade espacial total) e da verificação pela comparação do padrão observado nos resultados da área de estudo com o modelo teórico de referência.

Esta validação estatística pode ser feita por um de dois métodos: através da comparação com distribuições estatísticas e testes de hipóteses ou através de simulações do processo que resultam em gráficos com envelopes de simulação e valores de pseudo-significância.

Para validar os resultados obtidos pelo método de contagem de quadrantes utilizou-se o modelo teórico de CSR e estabeleceram-se duas hipóteses:  $H_0$  – a amostra em estudo corresponde a um padrão aleatório - e  $H_1$  – a amostra é diferente do padrão aleatório. Para validar estas hipóteses foi utilizada a distribuição de  $\chi^2$ , sendo primeiro encontrados os graus de liberdade de cada uma das amostras e analisados os valores críticos da distribuição de  $\chi^2$  ( $\chi^2$ -high e  $\chi^2$ -low) para um nível de significância de 0.01. Os valores de  $\chi^2$  para as distribuições em estudo, são calculados pelos graus de liberdade e pelo valor de VMR.

Estes valores foram depois comparados com os valores tabulados, correspondendo os resultados obtidos em todos os casos a amostras que não são o resultado de distribuições aleatórias, como pode ser visto nas tabelas 6 (Itália), 7 (Lazio) e 8 (Roma).

*Tabela 6 - Resultados do teste de  $\chi^2$  para o método de contagem por quadrantes para a região de Itália*

Variáveis / Grelhas	20 km x 20 km	30 km x 30 km	38 km x 38 km	50 km x 50 km
Teste $\chi^2$	163498.36	122904.44	90251.90	68065.21
$\chi^2$ high	1092.24	560.38	387.23	260.59
$\chi^2$ low	885.64	415.50	268.65	165.28

*Tabela 7 - Resultados do teste de  $\chi^2$  para o método de contagem por quadrantes para a região da Lazio*

Variáveis / Grelhas	5 km x 5 km	10 km x 10 km	15 km x 15 km	20 km x 20 km
Teste $\chi^2$	77486.61	63892.81	33520.26	38563.19
$\chi^2$ high	905.51	279.49	156.65	99.23
$\chi^2$ low	718.37	180.39	85.23	44.64

*Tabela 8 - Resultados do teste de  $\chi^2$  para o método de contagem por quadrantes para a região de Roma*

Variáveis / Grelhas	1 km x 1 km	2 km x 2 km	5 km x 5 km
Teste $\chi^2$	12033.85	22584.96	32877.77
$\chi^2$ high	111.14	469.81	1587.60
$\chi^2$ low	52.72	338.07	1336.28

Para validar estatisticamente os resultados da Função G utilizou-se o método de permutações, cujo resultado é obtido diretamente do ArcMap e apresentado na tabela 9:

*Tabela 9 - Resultados estatísticos Função G*

	Itália	Lazio	Roma
Z score	-219.71	-75.29	-46.68
P-value	0.00	0.00	0.00

O Z score observado indica que os valores correspondem a um padrão aglomerado (valor negativo de Z), que se encontra z vezes abaixo do valor esperado pela distribuição aleatória. O valor de significância (p-value) é menor do que 0.01 o que significa que existe menos de 1% de hipótese de que estes valores observados correspondam a um padrão aleatório.

Relativamente à Função de K-Ripley, a significância estatística vai depender das permutações conseguidas para cada conjunto de pontos. Como foi visto nos gráficos anteriores (figuras 24, 25, 26 no capítulo 4.4), foi obtido um resultado onde a reta vermelha, que indica a distribuição dos pontos em estudo, está acima da reta azul, que indica a distribuição esperada de um padrão aleatório, o que significa que os pontos estão distribuídos em clusters. No caso de Itália não foi possível obter o

Sílvia Mourão, FC 57541

resultado das permutações pelo que não é possível atribuir significância estatística a este resultado, no entanto para as áreas de Lazio e Roma este resultado foi obtido com nove permutações, o que corresponde a uma significância de 0.1.

Para todos os casos analisados foi possível concluir que a distribuição pontual em análise corresponde para as três áreas consideradas a uma distribuição do tipo aglomerado.

## 5. Análise de Áreas

A análise espacial de áreas tem como objetivo investigar a distribuição de uma variável associada a determinadas unidades geográficas, permitindo a identificação de outliers nos dados. Os dados trabalhados nesta etapa são geralmente dados pontuais generalizados por motivos de confidencialidade individual ou devido à necessidade de existência de uma maior aglomeração dos dados (Catita, 2022).

Um aspeto que é necessário ter em conta quando se faz a generalização da distribuição pontual para uma distribuição por áreas é o problema da unidade de área modificável, que surge devido ao facto de a definição das fronteiras das áreas poder afetar os resultados obtidos. Neste caso de estudo, as áreas correspondem a divisões administrativas utilizadas no país pois o objetivo do estudo é perceber a cobertura por regiões do sinal WiFi.

Nesta fase do projeto vai ser primeiramente feita um visualização dos dados através de mapas temáticos por diferentes métodos de corte da variável, seguindo-se uma análise exploratória da autocorrelação espacial através do índice global de Moran e dos indicadores locais de Moran e Getis-Ord e finalmente uma verificação estatística dos métodos utilizados.

### 5.1 - Organização dos dados em unidades de área

Nesta fase do projeto os dados que tinham sido obtidos na fase de análise pontual foram generalizados para áreas. Foram utilizadas para esta etapa as divisões regionais para a área total de Itália e as divisões por “comune”, a unidade administrativa mais pequena utilizada em Itália, para a região da Lazio. Devido ao facto de a área metropolitana de Roma ser considerada no todo como uma só “comune” e ser também a área que contém um maior número de pontos, fez-se uma subdivisão dessa área em municípios. Apenas a região da Lazio será trabalhada na análise da autocorrelação.

## 5.2 - Visualização dos Dados Espaciais (Métodos de corte da variável)

Com o objetivo de avaliar a autocorrelação espacial dos dados em estudo, foram primeiro criados mapas coropléticos utilizando diferentes métodos de corte de variável, nomeadamente a divisão por quantis, intervalos iguais e desvio padrão. Nenhuma destas divisões é perfeita - a divisão por quantis tem dificuldade em identificar valores extremos, devido à divisão em classes nas quais cada grupo tem o mesmo número de elementos, a divisão por intervalos iguais pode apresentar lacunas quando a distribuição está muito centrada no extremo superior ou inferior dos valores e o desvio padrão apresenta uma distribuição em torno da média, sendo que a média pode não ser uma boa representação dos valores observados.

No entanto, utilizando o conjunto das três abordagens podem ser inferidas algumas informações úteis sobre a área de estudo, que permitem uma compreensão inicial do tipo de dados que está a ser observado.

As figuras 27, 28 e 29 correspondem aos mapas temáticos obtidos para Itália.

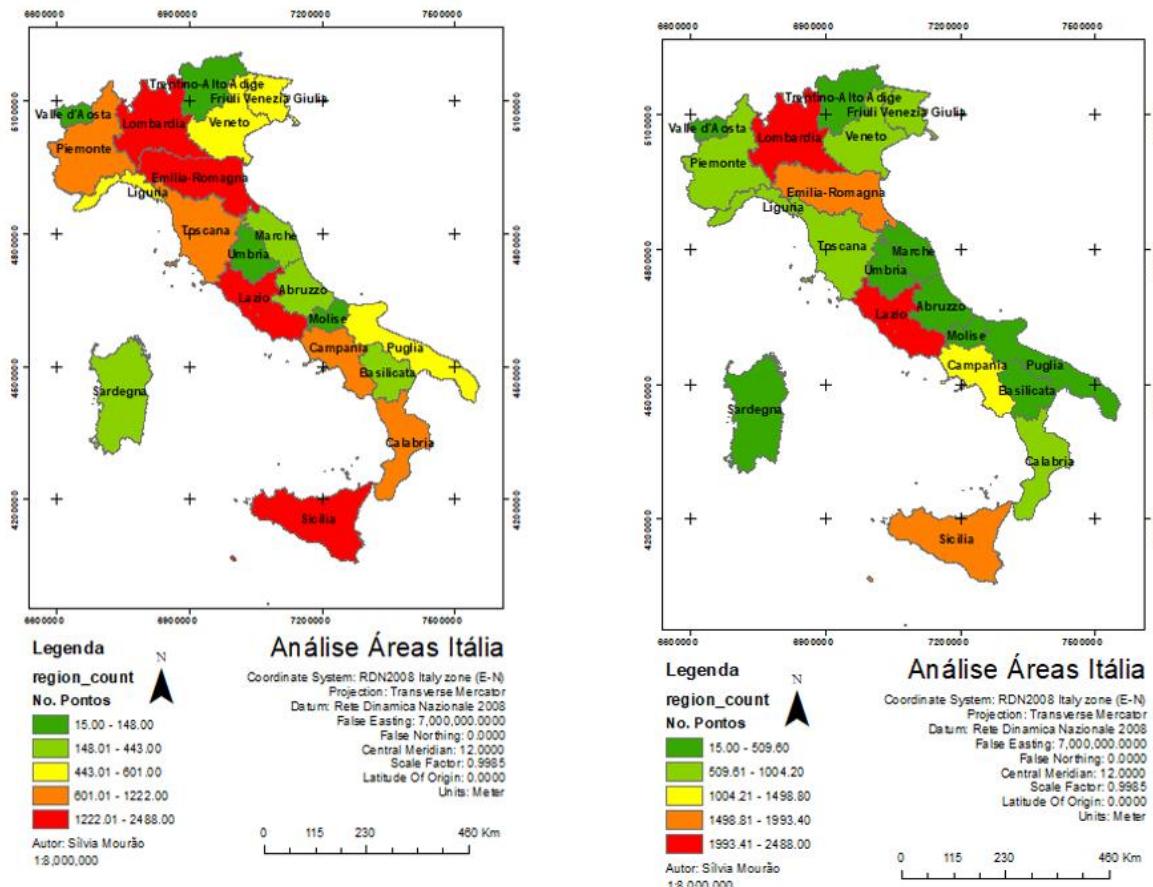


Figura 27 - Mapa temático com divisão por quantis para Itália (esquerda)

Figura 28 - Mapa temático com divisão por intervalos iguais para Itália (direita)

## Análise Áreas Itália

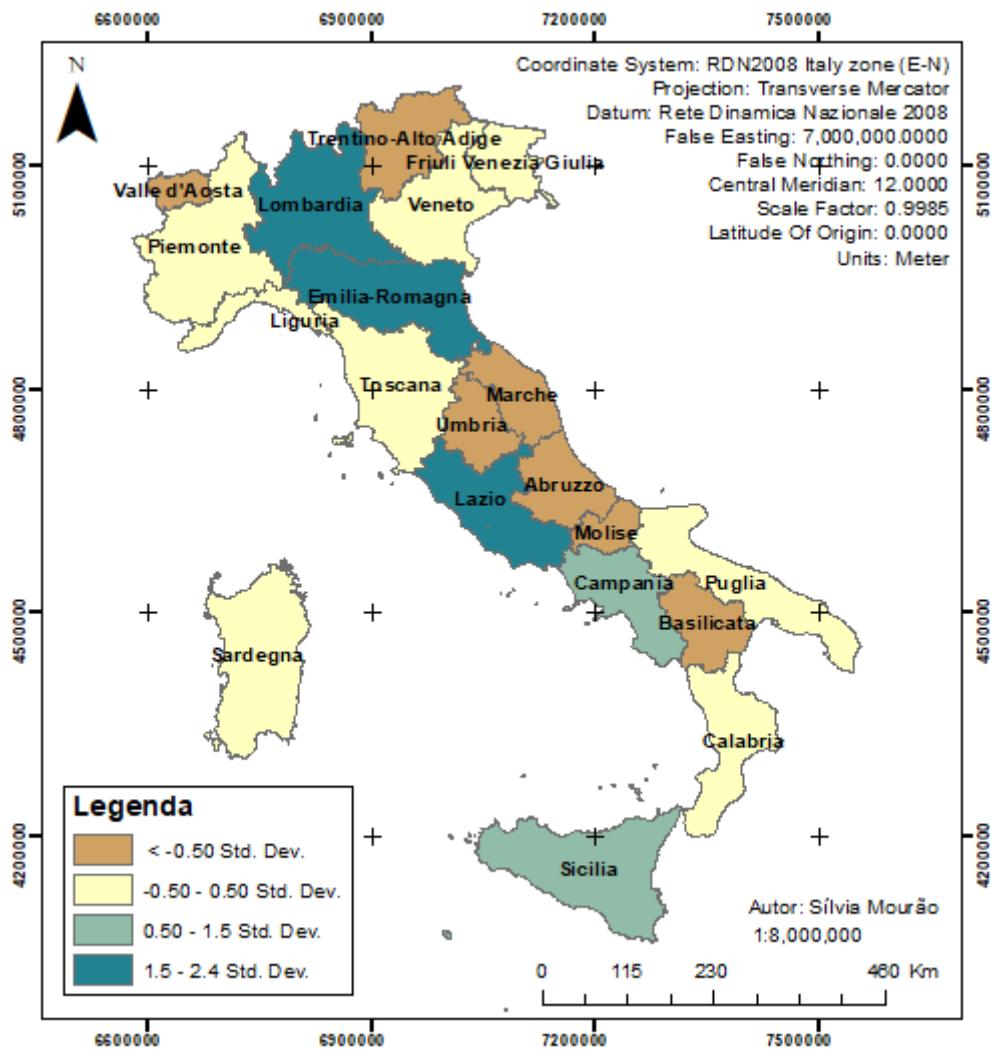


Figura 29 - Mapa temático com divisão por desvios padrões para Itália

Analizando as figuras 27, 28 e 29 é possível verificar que as regiões da Lazio e Lombardia são identificadas como tendo os valores no intervalo mais elevado, enquanto que as regiões de Valle d'Aosta, Trentino-Alto Adige, Umbria e Molise apresentam sempre os valores no intervalo mais baixo em todos os métodos do corte da variável.

Passando depois esta visualização para a região da Lazio obtiveram-se as figuras 30, 31 e 32:

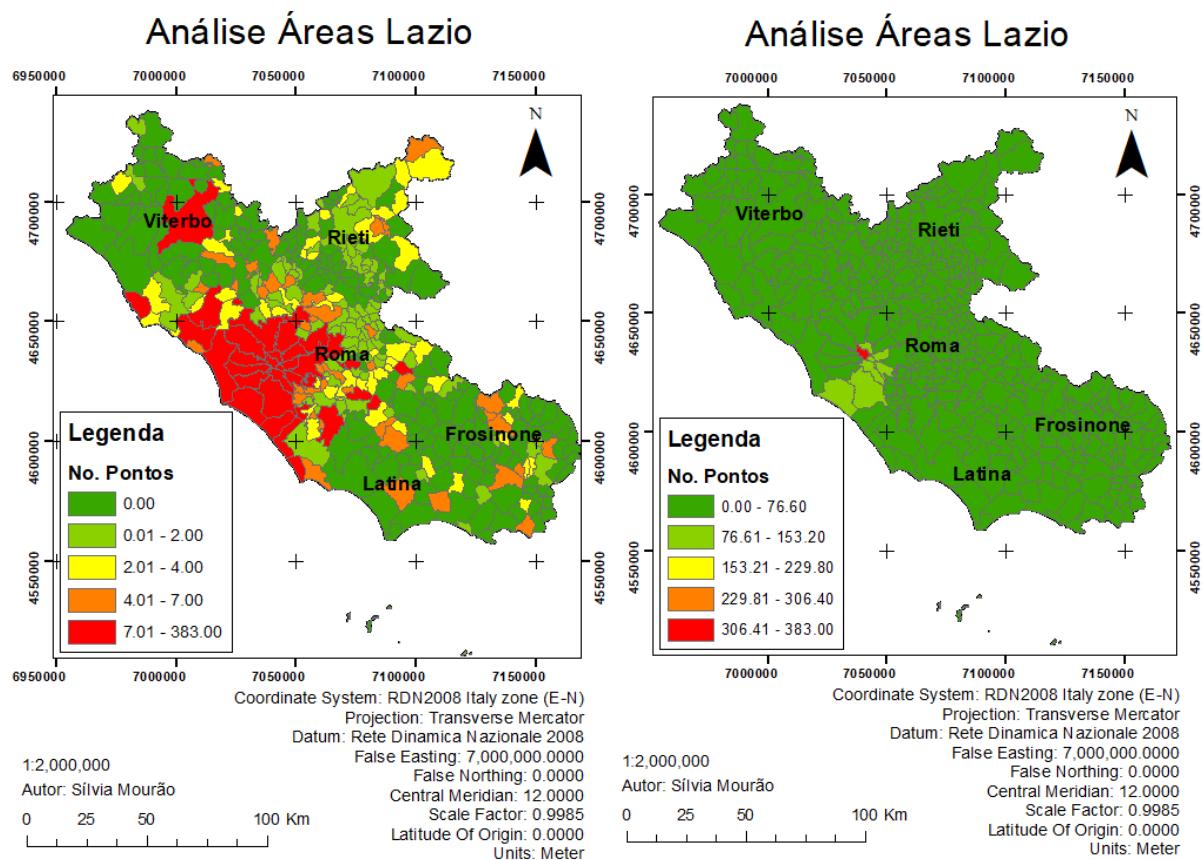


Figura 30 - Mapa temático com divisão por quantis para Lazio (esquerda)

Figura 31 - Mapa temático com divisão por intervalos iguais para Lazio (direita)

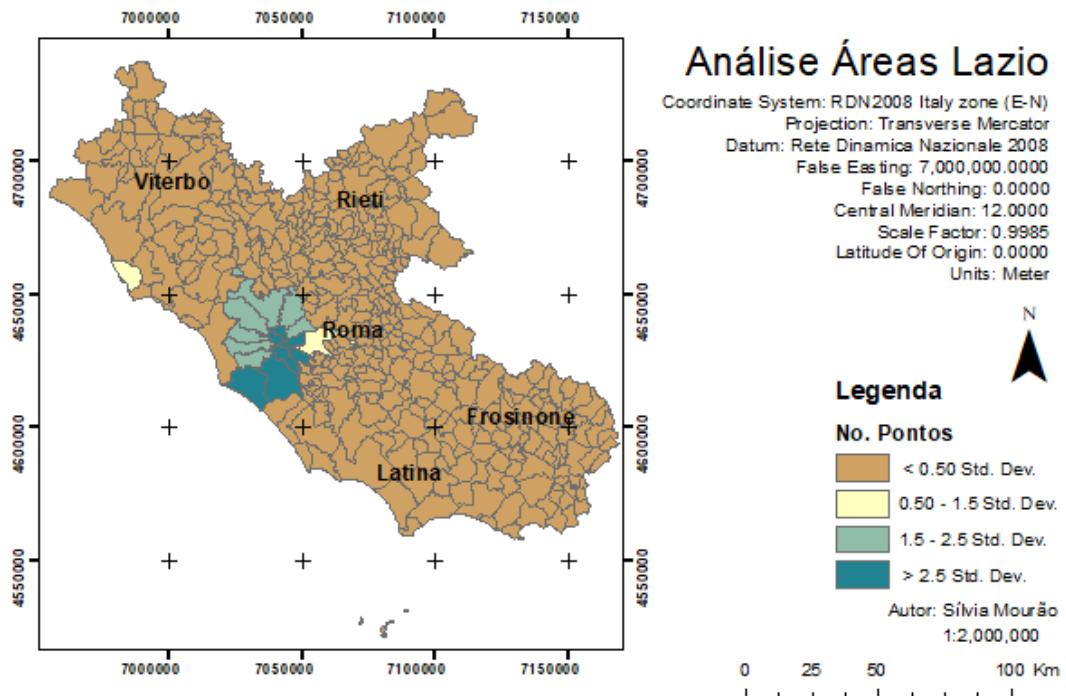


Figura 32 - Mapa temático com divisão por desvios padrões para Lazio

Analizando as figuras 30-32 verifica-se que na região da Lazio é a área da cidade de Roma, principalmente o centro histórico, que contém a grande maioria dos pontos de acesso WiFi, sendo que a grande concentração nestas zonas influencia de tal forma a dispersão dos dados que apenas duas “comune” estão próximas da média (figura 32).

### 5.3 - Análise Exploratória dos Dados

A fase da análise exploratória dos dados no contexto da análise de áreas consiste na medição da autocorrelação espacial, ou seja, a dependência espacial de uma variável numa determinada vizinhança, sendo neste caso medido o grau de influência que uma variável tem sobre si mesma.

Para determinar essa autocorrelação é necessário começar por definir uma matriz de proximidade. Para este projeto foi escolhida uma matriz de proximidade baseada na partilha de fronteiras entre as áreas de estudo, pois embora a região que vai ser tratada nesta fase contenha algumas observações em ilhas estas não serão relevantes para os resultados obtidos. A matriz considera a primeira ordem de vizinhança em todas as direções, também chamada de Queen contiguity. A área de estudo utilizada para esta análise exploratória foi apenas a região da Lazio.

Na figura 33 está representado o histograma de vizinhança obtido com esta matriz, onde as duas áreas sem vizinhos correspondem a duas ilhas.

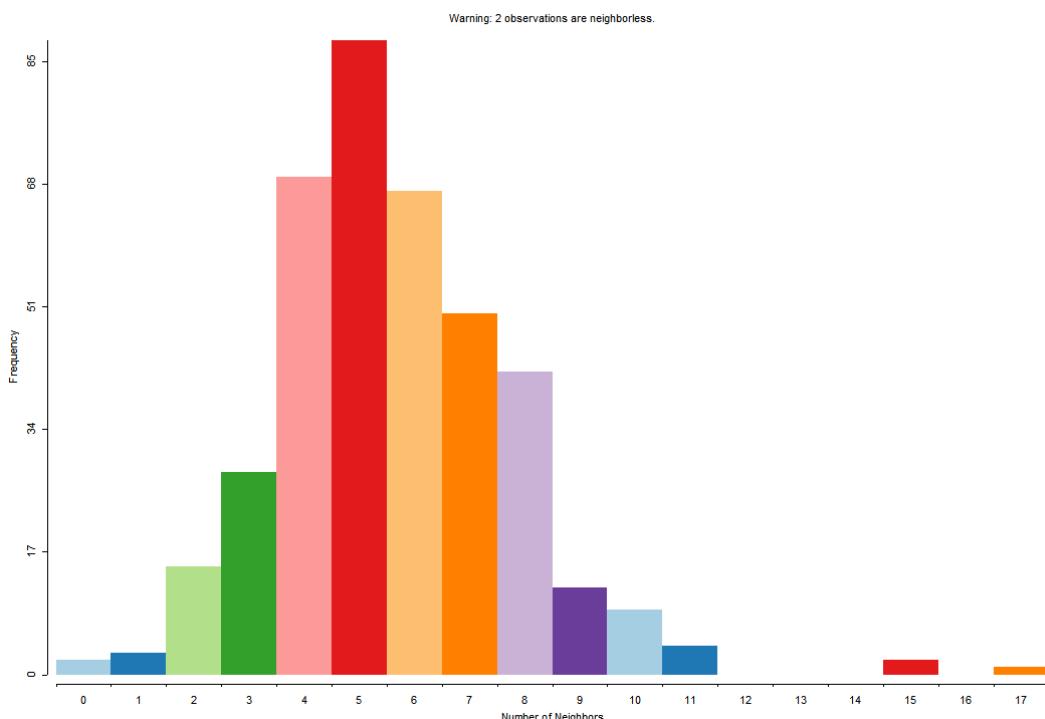
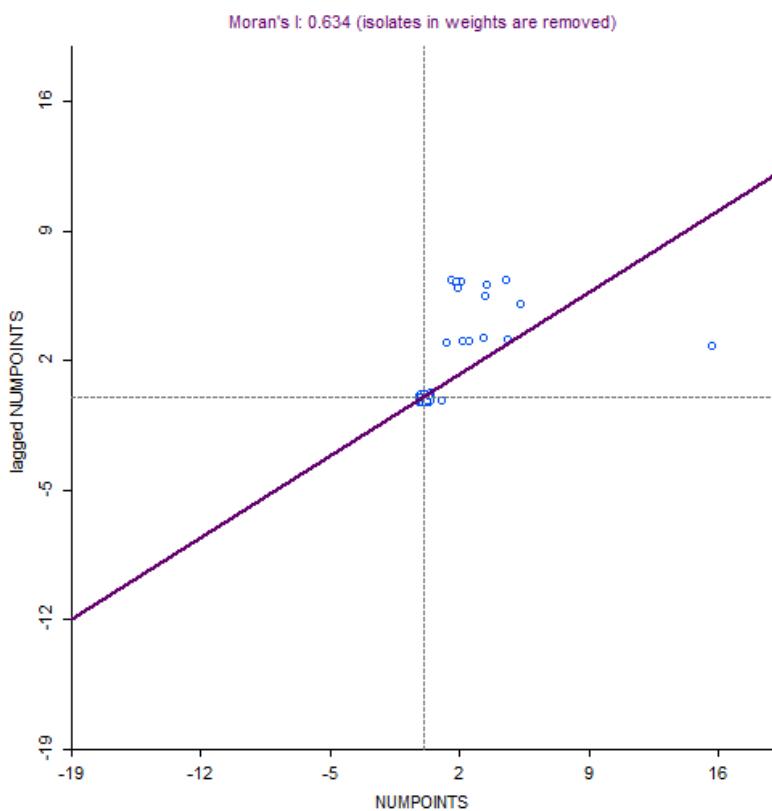


Figura 33 - Histograma de vizinhança

Sílvia Mourão, FC 57541

Em primeiro lugar foi aplicado o índice global de Moran à área de estudo. Os índices globais consideram todas as observações em estudo e têm como objetivo caracterizar a dependência espacial através de uma medida geral de associação – neste caso o índice de Moran. Este índice tem valores compreendidos no intervalo de -1 a 1, correspondendo o valor de -1 a uma autocorrelação espacial negativa, o valor de 1 a uma autocorrelação positiva e 0 a uma ausência de autocorrelação ou aleatoriedade do atributo no espaço.

Para a região em estudo o índice de Moran é de 0.634 o que indica a presença de uma autocorrelação positiva. Observando a figura 34 é possível ver que os valores das observações se encontram muito concentrados junto da origem, com algumas observações no quadrante superior direito (Quadrante I), que correspondem a observações de valores altos cercados por valores altos.



*Figura 34 - Scatterplot para o índice de Moran*

Em seguida, foram aplicados os índices locais de autocorrelação espacial (LISA). Estes índices medem a autocorrelação espacial local para cada área, permitindo identificar se existem agrupamentos ou locais de valores extremos que não são identificados pelo índice global.

Em primeiro lugar foi considerado o índice Local de Moran, uma medida que utiliza a covariância entre áreas para identificar zonas onde existam aglomerações de valores elevados (high-high), de valores baixos (low-low) e zonas de transição entre estes regimes espaciais.

Na figura 35 é visível o resultado da aplicação deste índice à área de estudo.

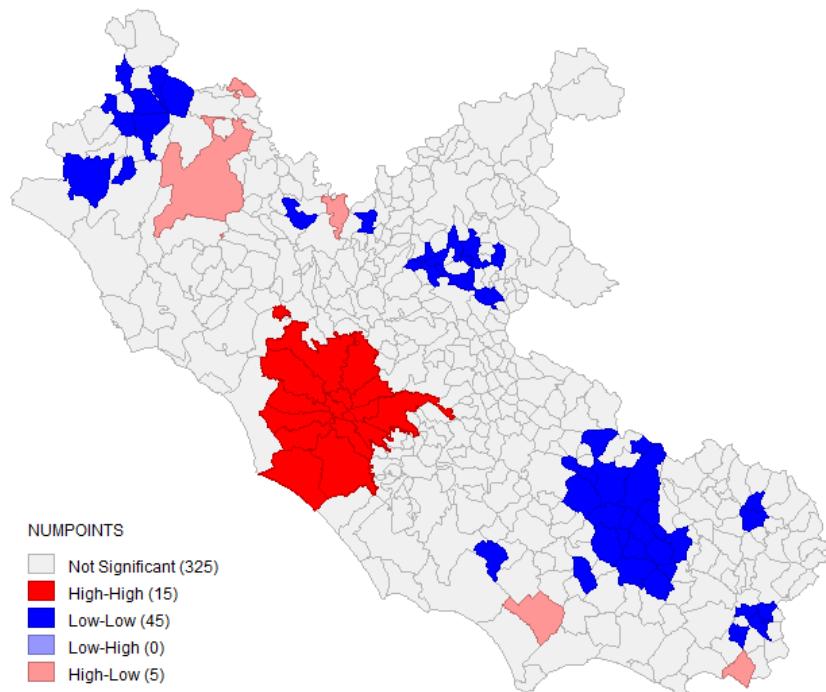


Figura 35 - Índice de Moran local para a Lazio

Analizando a figura 35 é possível constatar a existência de regiões onde estão presentes aglomerados do tipo high-high na zona que corresponde à área metropolitana de Roma e aglomerados do tipo low-low nas zonas norte e sul da região. Este índice produz também um mapa de significância, que será discutido no capítulo 5.4.

Outro método de índice local de autocorrelação espacial que é utilizado é a estatística de Getis-Ord, que não utiliza a matriz de covariância, mas sim a soma de valores vizinhos definidos pela matriz de vizinhança. Podem ser considerados os casos que excluem a observação central da matriz de vizinhança ( $G_i$ ) ou que incluem a observação central da soma ( $G_i^*$ ). Valores altos obtidos por esta estatística indicam a existência de índices altos da ocorrência do fenómeno (aglomerados do tipo high-high), enquanto que os valores baixos correspondem ao agrupamento de valores baixos.

A figura 36 apresenta o resultado a aplicação da estatística  $G_i$  e a figura 37 apresenta o resultado da estatística  $G_i^*$ :

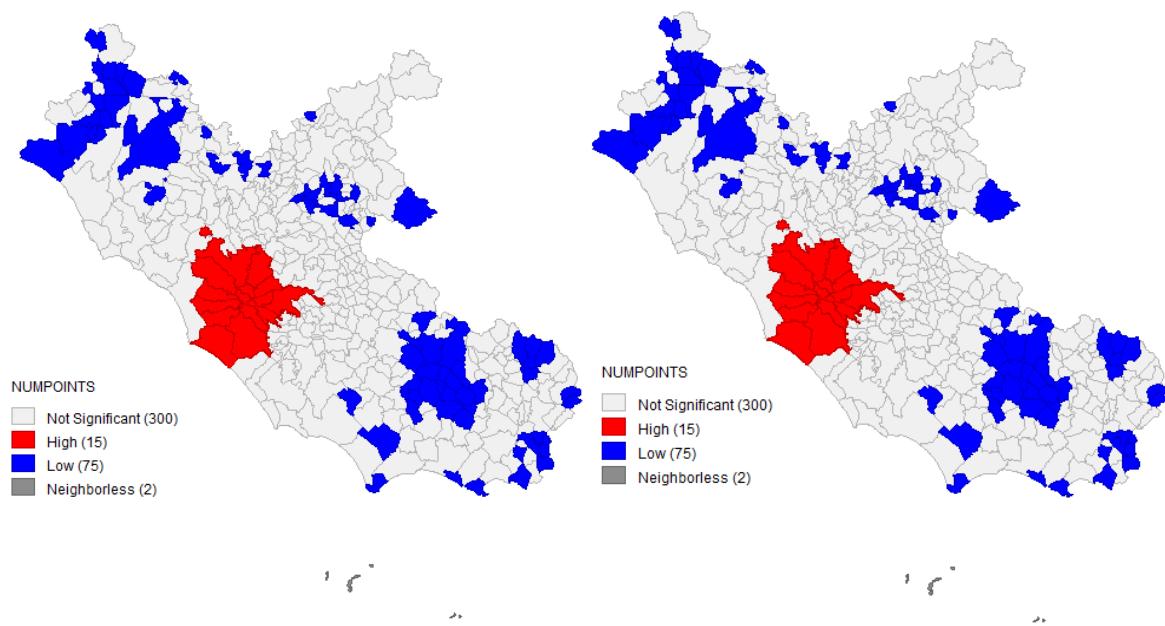


Figura 36 -Estatística Gi para a Lazio (esquerda)

Figura 37 - Estatística  $G^*$  para a Lazio (direita)

Os resultados obtidos pelas duas estatísticas são semelhantes. Mais uma vez, as regiões com autocorrelação espacial elevada encontram-se na zona do centro de Roma, com valores baixos nas zonas próximas dos limites da região a norte, este e sul.

#### 5.4 – Modelação dos dados

Os dados obtidos pelo na etapa de análise exploratória utilizaram permutações como base para tirar conclusões sobre a existência de autocorrelação espacial e por isso têm um nível de significância associado.

No caso do índice de Moran global foi feito um teste de pseudo-significância utilizando 999 permutações, tendo sido obtido um valor de significância de 0.001 e um z-score de 26.81 desvios padrões afastado da média da distribuição aleatória, como é visível na figura 38:

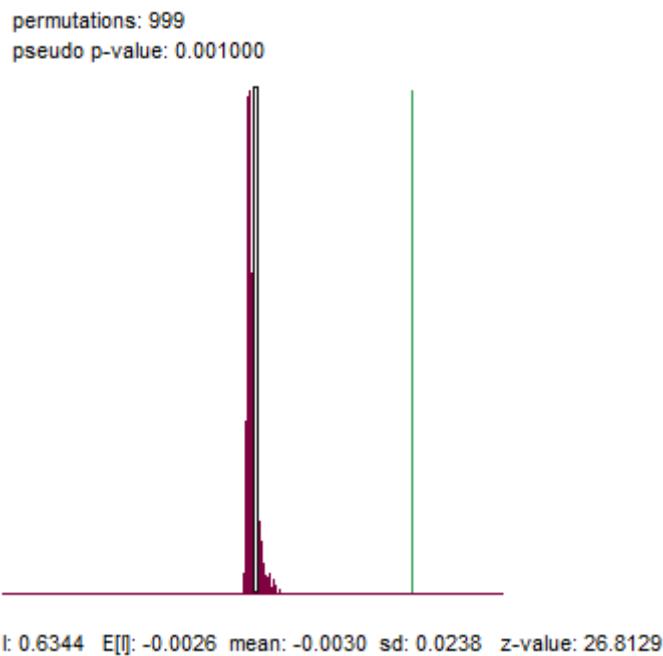
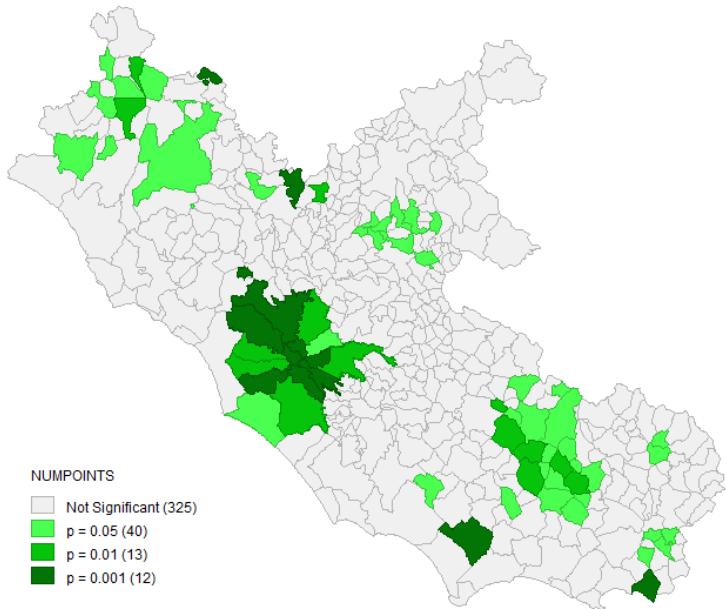


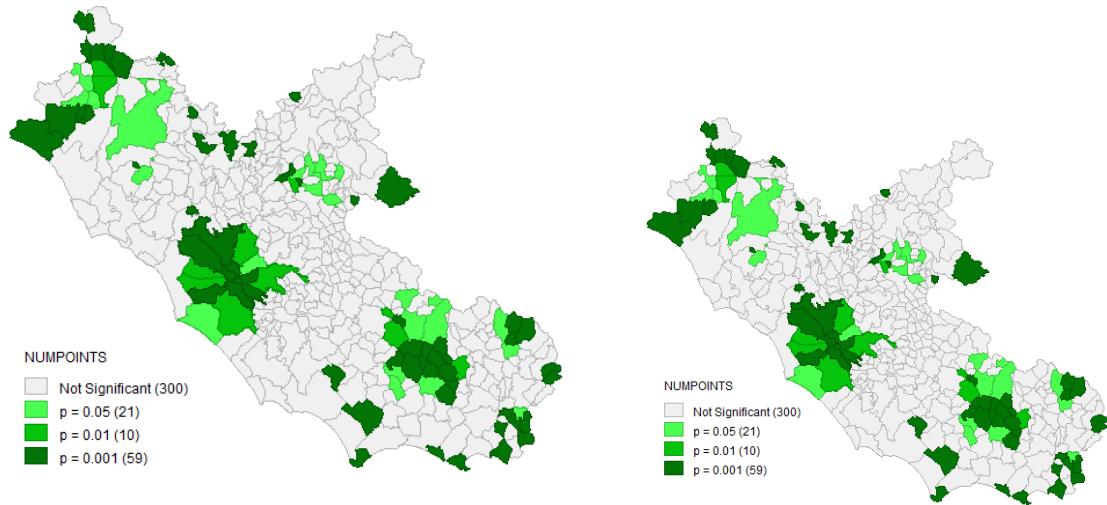
Figura 38 - Teste de pseudo-significância ao índice global de Moran

Para os casos dos índices locais os resultados são obtidos na forma de mapas de significância. Estes mapas são obtidos pela diferença do valor observado ao valor esperado estimado por 999 permutações de distribuições aleatórias, sendo que os valores com menor significância correspondem às regiões que mais afastadas se encontram da distribuição aleatória (valores muito maiores ou muito menores do que a distribuição aleatória) enquanto que as zonas não significantes correspondem a resultados mais próximos da distribuição aleatória (*Local Spatial Autocorrelation (1), n.d.; Project 4: Local Indicators of Spatial Association / GEOG 586: Geographic Information Analysis, n.d.*).

Nas figuras seguintes podem ser observados os resultados para o índice local de Moran (figura 39) e para as estatísticas de Getis-Ord Gi (figura 40) e G\* (figura 41), sendo que as últimas duas têm resultados iguais.



*Figura 39 - Mapa de significância para o índice local de Moran na Lazio*



*Figura 40 - Mapa de significância para a estatística Gi na Lazio (esquerda)*

*Figura 41 - Mapa de significância para a estatística G\* na Lazio (direita)*

Finalizando a etapa de análise de áreas é possível concluir que nesta região existe autocorrelação espacial com significância de 0.001, sendo que existem aglomerados de valores elevados na zona que corresponde à área metropolitana de Roma e zonas de aglomerado de valores baixos nas zonas mais periféricas da região.

## 6 - Métodos de Regressão

O modelo de regressão é uma ferramenta estatística que é utilizada para estudar a relação entre duas ou mais variáveis, de forma a que se possa descrever uma dessas variáveis, chamada de variável dependente, a partir de uma combinação linear das outras variáveis, as variáveis independentes. O objetivo da análise realizada neste capítulo é encontrar um modelo que se ajuste bem à variável em estudo, isto é, a diferença entre os valores estimados e observados é baixa e existe uma elevada correlação medida por  $R^2$  (Catita, 2022). Para modelar a regressão foi escolhida a região da Lazio como área de estudo.

O modelo de regressão linear simples (OLS) parte do pressuposto que as variáveis utilizadas não são colineares e que os resíduos da distribuição não apresentam autocorrelação espacial, têm distribuição aproximadamente normal, com média zero e variância zero. Estes pressupostos vão ser avaliados no capítulo x de forma a analisar se é necessário utilizar um modelo de regressão espacial.

### 6.1- Análise da correlação entre as variáveis independentes e dependente

Com o objetivo de construir um modelo de regressão foi primeiramente necessário procurar estabelecer a relação entre as variáveis e avaliar a linearidade entre essas variáveis. A variável dependente é a variável em estudo neste projeto, ou seja, o número de pontos de acesso gratuito à internet. Como variáveis independentes foram consideradas as seguintes treze variáveis:

- Turismo – Número de pontos de interesse turístico (monumentos, estátuas, museus) por área
- Transportes – Número de pontos de interesse relacionados com transportes (paragens de autocarro, estações de comboio) por área.
- Hotéis – Número de locais de alojamento (hotéis, hostels, airbnb) por área.
- Praias – Distância media dos hotspots numa área à praia mais próxima
- Natureza – Distância média dos hotspots numa área ao ponto de interesse natural mais próximo (picos de montanhas, parques naturais)
- Dist. Turismo - Distância média dos hotspots num polígono ao ponto de interesse turístico mais próximo.
- Comércio - Número de pontos de interesse comercial (supermercados, centros comerciais).
- Dist. Comércio – Distância média dos hotspots num polígono ao ponto de interesse comercial mais próximo.

Sílvia Mourão, FC 57541

- Noturna – Número de pontos de interesse relacionados com diversão noturna (bars, discotecas)
- Dist. Noturna - Distância media dos hotspots num polígono ao ponto de interesse relacionado com diversão noturna
- População – População de cada “*comune*” / município em 2019
- Densidade populacional – população dividida pela área de cada polígono
- Índice Turismo – Índice turístico do governo italiano (1 a 6, sendo 1 o valor com maior quantidade de turismo)

Para avaliar a relação entre estas variáveis foram criados diagramas de dispersão no software GeoDa, sendo os resultados gráficos apresentados na figura 42 e os valores obtidos para  $R^2$  apresentados de seguida na tabela 10:

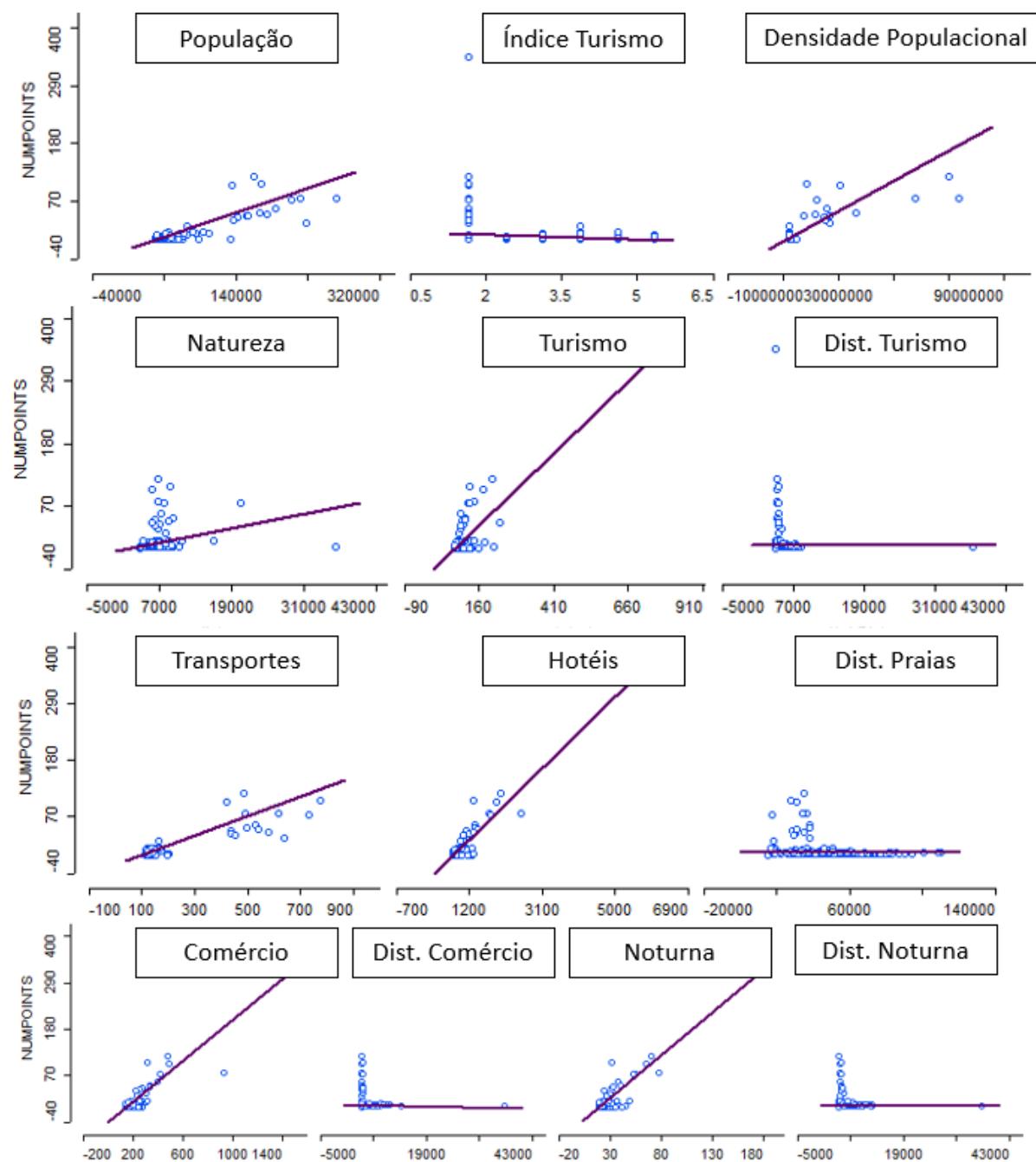


Figura 42 - Resultados dos testes de correlação entre o número de pontos de acesso e as variáveis independentes

*Tabela 10 - Resultados dos coeficientes de determinação obtidos para a variável dependente*

Variável	R <sup>2</sup>
Hotéis	0.903
Comércio	0.825
Noturna	0.808
Turismo	0.742
Densidade	0.675
Transportes	0.563
População	0.462
Natureza	0.048
Índice turismo	0.044
Dist. Praia	0.002
Dist. Turismo	0
Dist. Noite	0
Dist. Comércio	0

Os resultados obtidos mostram a existência de uma correlação elevada para as variáveis hotéis, comércio, diversão noturna, turismo, densidade populacional e transportes. As variáveis distância turismo, distância noturno e distância comércio não apresentam correlação.

## 6.2 - Modelo de Regressão Clássico (OLS) - Simples e Múltipla

Inicialmente efetuou-se uma regressão linear simples, escolhendo apenas uma variável independente. Para fazer esta análise foram utilizadas duas variáveis, Hotéis, que apresenta o maior valor de R<sup>2</sup> e Turismo, pois o objetivo deste projeto é a sua aplicação ao turismo e esta variável apresenta também um valor elevado de R<sup>2</sup>. Todas estas correlações foram realizadas com recurso ao GeoDa.

Em primeiro lugar foi realizada a regressão simples com a variável hoteis. Os resultados obtidos são visíveis na figura 43 e os resíduos do modelo na figura 44:

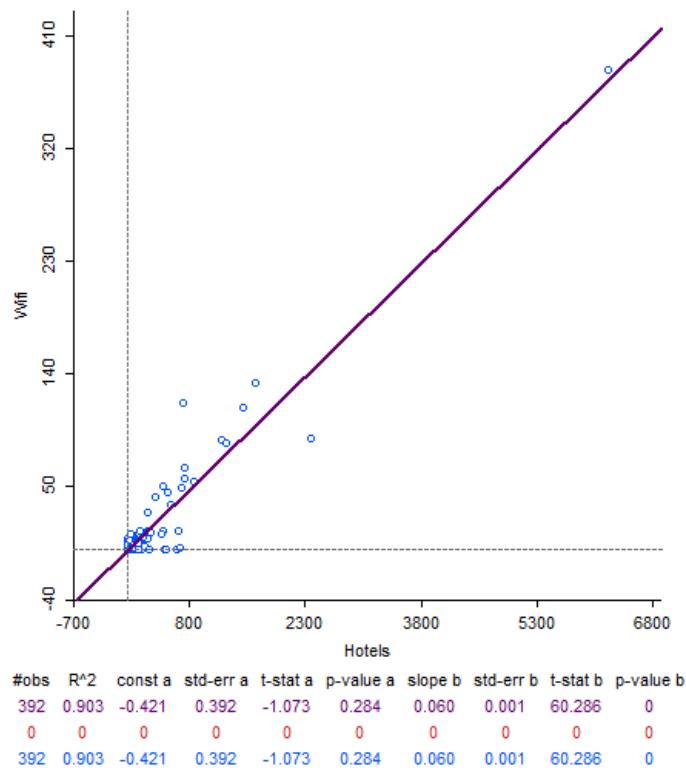


Figura 43 - Coeficiente de determinação para a regressão com a variável independente hotéis

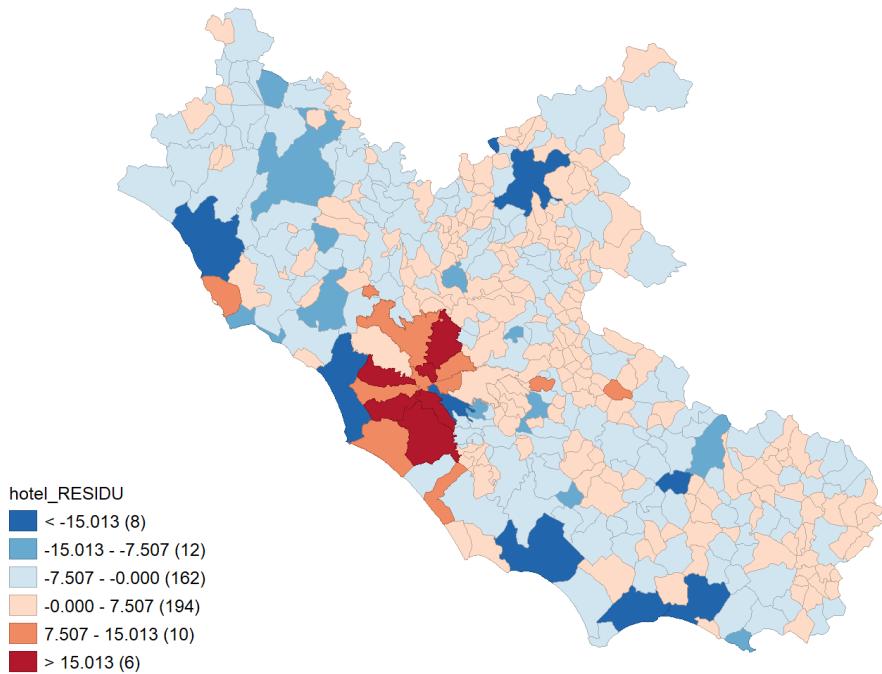


Figura 44 - Resíduos do modelo de regressão simples para a variável hotéis

Sílvia Mourão, FC 57541

As zonas nos tons mais claros representam as zonas onde os valores foram estimados com melhor precisão enquanto que as cores mais fortes representam as zonas com maior diferença entre a estimativa pelo modelo e o valor real. Aqui é possível constatar que o modelo tem dificuldade em estimar os valores para as zonas que tinham anteriormente sido identificadas como aglomerados de valores altos e baixos.

No que diz respeito à performance do modelo foram obtidos os relatórios visíveis na figura 45 e 46:

```
R-squared      : 0.903092 F-statistic       : 3634.45
Adjusted R-squared : 0.902844 Prob(F-statistic)   : 0
Sum squared residual: 22033.1 Log likelihood     : -1345.92
Sigma-square    : 56.4951 Akaike info criterion : 2695.83
S.E. of regression : 7.51632 Schwarz criterion   : 2703.77
Sigma-square ML  : 56.2069
S.E of regression ML: 7.49713
```

*Figura 45 - Estatística da Regressão Simples com a variável independente hotéis*

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-0.42088	0.392162	-1.07323	0.28383
Hotels	0.0603585	0.0010012	60.2864	0.00000

*Figura 46 - Coeficientes do modelo de regressão simples para a variável hotéis*

Aqui é possível verificar que o modelo apresenta um R2 elevado e um p-value baixo, o que significa que o coeficiente é importante para o modelo. O modelo de regressão para esta variável é então dado por:

$$\text{Pontos de Acesso} = -0.42088 + 0.0603585 \times \text{hotéis}$$

A mesma abordagem foi utilizada para a variável turismo, sendo os resultados apresentados em seguida nas figuras 47 e 48.

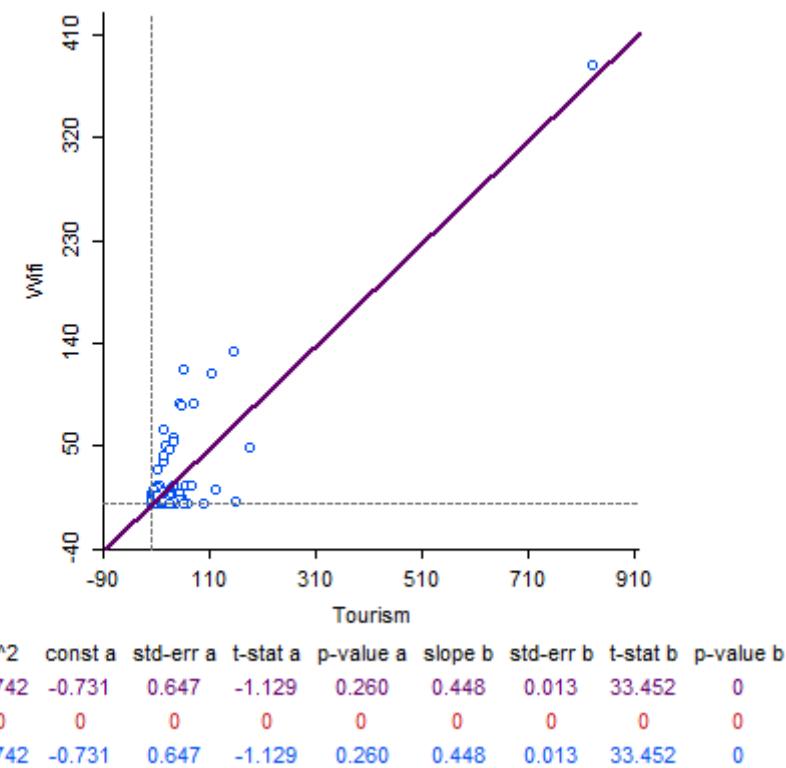


Figura 47 - Coeficiente de determinação para a regressão com a variável independente turismo

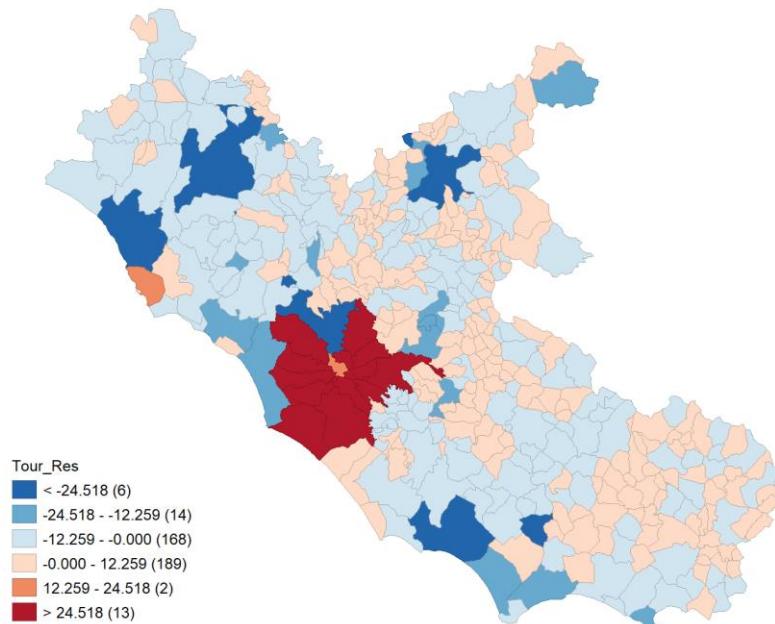


Figura 48 - Resíduos do modelo de regressão simples para a variável turismo

Sílvia Mourão, FC 57541

Analizando o mapa de resíduos é possível perceber que este modelo tem um desempenho pior no que toca à estimativa de valores, principalmente na zona central onde existe a maior concentração de pontos.

No que diz respeito à performance do modelo foram obtidos os relatórios visíveis na figura 49 e 50:

R-squared	:	0.741555	F-statistic	:	1119.02
Adjusted R-squared	:	0.740892	Prob(F-statistic)	:	0
Sum squared residual	:	58760.6	Log likelihood	:	-1538.18
Sigma-square	:	150.668	Akaike info criterion	:	3080.35
S.E. of regression	:	12.2747	Schwarz criterion	:	3088.3
Sigma-square ML	:	149.899			
S.E of regression ML	:	12.2433			

Figura 49 - Estatística da Regressão Simples com a variável independente turismo

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-0.730998	0.647409	-1.12911	0.25955
Tourism	0.447821	0.013387	33.4518	0.00000

Figura 50 - Coeficientes do modelo de regressão simples para a variável turismo

O modelo apresenta um  $R^2$  elevado e um p-value baixo o que indica que a variável independente é muito importante para o modelo, no entanto tanto  $R^2$  como o mapa de resíduos demonstram que o primeiro modelo se ajusta melhor a estes dados. O modelo de regressão é dado por:

$$\text{Pontos de Acesso} = -0.732998 + 0.447821 \times \text{turismo}$$

Em seguida, considerou-se a abordagem da regressão linear múltipla. Para isso considerou-se um limite para a regressão múltipla de  $R^2 > 0.65$  e fez-se a regressão múltipla para estas 5 variáveis. No entanto, os resultados desta regressão mostram um coeficiente muito baixo para a variável densidade de população, pelo que esta variável foi removida e uma nova regressão múltipla efetuada para um limite de  $R^2 > 0.7$ , como é visível na figura 51.

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.404655	0.374861	1.07948	0.28105
Hotels	0.0945048	0.00880382	10.7345	0.00000
Commerce	-0.086074	0.0195909	-4.39357	0.00001
Nightlife	-0.503105	0.106786	-4.71133	0.00000
Pop_dens	5.45753e-07	7.95372e-08	6.8616	0.00000
Tourism	-0.0526508	0.0298954	-1.76117	0.07900

Figura 51 - Primeira tentativa de regressão múltipla

O mapa dos resíduos obtido por este modelo é visível na figura 52.

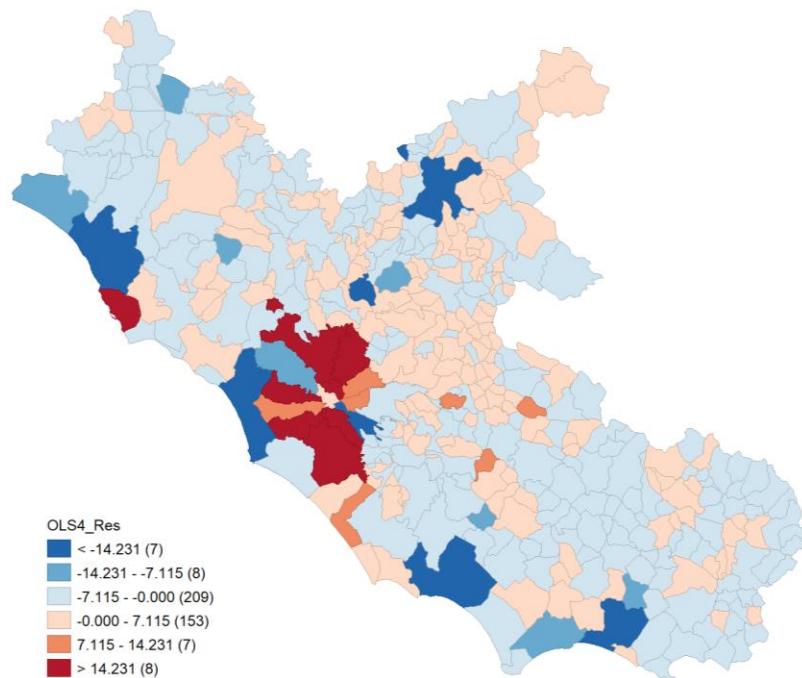


Figura 52 - Resíduos do modelo de regressão múltipla

Os valores estimados para este modelo mais uma vez continuam a ter dificuldades em aproximar os valores observados em certas regiões, no entanto existem já alguns municípios na zona da área metropolitana de Roma que estão a ser sobreestimados em vez de subestimados.

No que diz respeito à performance do modelo foram obtidos os relatórios visíveis nas figuras 53 e 54:

R-squared	:	0.912931	F-statistic	:	1014.44
Adjusted R-squared	:	0.912031	Prob(F-statistic)	:	0
Sum squared residual:		19796.1	Log likelihood	:	-1324.93
Sigma-square	:	51.1527	Akaike info criterion	:	2659.86
S.E. of regression	:	7.15211	Schwarz criterion	:	2679.72
Sigma-square ML	:	50.5002			
S.E of regression ML:		7.10635			

Figura 53 - Estatística da regressão múltipla

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.431081	0.396531	1.08713	0.27765
Hotels	0.115617	0.00872593	13.2498	0.00000
Commerce	-0.0998495	0.0206154	-4.84345	0.00000
Nightlife	-0.448424	0.11265	-3.98068	0.00008
Tourism	-0.137309	0.0288063	-4.76661	0.00000

Figura 54 - Coeficientes do modelo de regressão múltipla

Mais uma vez observa-se um valor de  $R^2$  elevado e valores de significância baixos, sendo o maior o da variável de vida noturna. O modelo de regressão é dado por:

$$\text{Pontos de Acesso} = 0.431081 + 0.115617 \times \text{hotéis} - 0.0998495 \times \text{comércio} - 0.448424 \times \text{noturna} - 0.137309 \times \text{turismo}$$

### 6.3 Avaliação do desempenho dos modelos numa área de estudo diferente

Um outro método de validação dos modelos é a possibilidade de estes poderem ser aplicados a outras regiões. Para este estudo foi escolhida a região da Lombardia, que contém a segunda cidade com maior população em Itália, Milão.

Foram obtidos os dados para as quatro variáveis que foram utilizadas para a regressão múltipla e que são também utilizadas para as duas regressões lineares simples.

Os valores estimados foram calculados através da expressão obtida para cada um dos modelos e os resíduos calculados pela diferença entre os valores calculados e os valores observados. Em seguida são apresentados os resultados da regressão com o modelo dos hotéis (figura 55), o modelo do turismo (figura 56) e o modelo da regressão múltipla (figura 57).

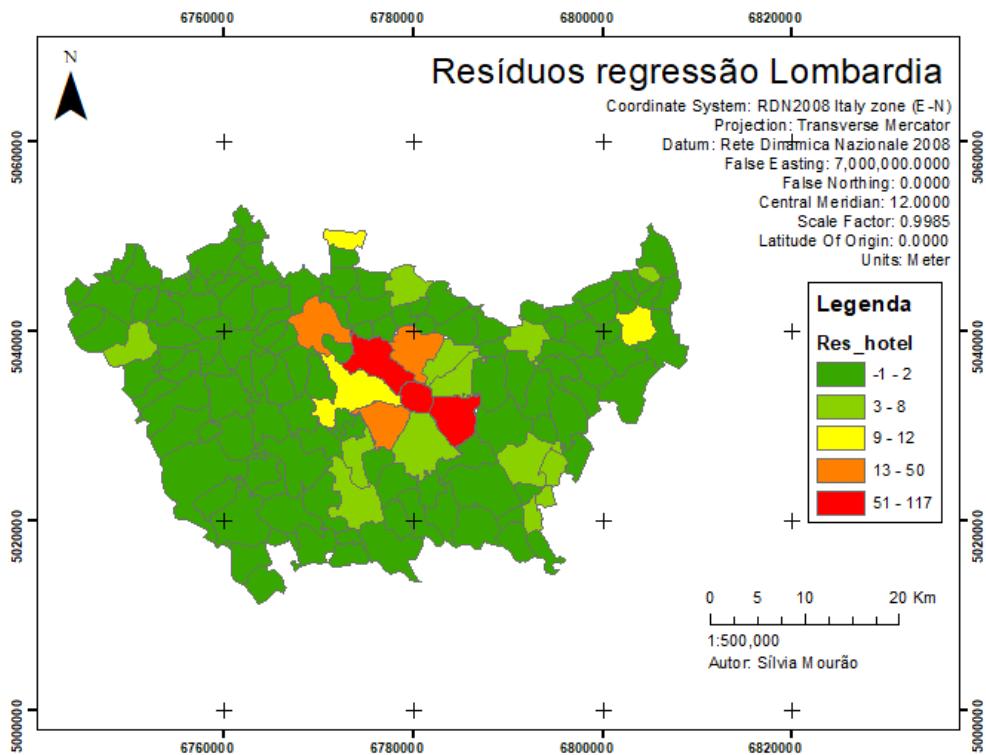


Figura 55 - Resíduos do modelo de regressão simples com a variável hotéis aplicada à região da Lombardia

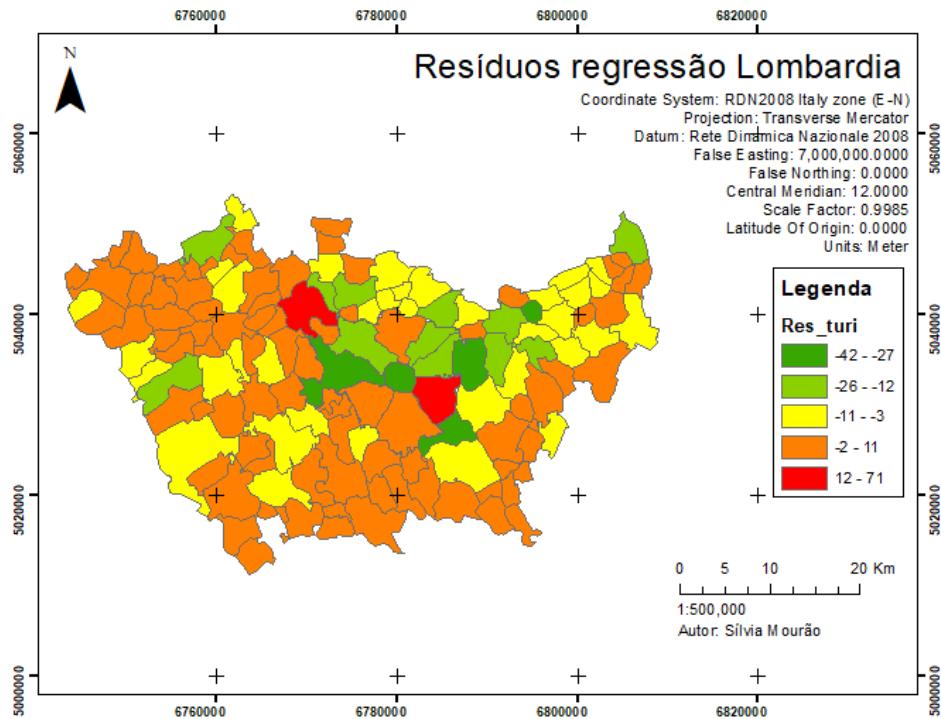


Figura 56 - Resíduos do modelo de regressão simples com a variável turismo aplicada à região da Lombardia

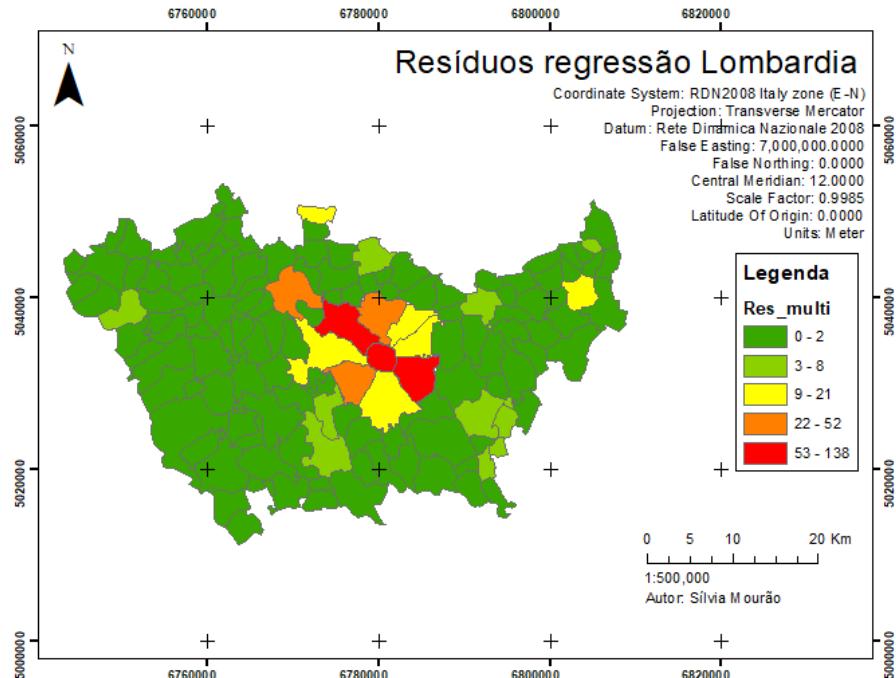


Figura 57 - Resíduos do modelo de regressão múltipla aplicado à região da Lombardia

Analizando estas figuras é possível constatar que o modelo dos hoteis parece ter uma melhor estimativa dos valores, no entanto é necessário referir que para o modelo do turismo são as áreas a laranja que correspondem a uma melhor estimativa. Ainda assim esse modelo aparenta ter um pior desempenho no que toca à visualização de cada área individualmente.

Como forma de comparar os modelos foi elaborada a tabela 11:

Tabela 11 - Análise da performance dos modelos na região da Lombardia

	Resíduos	Soma	Média	Desvio Padrão
Milão	OLS hotéis	502	3.59	15.17
	OLS turismo	-430	-3.07	10.57
	OLS múltipla	602	4.3	17.21
Roma	OLS hotéis	-24	-0.06	7.49
	OLS turismo	-12	-0.03	12.23
	OLS múltipla	4	0.01	7.12

O modelo tem um desempenho muito pior na região da Lombardia do que na região da Lazio, no entanto o modelo do turismo parece ser aquele que melhor se ajusta aos dados. Devido à elevada correlação observada nos modelos da regressão com a variável independente hotéis e com as variáveis múltiplas existe um sobre ajustamento do modelo à região de estudo e por isso o modelo tem um desempenho pior quando é colocado numa nova região.

#### 6.4 Avaliação dos resultados e diagnóstico dos resíduos do modelo OLS

Para avaliar se as regressões lineares obtidas cumprem os pré-requisitos estabelecidos para o modelo OLS, para todos os modelos de regressão foram realizados seis testes, sendo os resultados totais apresentados na tabela 12 e os resultados para cada teste no modelo da regressão múltipla descritos em seguida:

- Análise da Linearidade

Para validar este critério é necessário verificar se os resíduos apresentam uma distribuição linear em torno do eixo de x. A distribuição dos resíduos deve ser aleatória, o que não se verifica neste caso (figura 58):

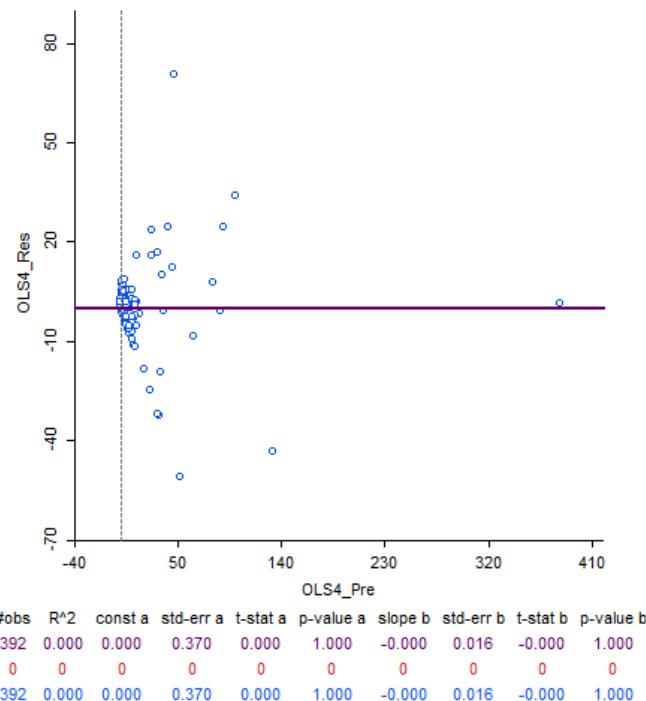


Figura 58 - Distribuição dos resíduos

- Análise da Homoscedasticidade

Para validar este critério é necessário que os dados apresentem variância nula, que se distribuam de forma uniforme em torno do eixo de x, no entanto a figura 58 demonstra que existe um afunilamento dos resultados. Para este critério pode também ser aplicado o teste de Breusch-Pagan, onde se estabelece um teste de hipótese sobre a constância da variância, sendo depois validada a hipótese através da distribuição de  $\chi^2$ . Neste caso, o teste resultou num valor de 859.67 e significância <0.001, concluindo-se que a hipótese nula deve ser rejeitada, pois existe um aumento da variância à medida

que a variável independente aumente, o que significa que este modelo de regressão não passa o teste da Homoscedasticidade.

- Análise da Normalidade dos Resíduos

Para avaliar a normalidade é graficado o histograma dos resíduos e verifica-se se este segue uma distribuição normal. Como é visível na figura 59 este facto não se verifica.

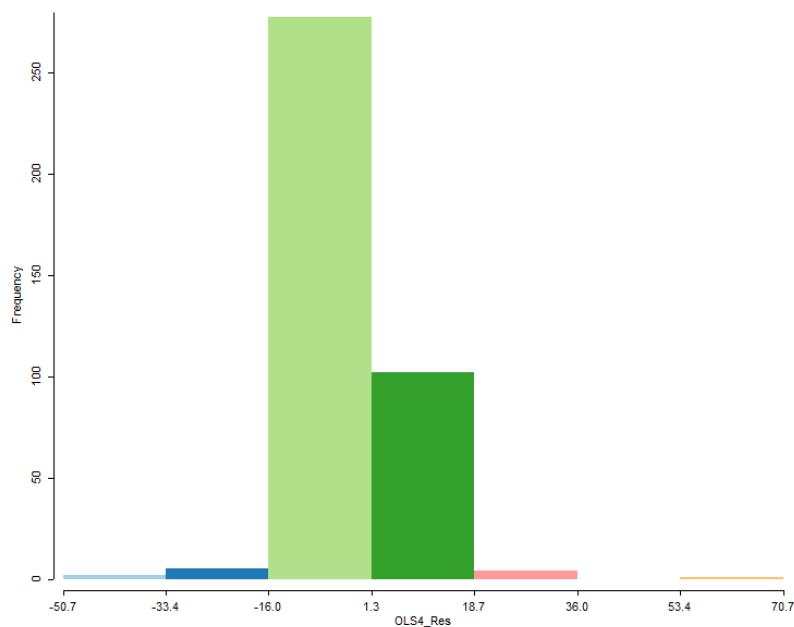


Figura 59 - Histograma dos resíduos

Um outro teste que se pode realizar é o QQ plot dos resíduos. Este gráfico compara o quantil observado com o quantil teórico de uma distribuição normal, ou seja, para uma distribuição que siga a distribuição normal é esperado que os valores observados se encontrem em torno de uma reta. Tal não é observado neste caso, como é visível na figura 60:

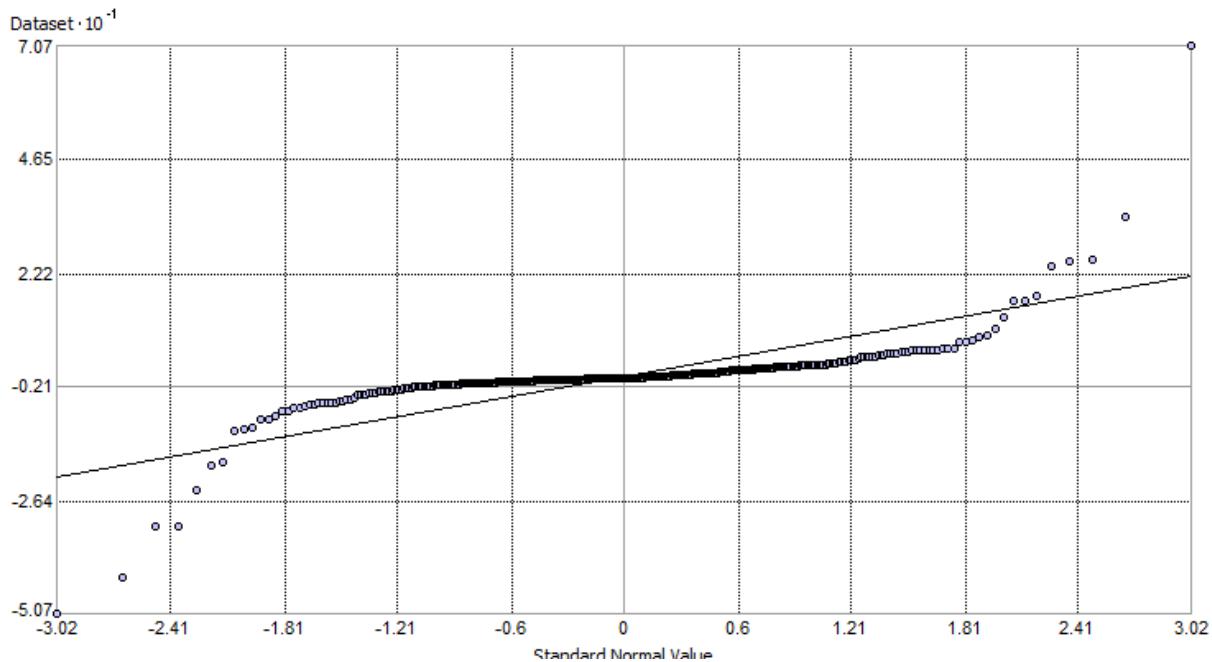


Figura 60 - QQ plot dos resíduos

O último teste à normalidade é o teste de Jarque-Bera, que é também uma medida que permite interpretar se os resultados seguem uma distribuição normal. Os valores obtidos das observações são mais uma vez comparados com os valores teóricos através da distribuição de  $\chi^2$ . Neste caso o valor obtido foi de 23192.83 para uma significância inferior a 0.001, pelo que a hipótese nula é rejeitada e conclui-se que os resíduos não apresentam uma distribuição normal.

- Independência dos resíduos

Para testar a independência dos resíduos é necessário utilizar a figura 58 e perceber se estes se encontram distribuídos de forma aleatória. Neste caso, os resíduos formam um aglomerado de valores junto à origem, pelo que este critério não se verifica.

- Multicolinearidade

A multicolinearidade é um termo que indica a existência de elevada correlação entre as variáveis independentes. O valor é calculado durante a regressão, tendo sido obtido para esta regressão um valor de 22.458403, o que é inferior ao limite de 30 e por isso aceitável para o estudo.

- Análise de autocorrelação espacial dos resíduos

Para efetuar a análise da autocorrelação espacial dos resíduos foi utilizado o índice de Moran global e local, tal como tinha sido feito no estudo da análise de áreas, utilizando uma matriz de vizinhança do tipo queen contiguity. Foi obtido um resultado de 0.167, o que é ainda bastante próximo de zero e pode indicar que não existe autocorrelação global dos resíduos (figura 61).

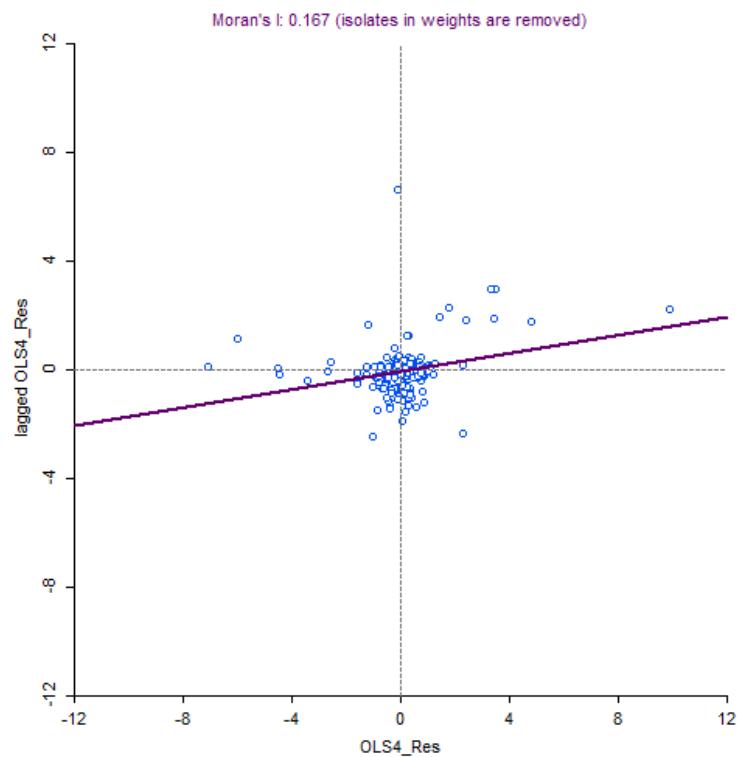


Figura 61 - Índice de Moran Global para a autocorrelação espacial dos resíduos

A nível local, figura 62, o índice de Moran revela a existência de aglomerados do tipo high-high na zona de Roma, e low-low a norte, este e sul de Roma, o que significa que estas são as zonas onde o modelo tem maior dificuldade a ajudar-se aos dados observados. As zonas costeiras parecem ter maior impacto na autocorrelação, como é visível no mapa de significância na figura 63:

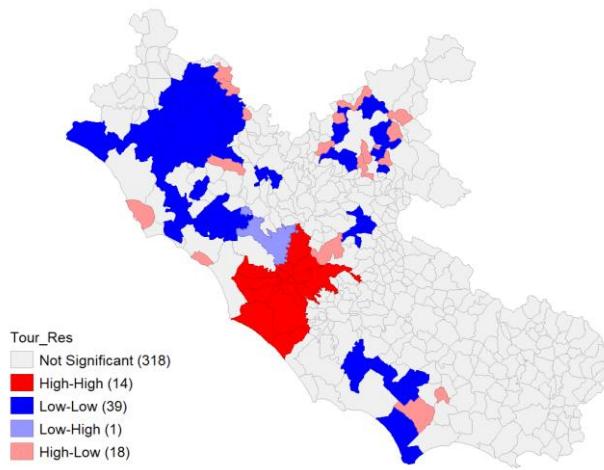


Figura 62 - Índice de Moran local para os resíduos

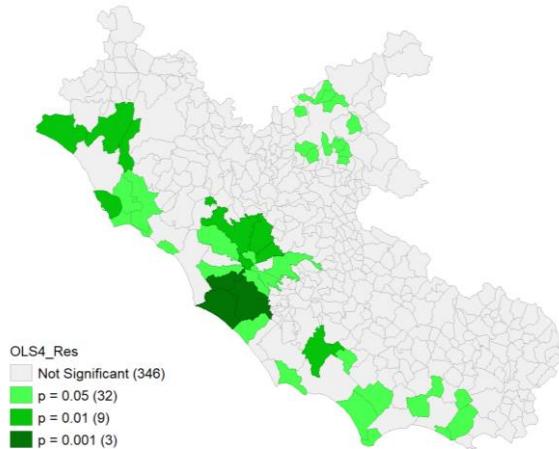


Figura 63 - Mapa de significância para os resíduos

- Conclusões sobre a necessidade de utilização de Modelos de Regressão Espacial

A tabela 12 mostra um resumo dos critérios do modelo de regressão simples aplicados às três regressões efetuadas. Em todos os casos existem pelo menos três critérios que não correspondem à expectativa teórica, parecendo existir uma correlação espacial para todas as regressões lineares em estudo, pelo que deve ser feita uma regressão espacial. Para as regressões lineares o caso da regressão múltipla apresenta o melhor resultado.

Tabela 12 - Sumário dos resultados obtidos para as regressões lineares

Indicadores	OLS Hotéis	OLS Turismo	OLS Múltipla
<b>Linearidade</b>	R2 = 0.903	R2 = 0.742	R2 = 0.91
<b>Homoscedasticidade</b>	H0 Rejeitado (Teste Breusch-Pagan)	H0 Rejeitado (Teste Breusch-Pagan)	H0 Rejeitado (Teste Breusch-Pagan)
<b>Normalidade</b>	H0 Rejeitado (Teste de Jarque-Bera)	H0 Rejeitado (Teste de Jarque-Bera)	H0 Rejeitado (Teste de Jarque-Bera)
<b>Independência</b>	Distribuição de Resíduos não é aleatória	Distribuição de Resíduos não é aleatória	Distribuição de Resíduos não é aleatória
<b>Multicolinearidade</b>	1.292	1.345	22.5
<b>Dependência Espacial</b>	Alguma correlação espacial	Baixa Autocorrelação	Baixa Autocorrelação

## 6.5. Regressão Espacial

Tendo sido constatado que existe uma componente espacial que influencia os resultados da regressão linear, é então necessário utilizar métodos de regressão espacial para as mesmas variáveis e verificar como é que os modelos se comportam.

Para estes modelos é novamente considerada uma matriz de vizinhança do tipo queen contiguity de 1<sup>a</sup> ordem para realizar a análise e utilizaram-se três modelos de regressão: os primeiros dois modelos, do tipo global, são o modelo spatial lag, que atribui a autocorrelação à variável dependente, e o modelo spatial error, que atribui a autocorrelação espacial ao erro. Existe ainda um modelo de regressão em regime local, denominada GWR (geographically weighted regression - regressão geográfica ponderada), onde existe uma subdivisão da área de estudo e são realizadas regressões para cada sub-região, o que significa que cada uma vai ter um modelo de regressão diferente. O modelo de GWR geralmente tem um ajuste muito bom à área de estudo, tendo a vantagem de permitir perceber quais os fatores que têm maior impacto no resultado de cada área, no entanto não são diretamente transferíveis para outras regiões de estudo.

Foram realizados em primeiro lugar as regressões do tipo global, sendo que os resultados obtidos para o modelo spatial lag são visíveis na figura 64 e os resultados obtidos pelo modelo spatial error são visíveis na figura 65:

R-squared	:	0.927496	Log likelihood	:	-1290.01
Sq. Correlation	:	-	Akaike info criterion	:	2592.02
Sigma-square	:	42.0524	Schwarz criterion	:	2615.85
S.E of regression	:	6.48479			

Variable	Coefficient	Std.Error	z-value	Probability
W_Wifi	0.159043	0.0184773	8.60744	0.00000
CONSTANT	-0.285655	0.370611	-0.770767	0.44084
Hotels	0.103998	0.00800102	12.9981	0.00000
Tourism	-0.0711438	0.0269441	-2.64042	0.00828
Nightlife	-0.390925	0.102415	-3.81707	0.00014
Commerce	-0.112362	0.0187587	-5.98984	0.00000

Figura 64 - Resultados da regressão spatial lag

R-squared	:	0.926092	R-squared (BUSE)	:	-
Sq. Correlation	:	-	Log likelihood	:	-1304.236483
Sigma-square	:	42.8671	Akaike info criterion	:	2618.47
S.E of regression	:	6.5473	Schwarz criterion	:	2638.33

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	1.10121	0.681981	1.61473	0.10637
Hotels	0.0995276	0.00791569	12.5735	0.00000
Tourism	-0.0583193	0.0261981	-2.22609	0.02601
Nightlife	-0.461198	0.102137	-4.51549	0.00001
Commerce	-0.0945205	0.0179679	-5.26053	0.00000
LAMBDA	0.512804	0.0570097	8.99503	0.00000

Figura 65 - Resultados da regressão spatial error

Uma discussão mais profunda destes resultados será feita no final deste capítulo.

Foi de seguida realizada a regressão local GWR e os resultados dos resíduos obtidos estão mapeados na figura 66:

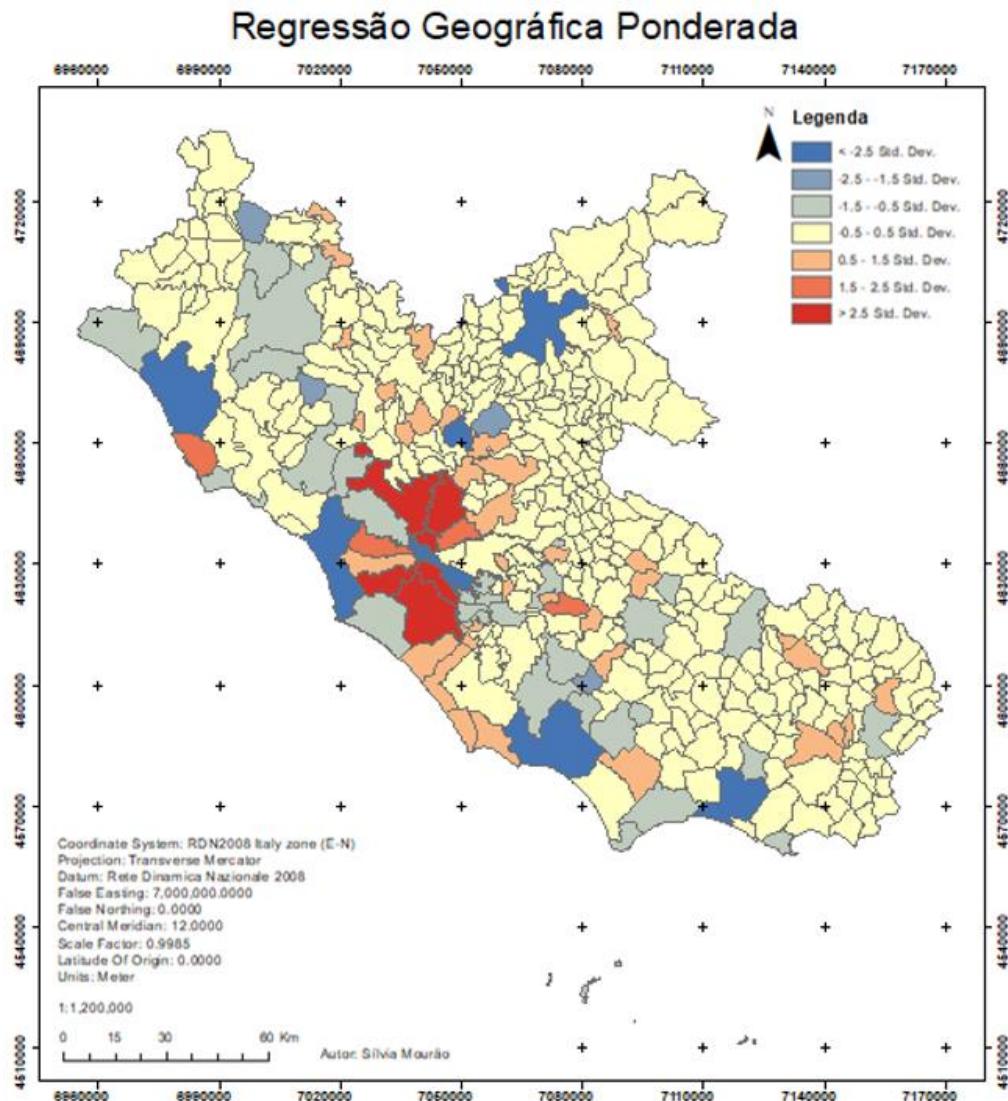


Figura 66 - Mapeamento do desvio padrão dos resíduos obtidos pelo GWR

É possível constatar que o desvio padrão dos resíduos é elevado (positivo) em alguns municípios da cidade de Roma e também elevado (negativo) em zonas junto à costa. Nas figuras seguintes é possível ver uma comparação entre os valores reais dos pontos de acesso (figura 67) e os valores estimados pelo modelo de regressão GWR (figura 68) tendo sido utilizados aproximadamente os mesmos intervalos de escala.

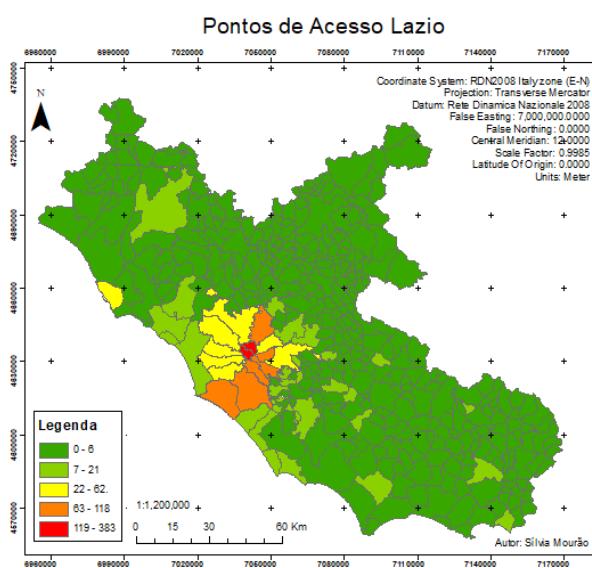


Figura 67 - Valores reais de pontos de acesso na Lazio (esquerda)

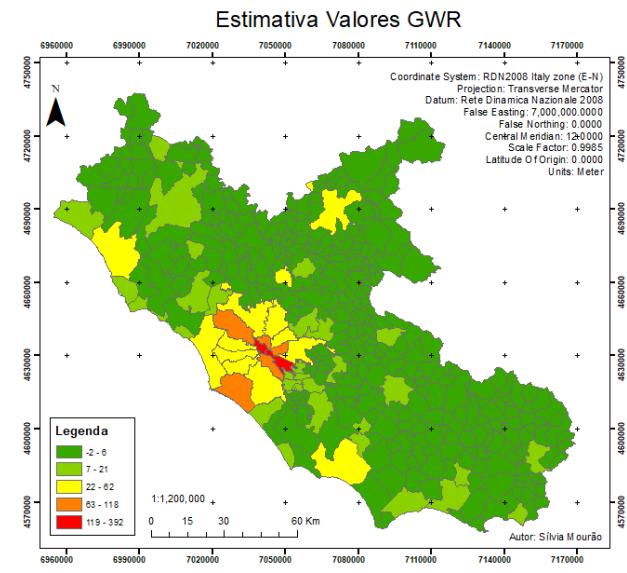


Figura 68 - Valores estimados de pontos de acesso pelo modelo de regressão geográfica ponderada (direita)

O modelo parece no geral fornecer uma boa aproximação aos valores reais.

Uma outra análise que é possível fazer com este modelo é o comportamento de  $R^2$  localmente para a área em estudo. Neste caso verifica-se que existe melhor correlação na zona central, diminuindo à medida que se afasta do centro, como é visível na figura 69:

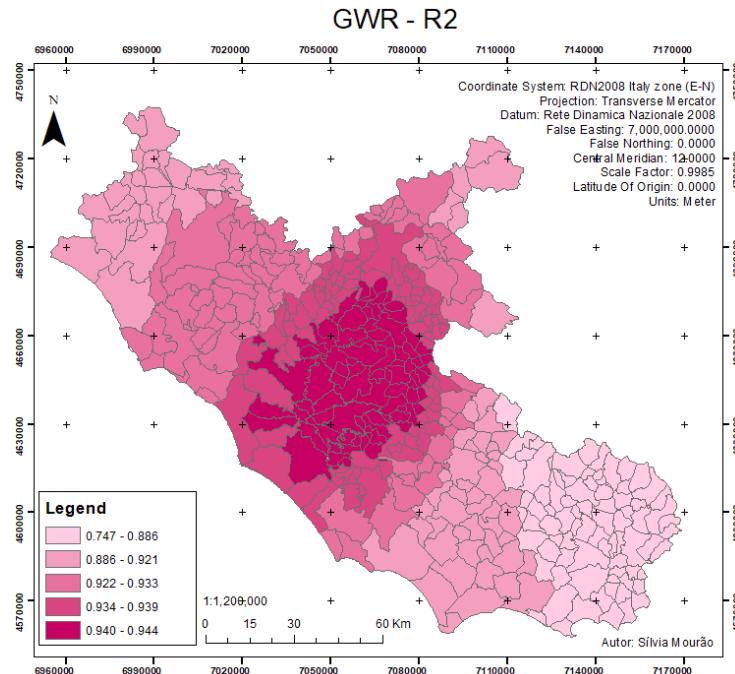


Figura 69 - Variação de  $R^2$  calculado pelo modelo GWR

Ainda assim, a variação de  $R^2$  está contida no intervalo de [0.747 - 0.944] pelo que existe para toda a região um bom modelo de regressão utilizando estas variáveis independentes.

Em seguida foi feito um estudo sobre a contribuição de cada uma das variáveis para o modelo, tendo sido obtido o resultado visível na figura 70:

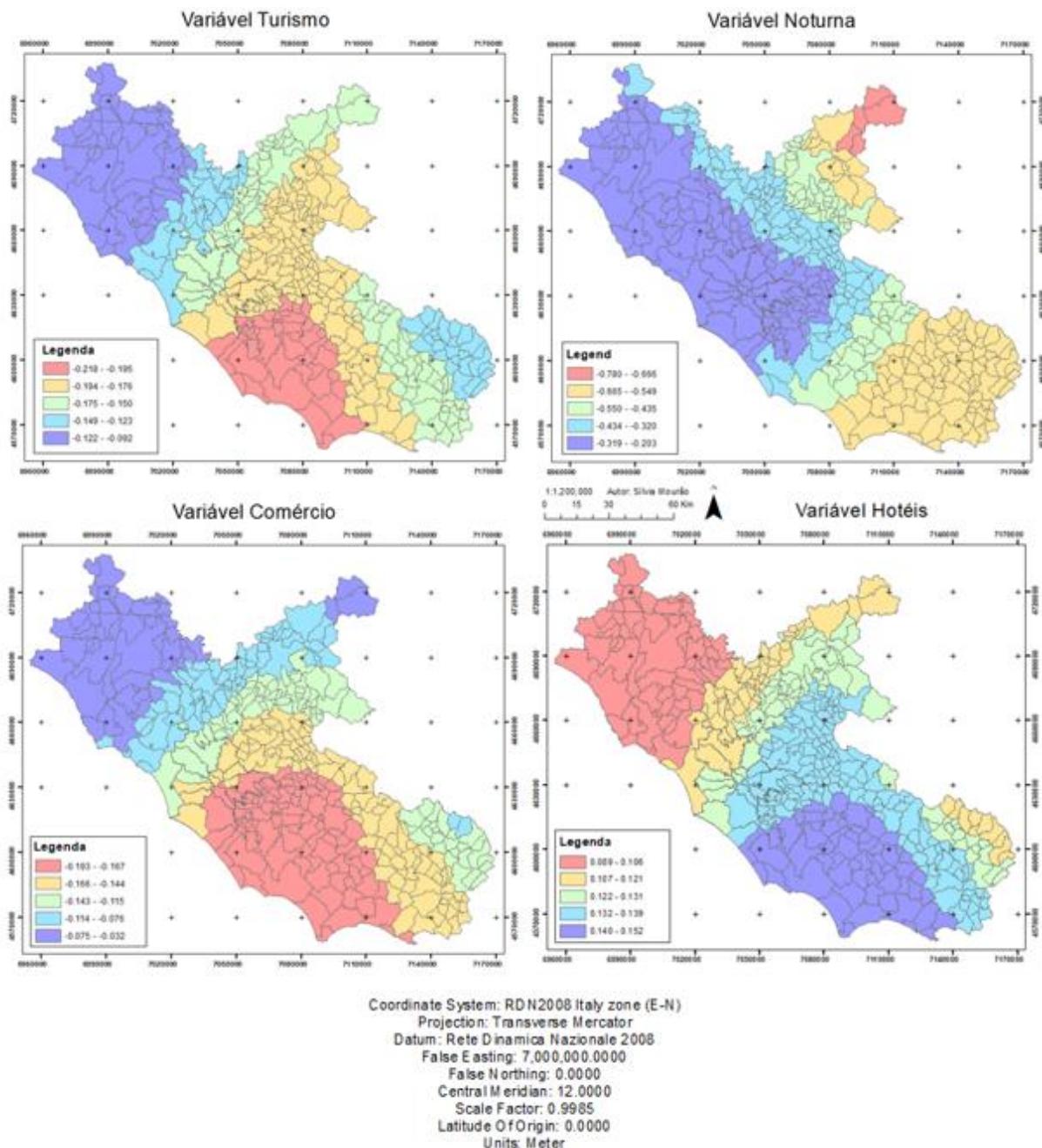


Figura 70 - Regiões de maior e menor contribuição para cada uma das variáveis independentes

Sílvia Mourão, FC 57541

As regiões em tons de roxo correspondem a uma maior contribuição da variável em questão enquanto que as zonas a vermelho correspondem a uma menor contribuição. A variável hotéis é sempre a que tem maior relevância em todas as zonas pois é a única que apresenta valores positivos, o que é esperado devido ao valor elevado de correlação entre esta variável e a variável dependente. A área metropolitana de Roma, ao centro, apresenta uma grande variabilidade de contribuições em todas as variáveis, à exceção da variável vida noturna.

## 6.6 - Comparação dos modelos

Após terem sido realizadas regressões com os vários modelos estudados, compilou-se a seguinte tabela que sumariza os resultados obtidos para os vários casos:

*Tabela 13 - Comparação dos modelos de regressão*

	<b>OLS1 - Hotéis</b>	<b>OLS2 - Turismo</b>	<b>OLS-Multi</b>	<b>LAG</b>	<b>ERROR</b>	<b>GWR</b>
<b>R<sup>2</sup></b>	0.903	0.742	0.913	<b>0.927</b>	0.926	0.923
<b>Log Likelihood</b>	-1345.92	-1538.18	-1324.93	<b>-1290.01</b>	-1304.24	-
<b>Akaike info criterion</b>	2695.83	3080.35	2659.86	<b>2592.02</b>	2618.47	2628.96
<b>Schwarz Criterion</b>	2703.77	3088.3	2679.72	<b>2615.85</b>	2638.33	-

Analizando a tabela 13 é possível verificar que todos os modelos apresentam muito bons resultados, sendo que o pior foi o modelo da regressão linear simples com a variável turismo e o modelo de regressão que melhor se ajusta às observações deste projeto é o Spatial LAG. As regressões espaciais obtiveram melhores resultados do que as regressões clássicas devido ao facto de se ter provado a existência de autocorrelação espacial dos resíduos. A distribuição dos pontos de acesso WiFi gratuitos na região da Lazio pode claramente ser explicada a partir da distribuição dos hotéis, locais de comércio, pontos de interesse turístico e estabelecimentos de diversão noturna.

## 7. Interpolação Espacial

A interpolação espacial é um procedimento que converte dados de observações pontuais em superfícies continuas, de forma a estimar os valores nas regiões desconhecidas.

A interpolação pode ser classificada como pertencendo a vários grupos, podendo ser do tipo global ou local no que toca à extensão espacial, do tipo exato ou aproximado relativamente ao ajustamento e utilizar um modelo do tipo determinístico ou probabilístico (Catita, 2022). Para este capítulo vão ser utilizado um modelo global determinístico, vários modelos locais determinísticos e um modelo local probabilístico.

A etapa da análise de superfícies teve dois objetivos: a determinação de um mapa de intensidade do sinal WiFi para depois poder ser utilizado para determinar a impedância das estradas na fase de análise de redes e a interpolação de modelo digital do terreno com o objetivo de verificar se é necessário considerar também a altimetria como impedância para os trajetos.

### 7.1 Preparação / Tratamento dos dados

Com o objetivo de realizar a interpolação, foram primeiro estabelecidas as condições necessárias para a interpolação do sinal WiFi. Para este estudo utilizou-se apenas a área que corresponde à “comune” de Roma, dividida em municípios e que será também a área a utilizar para a última parte do projeto. O sinal WiFi tipicamente obtido pelos pontos de acesso gratuitos disponíveis na cidade tem um alcance garantido de 100 m. O sinal WiFi tem geralmente uma boa propagação nos primeiros 50 m, podendo no máximo o sinal propagar-se até aos 150 m de distância do ponto de acesso (*How Far Can a WiFi Signal Travel? - Actiontec.Com*, n.d.).

Tendo em atenção estas propriedades do sinal WiFi, foi utilizado o QGIS para definir zonas circulares de propagação, tendo sido estabelecido a seguinte lógica para a percentagem da intensidade de sinal recebida a uma relativa distância do ponto de acesso: ponto de acesso = 100%, 50 m envolventes = 100%, 100 m envolventes = 50%, 150 m = 0%. Os círculos envolventes foram depois convertidos para pontos com o objetivo de interpolar a qualidade do sinal visto que a distribuição não é linear e tem uma granularidade grande. É importante referir que o caso que está aqui a ser estudado é simplificado – normalmente teriam de ser consideradas outras fontes de interferência do sinal como barreiras naturais ou construções.

Esta grelha de pontos foi depois importada para o ArcMap onde foi lhe foi retirada uma pequena amostra de pontos para posteriormente validar os resultados da interpolação, estando assim os dados prontos para serem utilizados no interpolador.

Para o caso da interpolação do modelo digital de terreno, foram utilizadas as curvas de nível que tinham sido obtidas no conjunto de dados iniciais e destas foram extraídas os pontos de cada curva de nível. Infelizmente não foi possível obter dados da rede de nivelamento de alta precisão italiana pois estes não estão disponíveis como dados abertos, pelo que mais uma vez foi realizada uma amostra de pontos para validação posterior. Estes dados foram também utilizados para interpolação.

## 7.2 Métodos de Interpolação Determinísticos (Locais e Globais) - Sinal WiFi

Os métodos determinísticos de interpolação utilizam principalmente estimativas a partir de modelos matemáticos com base em médias ponderadas de valores conhecidos e interpolam a superfície a partir da proximidade – os pontos mais próximos daquele que pretendemos estimar têm um maior peso do que os valores mais afastados. Estes métodos têm a desvantagem de não fornecerem uma estimativa dos erros dos valores estimados.

Os interpoladores globais utilizam todos os pontos da amostra para gerar uma única função, que é depois aplicada por toda a região. Este método tende a produzir superfícies suavizadas, que permitem observar a tendência geral da superfície, mas que não têm boa resolução a nível pontual. A superfície de tendência é um método de interpolação determinístico global, que gera uma superfície suavizada que se ajusta ao conjunto de pontos da amostra através de uma função matemática polinomial. O grau do polinómio vai resultar em inclinações e deformações da superfície à medida que esta tenta criar um maior ajuste aos dados observados. Utilizando o valor de erro médio quadrático obtido pela validação que o ArcMap efetua com base nos pontos a serem interpolados, utilizou-se neste caso a interpolação para um polinómio de grau 10, sendo os resultados obtidos visíveis na figura 71:

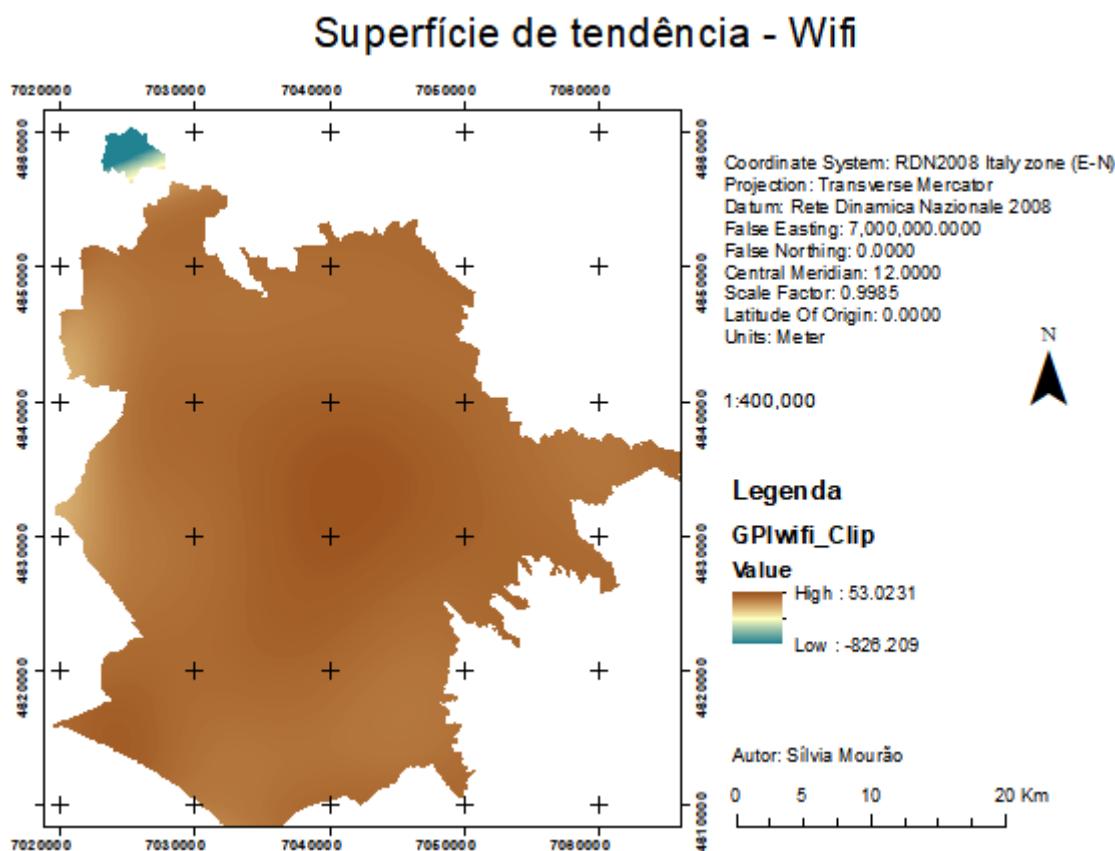


Figura 71 - Superfície de tendência interpolada para o sinal WiFi

A superfície de tendência fornece um muito mau ajuste aos dados, sendo que existem de facto valores de castanho mais intensos nas regiões centrais e costeiras o valor dispara quando chega à zona limite onde não existem dados.

De seguida foram considerados métodos de interpolação determinística local. Os interpoladores locais aplicam repetidamente o processo de interpolação a um pequeno conjunto de pontos vizinhos entre si e conseguem captar a variação local da superfície. Antes de passar à própria interpolação, foi primeiramente feita uma abordagem de gerar polígonos de Thiessen. Estes não são um método de interpolação, determinando as áreas mais próximas de um determinado ponto conhecido. O resultado obtido para a área é visível na figura 72, com uma versão ampliada da zona central na figura 73:

### Polígonos de Thiessen - WiFi

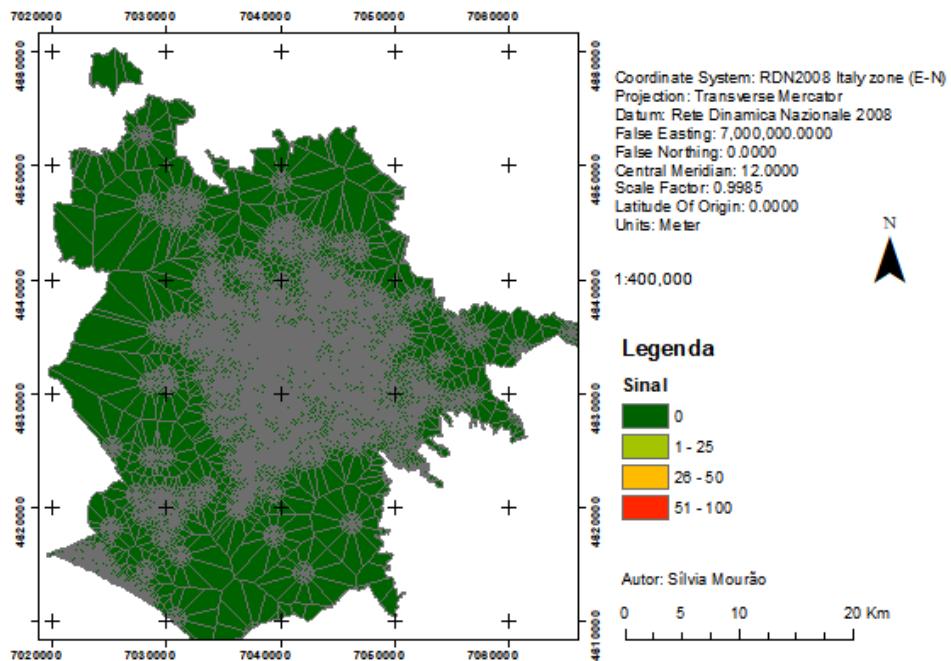


Figura 72 - Polígonos de Thiessen para o sinal WiFi

### Polígonos de Thiessen - WiFi

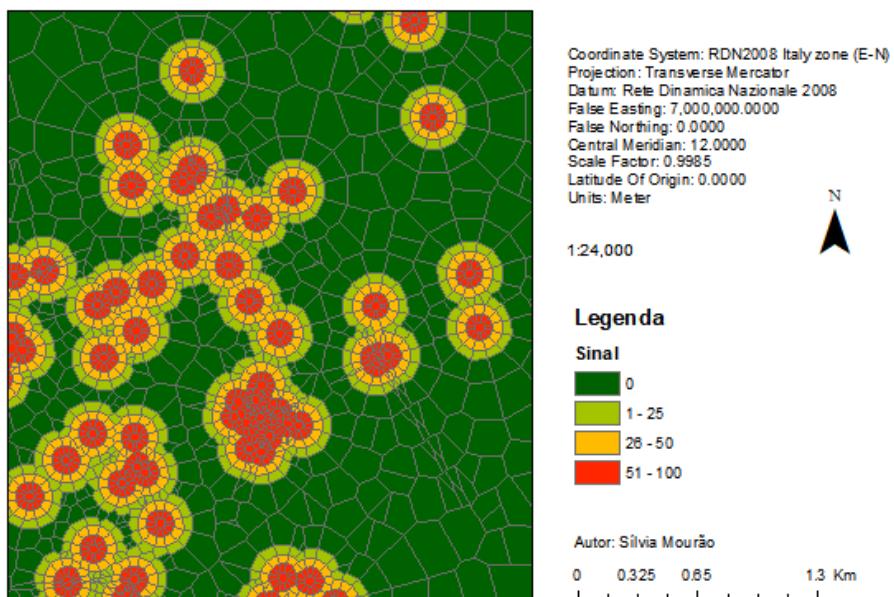


Figura 73 - Detalhe dos polígonos de Thiessen para o sinal WiFi

Passando-se então aos métodos interpoladores propriamente ditos, considerou-se inicialmente a Interpolação Polinomial Local (LPI), que é um interpolador aproximado, que utiliza polinómios como a superfície de tendência, mas exige mais decisões a nível de parâmetros a utilizar para o modelo.

Interpolador local determinístico (LPI). Neste caso otimizaram-se mais uma vez os valores tendo em conta os valores previstos de RMS para a interpolação. A superfície resultante encontra-se na figura 74:

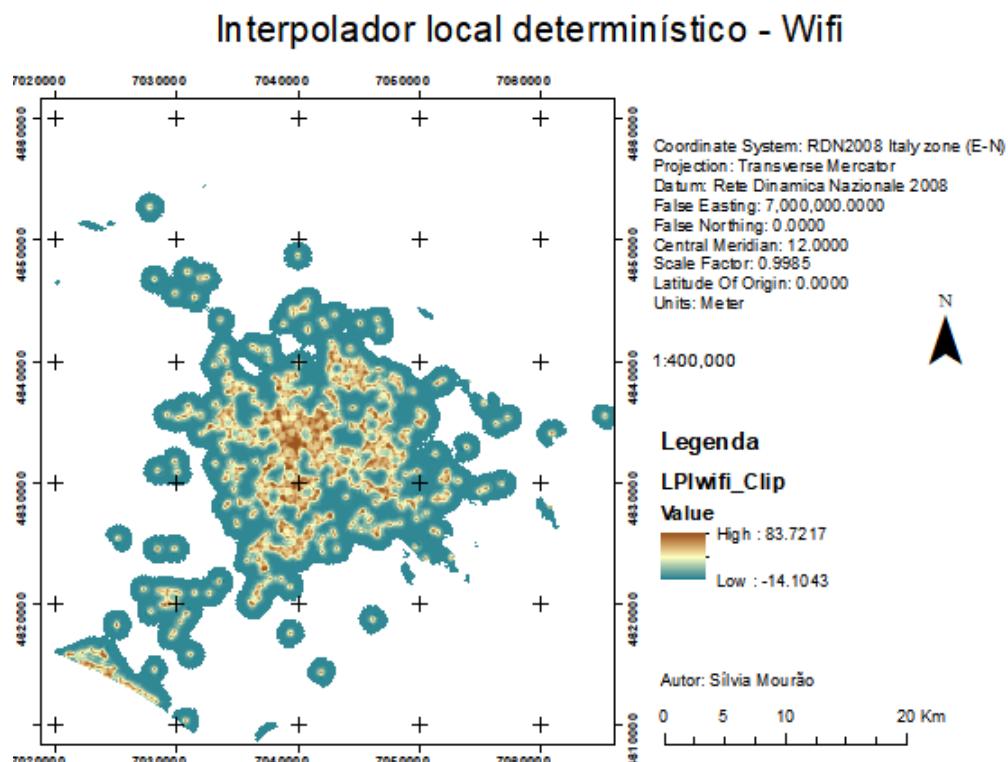


Figura 74 - Interpolador local determinístico para o sinal WiFi

Este método interpolador consegue apenas fazer estimativas na região onde existe densidade de pontos, sendo não conseguindo extrapolar para a área mais periférica onde não existem pontos de acesso.

Um outro método de interpolação determinística local que foi considerado foi o Inverso da distância ponderada (IDW). Este interpolador assume que o valor do atributo desconhecido é influenciado pelos pontos conhecidos mais próximos do ponto a estimar, considerando um maior peso para os valores da amostra mais próximos do valor a estimar. Essa relação assume normalmente o inverso do quadrado da distância, no entanto é possível definir uma outra potência para esta interpolação. No caso desta

interpolação foi utilizado o inverso da distância à quarta. A superfície obtida encontra-se na figura 75, sendo também visível a zona central a uma escala maior na figura 76:

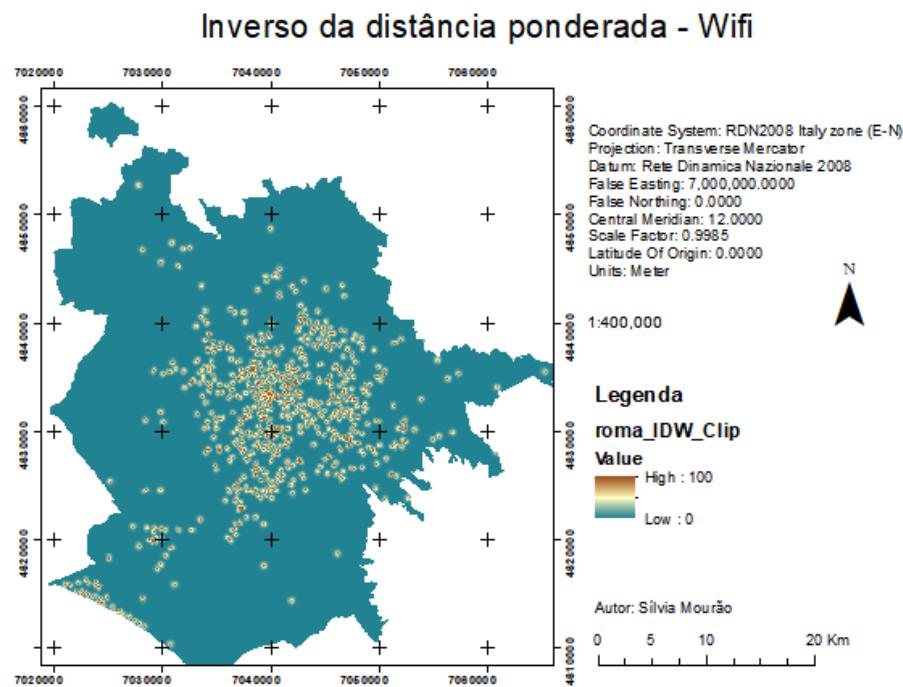


Figura 75 - Interpolação pelo método do inverso da distância ponderada para o sinal WiFi

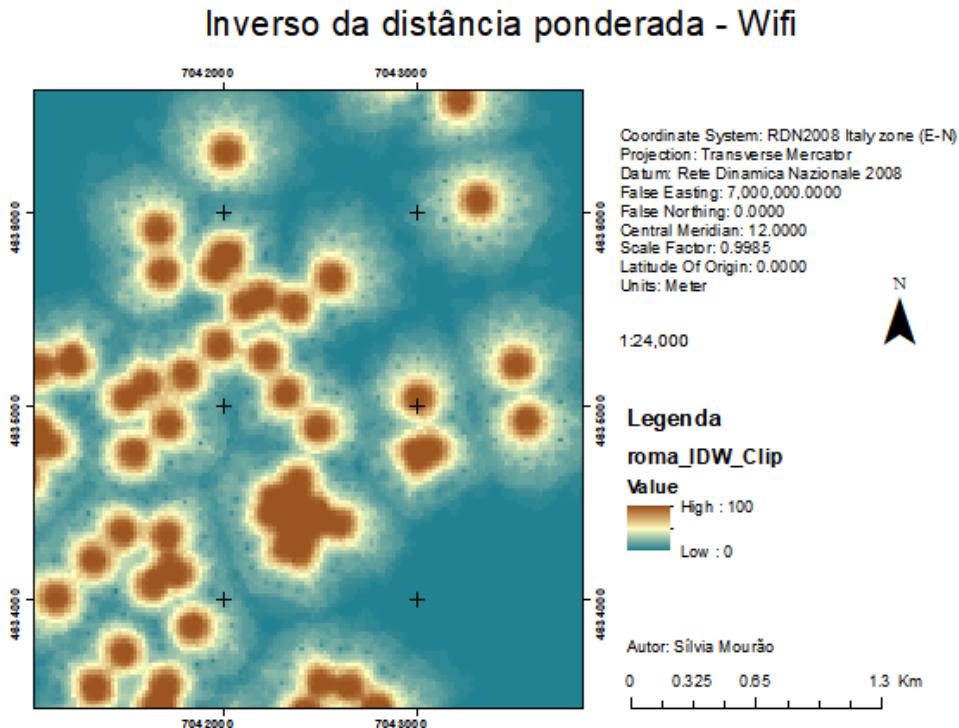


Figura 76 - Detalhe da interpolação pelo método do inverso da distância ponderada para o sinal WiFi

O resultado obtido demonstrado nestas imagens corresponde a uma variação do sinal que é bastante constante numa zona central, mas que se comporta de uma forma bastante estranha na periferia da zona de influência do ponto de acesso,

O último método de interpolação considerado nesta fase de interpoladores determinísticos locais foi o da Média Móvel Espacial (Radial Basis). Este interpolador cria um círculo num raio definido nos parâmetros de interpolação, calculando depois o ponto a ser interpolado a partir da média de todos os valores que se encontram dentro do círculo. Para este caso foi definido um círculo com um raio de 200 m, sendo os resultados obtidos visíveis na figura 77 e mais uma vez uma visão mais detalhada da zona central na figura 78:

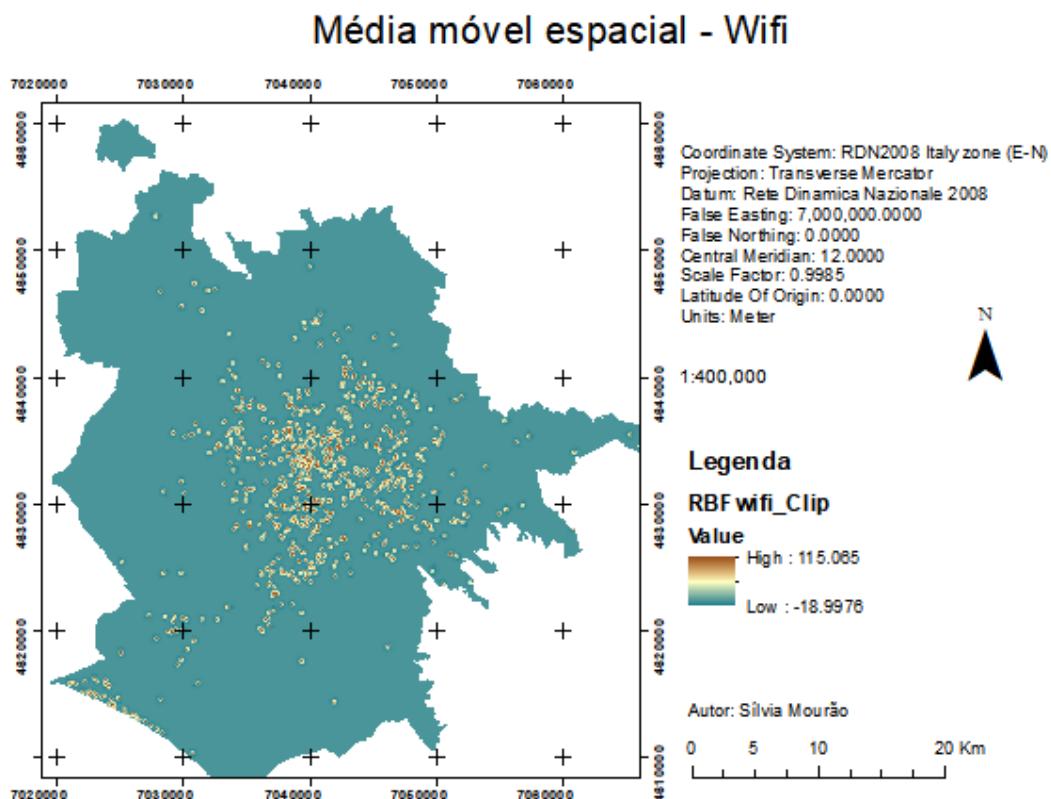


Figura 77 - Interpolação pelo método da média móvel espacial para o sinal WiFi

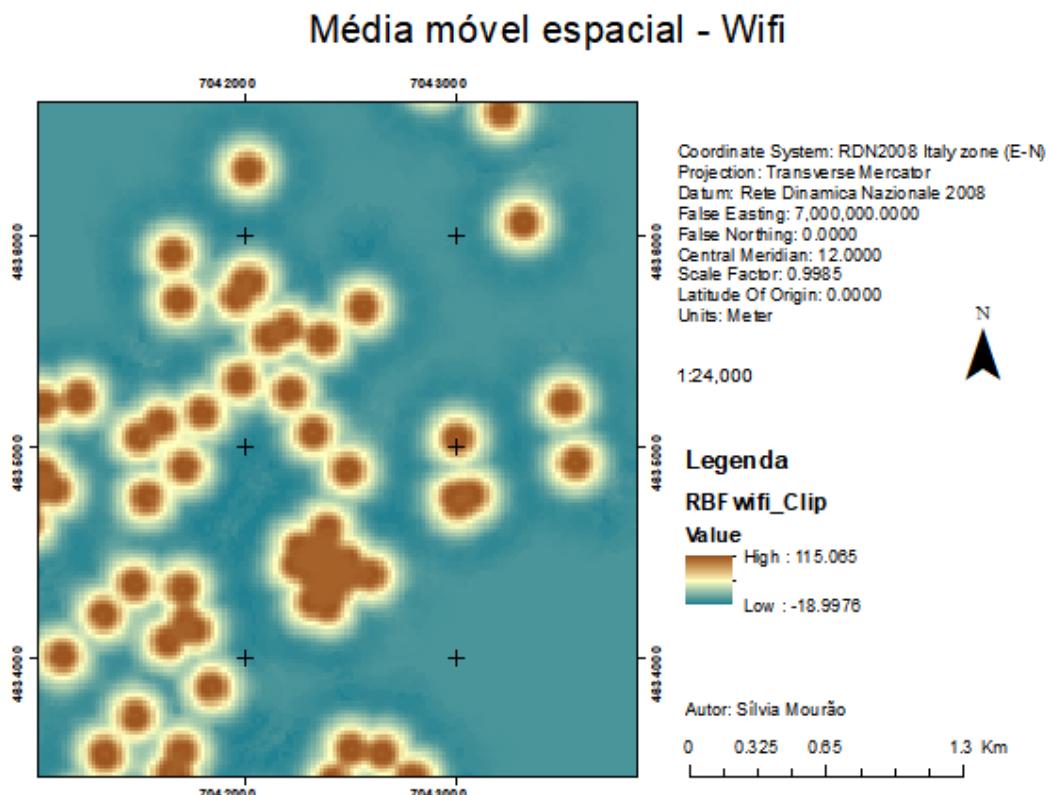


Figura 78 – Detalhe da interpolação pelo método da média móvel espacial para o sinal WiFi

Os resultados obtidos por este método são semelhantes aos obtidos pelo interpolador IDW no entanto este gera zonas de sinal negativo nas fronteiras entre a região de influência dos pontos de acesso e a região onde não existe sinal.

### 7.3 Métodos de Interpolação Probabilísticos (Kriging) - Sinal WiFi

Os métodos de interpolação probabilística, tal como os determinísticos, utilizam médias ponderadas dos valores da vizinhança, no entanto estes têm em consideração a existência de correlação espacial, utilizando a matriz de covariâncias para atribuir pesos a essa vizinhança e não apenas a distância ao ponto a estimar. Estes modelos têm também a possibilidade de considerar o efeito de anisotropia (tendência de um fenómeno não se propagar de igual modo em todas as direções).

O Kriging é um exemplo de método de interpolação probabilística, que segue um processo de quatro etapas para conseguir interpolar uma superfície:

- Análise exploratória dos dados

- Analise estrutural (modelação do semi-variograma experimental)
- Ajuste do modelo teórico
- Interpolação estatística pelo método de Kriging para produzir a superfície.

Um variograma é uma apresentação visual da covariância exibida entre cada par de pontos dos dados amostrados. Para cada par de pontos o valor de semi-variância é graficado contra a distância entre eles. Este variograma experimental corresponde aos valores observados, sendo a fase seguinte a da procura de um modelo teórico que melhor se ajuste aos dados observados (Waller & Gotway, 2004). Os parâmetros utilizados para realizar o Kriging para estes dados foram mais uma vez determinados a partir dos valores previstos de EMQ, tendo sido utilizado o modelo ordinário com a função de Epanechnikov, que corresponde a uma função parabólica descontínua. Na figura 79 podemos ver o semi-variograma obtido, tendo sido utilizado apenas um modelo do tipo estável fazer a aproximação, visível na figura 80, tendo mais uma vez sido mapeada uma vista detalhada na figura 81:

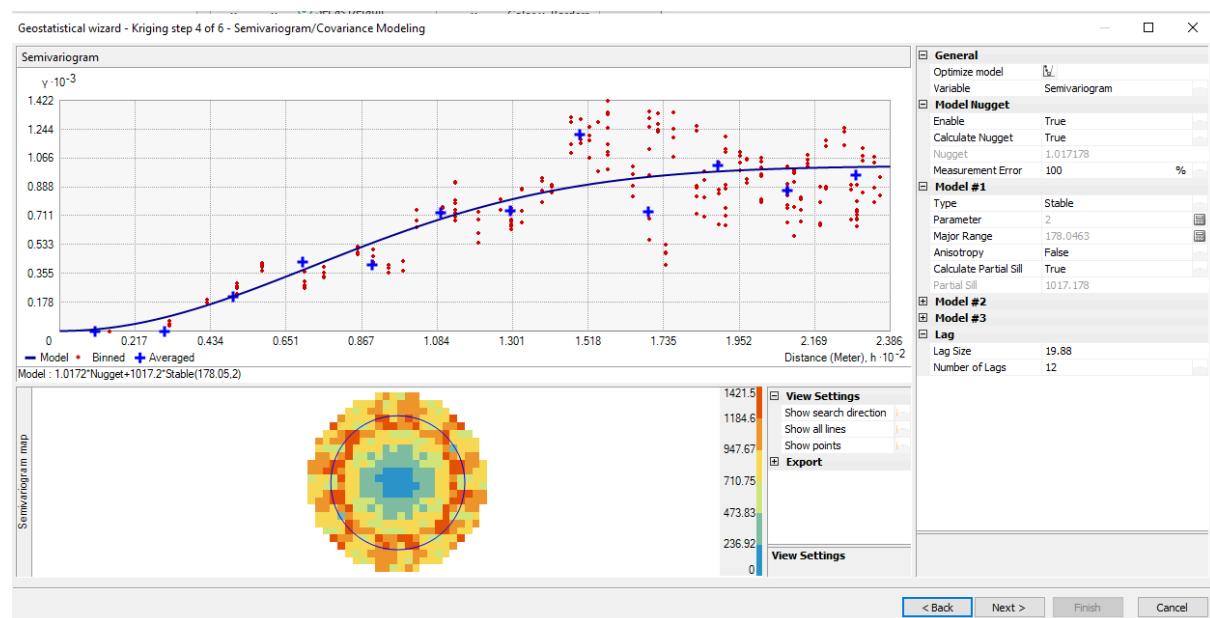


Figura 79 - Semi-variograma e aproximação do modelo

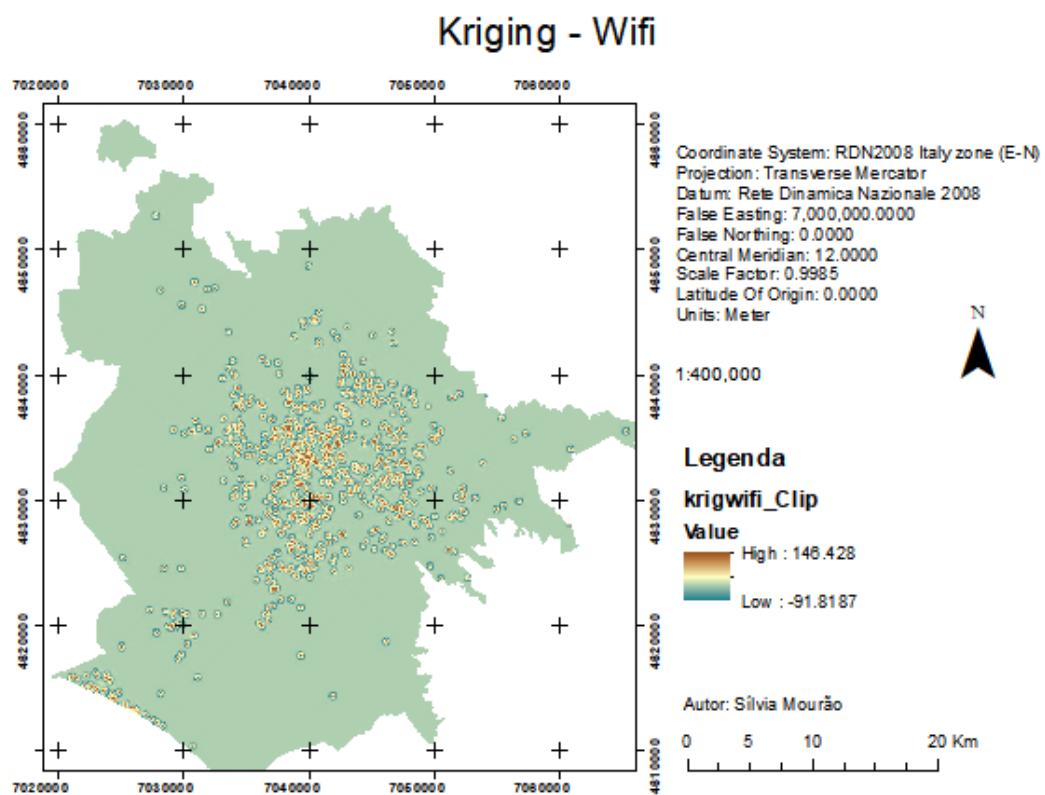


Figura 80 - Interpolação pelo método Kriging para o sinal WiFi

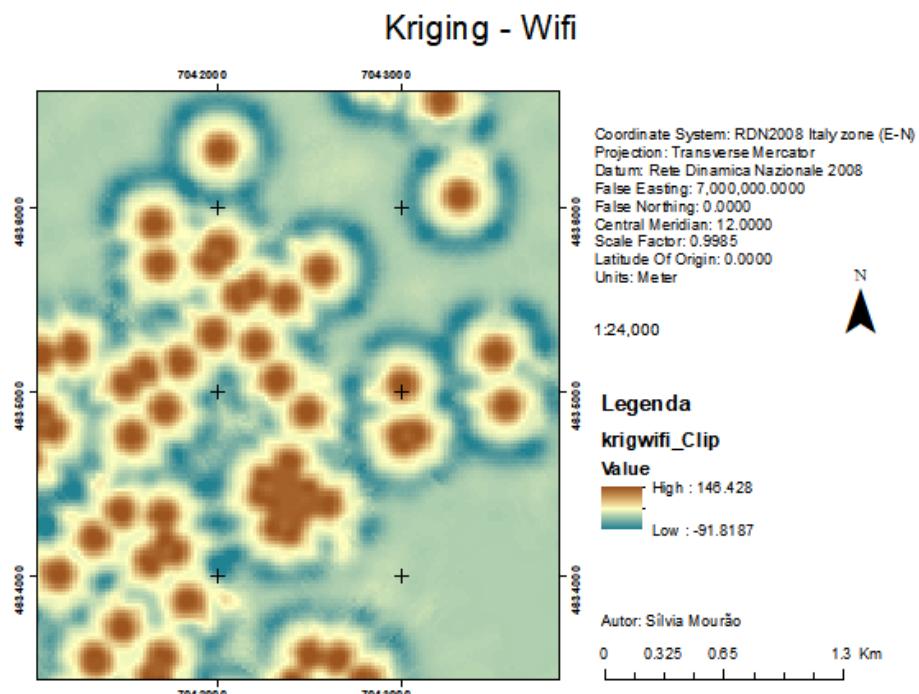


Figura 81 - Detalhe da interpolação pelo método Kriging para o sinal WiFi

Sílvia Mourão, FC 57541

Analisando as duas imagens obtidas, verifica-se que mais uma vez este interpolador tem tendência a criar zonas de intensidade de sinal negativa nas fronteiras entre as zonas de influência dos pontos de acesso e as zonas sem pontos de acesso.

#### 7.4 - Validação e análise comparativa dos resultados dos métodos de interpolação

A validação dos modelos foi feita a partir da amostra de pontos que tinha sido retirada do conjunto de dados antes do processo de interpolação e foi utilizado o ArcMap para extrair o valor calculado para esses pontos por cada um dos métodos interpoladores.

Os valores obtidos foram depois comparados com os valores observados, estando as conclusões dessa análise refletidas na tabela 14, onde EMQ Geostatistical diz respeito ao erro médio quadrático calculado pelo ArcMap e validado nos pontos da amostra, a média das diferenças é apenas uma medida de quão afastados os valores observados e estimados estão em média e o EMQ Set Validação corresponde então ao valor de erro médio quadrático encontrado analisando os pontos de validação.

*Tabela 14 - Comparação de resultados obtidos pelos vários métodos de interpolação para o sinal WiFi*

Método	GPI	LPI	IDW	RBF	Krig
<b>EMQ Geostatistical</b>	38.4106	27.1035	21.4239	<b>5.6027</b>	11.2248
<b>Média das Diferenças</b>	0.5889	95.7806	-3.0783	-0.2055	-1.7260
<b>EMQ Set Validação</b>	38.6113	972.6761	8.2340	<b>6.1078</b>	11.2836

Analisando esta tabela é possível concluir que o interpolador da média móvel espacial é o que consegue obter melhores resultados para a intensidade do sinal WiFi.

#### 7.5 - Produção de um Modelo Digital de Elevação a partir de métodos de interpolação

Um processo semelhante ao realizado para a intensidade do sinal WiFi foi realizado para a produção de um modelo digital de terreno. A principal alteração foi a adição da utilização de um modelo TIN, que foi também comparado com os outros métodos utilizados, e cujos resultados obtidos podem ser vistos nas figuras 82 e 83.

Após terem sido trabalhados todos os métodos, concluiu-se que o melhor resultado neste caso foi obtido para a interpolação do inverso da distância ponderada, figura 84. A comparação entre os vários modelos encontra-se na tabela 15.

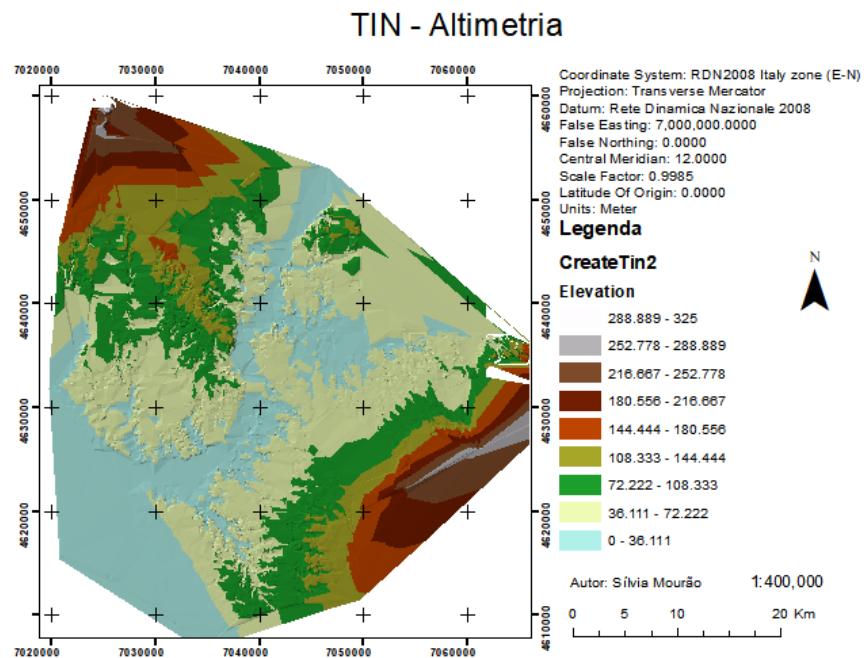


Figura 82 - Geração de MDT a partir de uma TIN

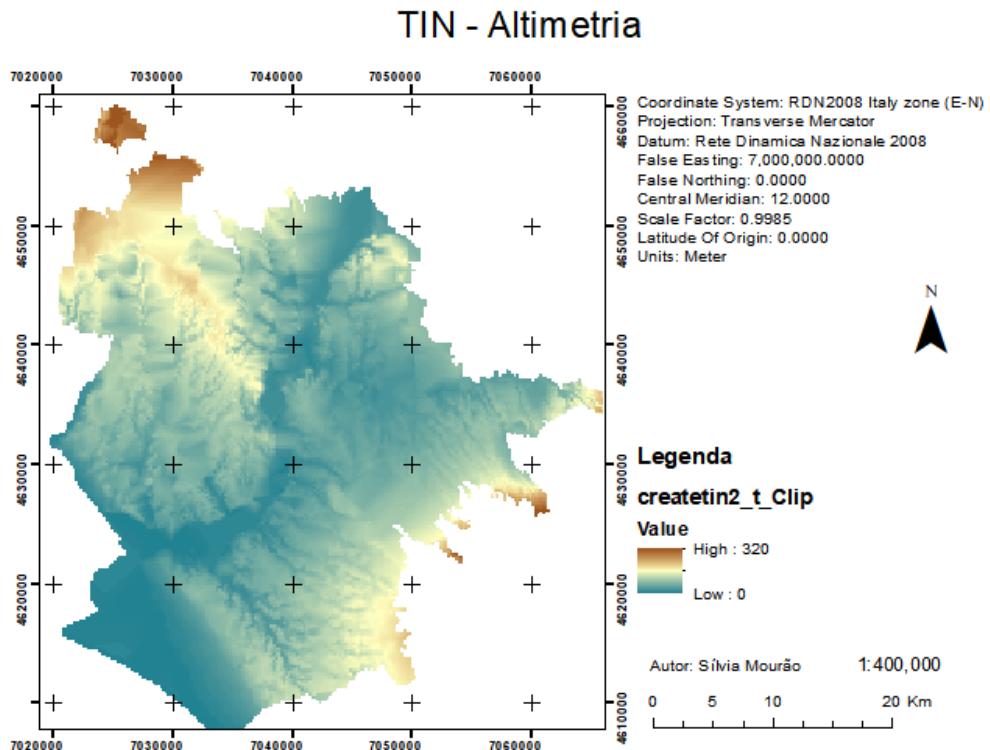


Figura 83 - Interpolação de altimetria resultante da TIN

### Inverso da distância ponderada - Altimetria

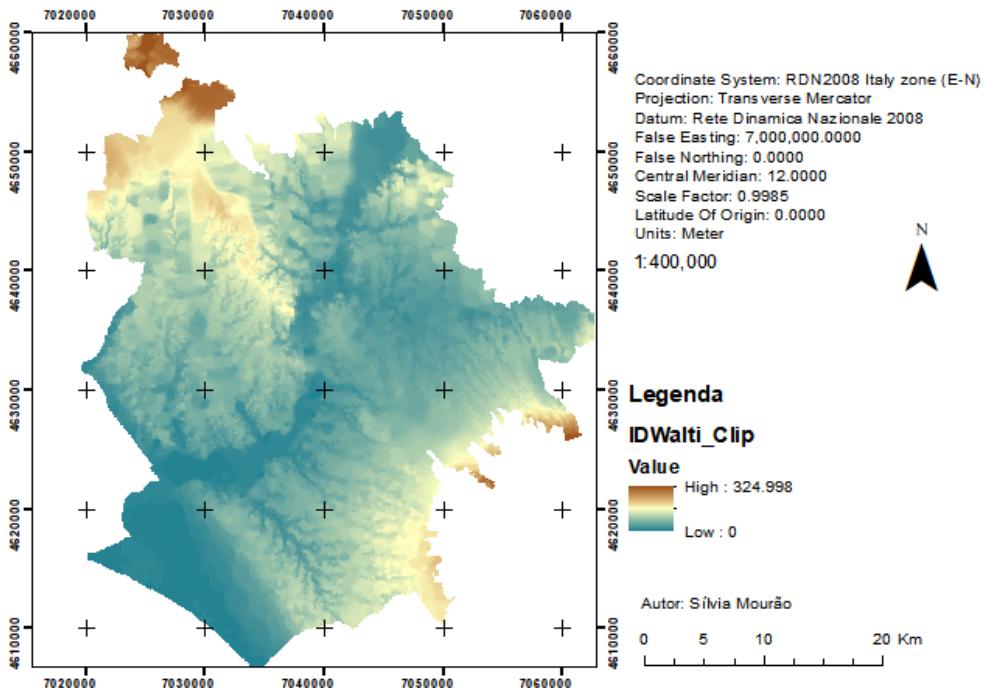


Figura 84 - Interpolação da altimetria pelo inverso da distância ponderada

Tabela 15 - Comparação de resultados obtidos pelos vários métodos de interpolação para a altimetria

Método	GPI	LPI	IDW	RBF	Krig	TIN
<b>EMQ Geostatistical</b>	14.0854	2.6324	<b>0.4013</b>	0.9333	0.7812	
<b>Média das Diferenças</b>	-0.0264	20.5448	<b>-0.0181</b>	-0.0196	-0.0262	2.3190
<b>EMQ Set Validação</b>	14.0252	456.2546	<b>1.5048</b>	1.5839	2.3437	166.3558

O modelo obtido através da interpolação do inverso da distância ponderada foi então utilizado para criar mapas de slope (figura 85), aspect (figura 86) e hillshade (figura 87), de forma a ser possível compreender a inclinação do terreno, a sua orientação cardinal e zonas de sombra.

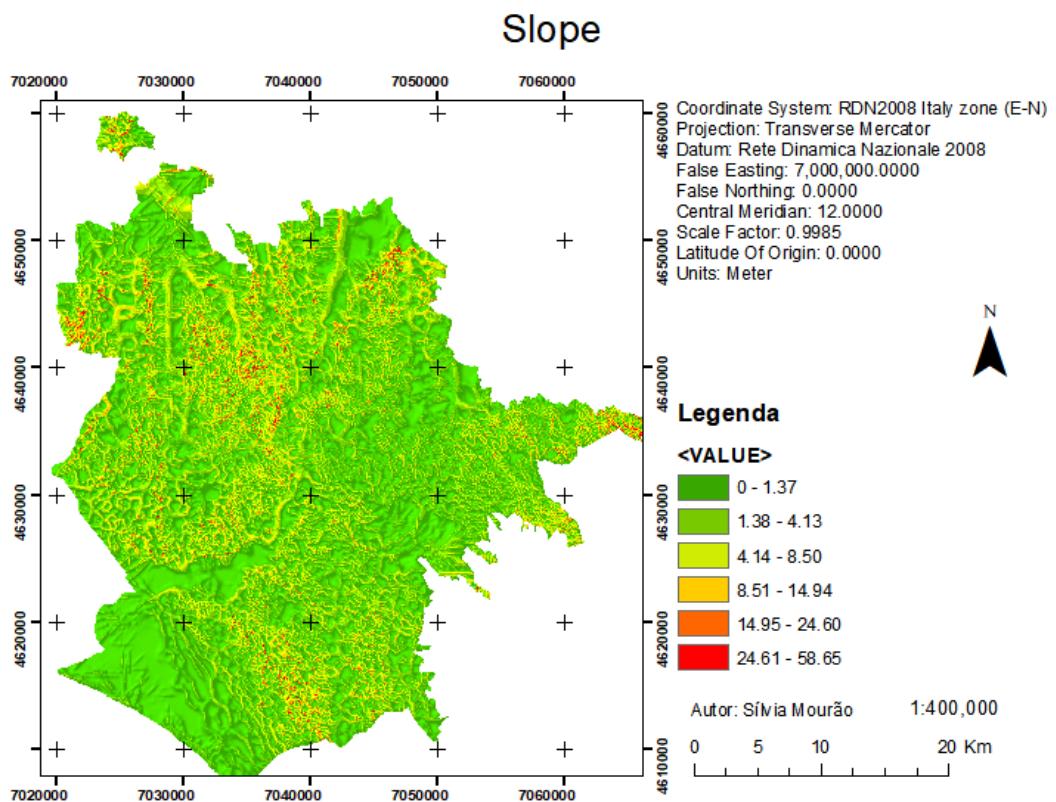


Figura 85 - Inclinação do terreno

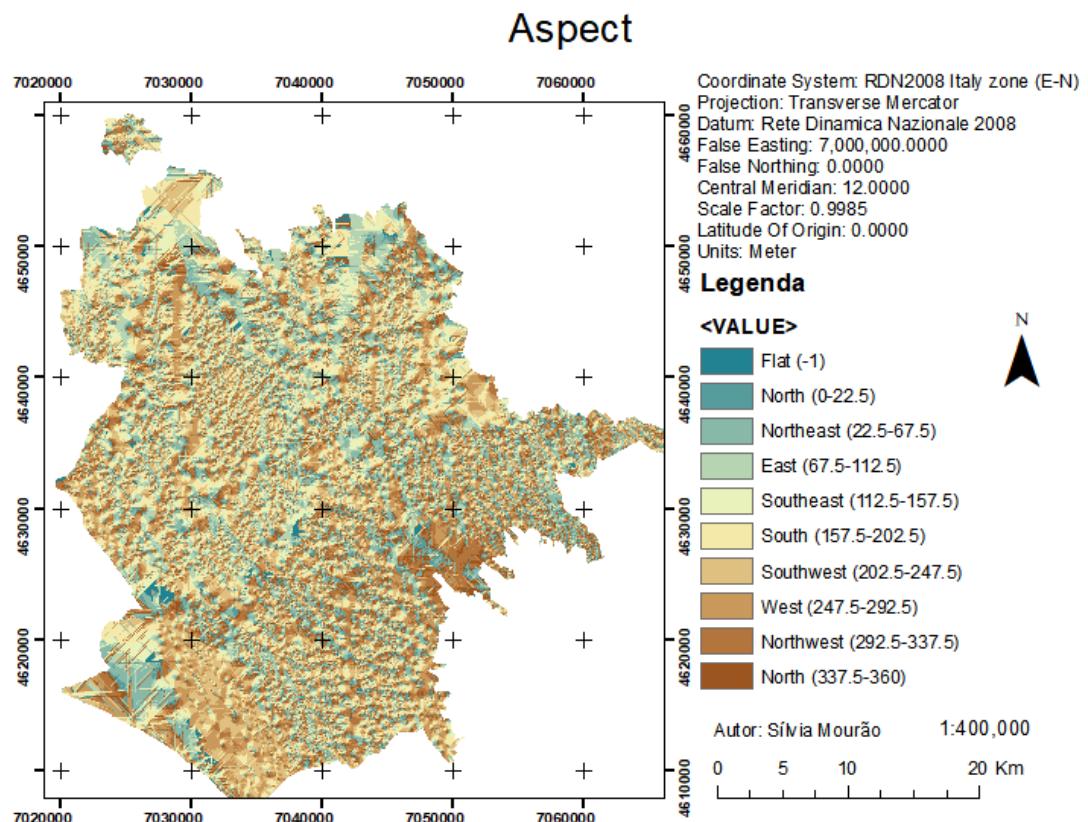


Figura 86 - Orientação do relevo do terreno

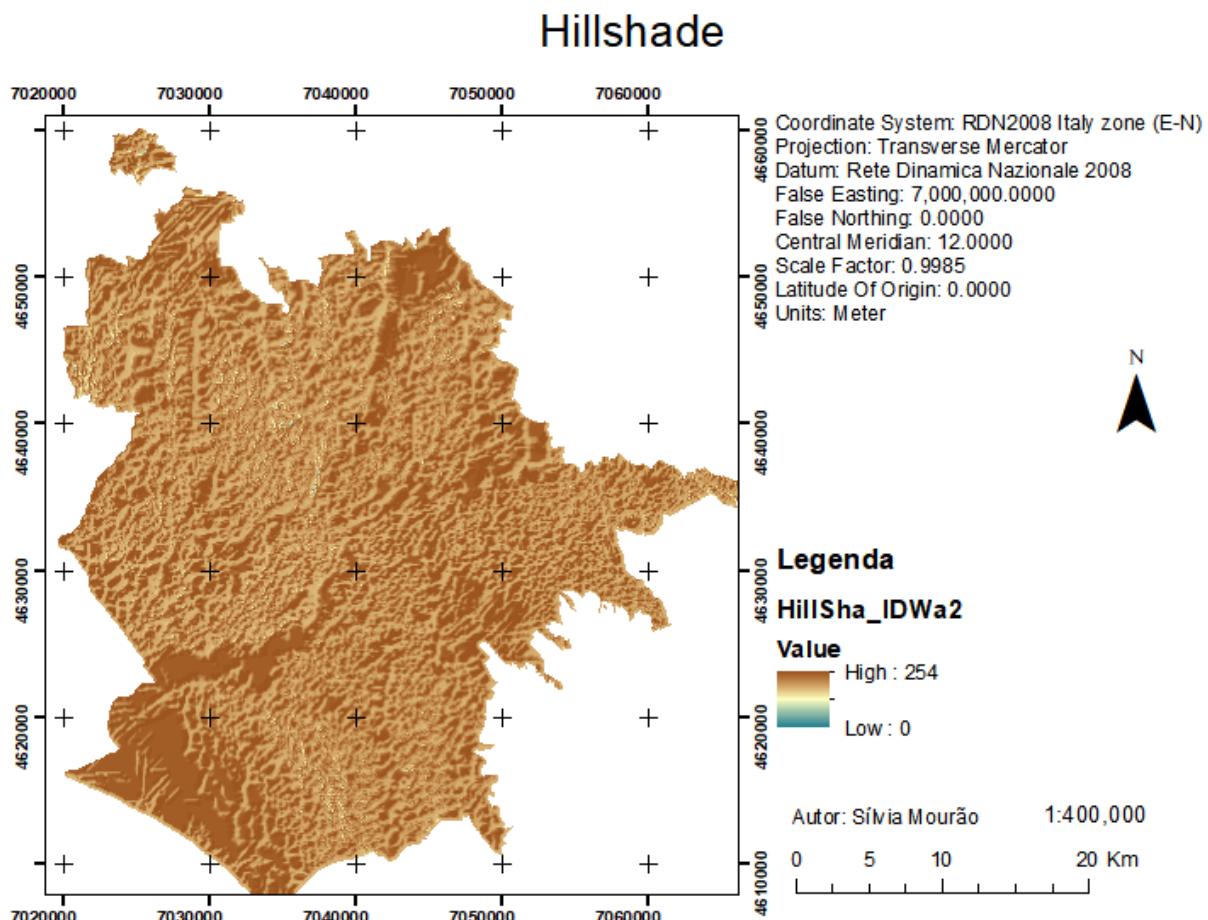


Figura 87 - Zonas de sombra

A criação do modelo digital de terreno teve como objetivo estudar se a variação altimétrica na zona de estudo para a próxima fase do projeto necessitará da utilização da elevação como fator de impedância para a rede. Essa zona consistirá apenas na zona histórica de Roma, onde se encontram a maioria das atrações turísticas, por isso interessa saber como é a altimetria nessa zona.

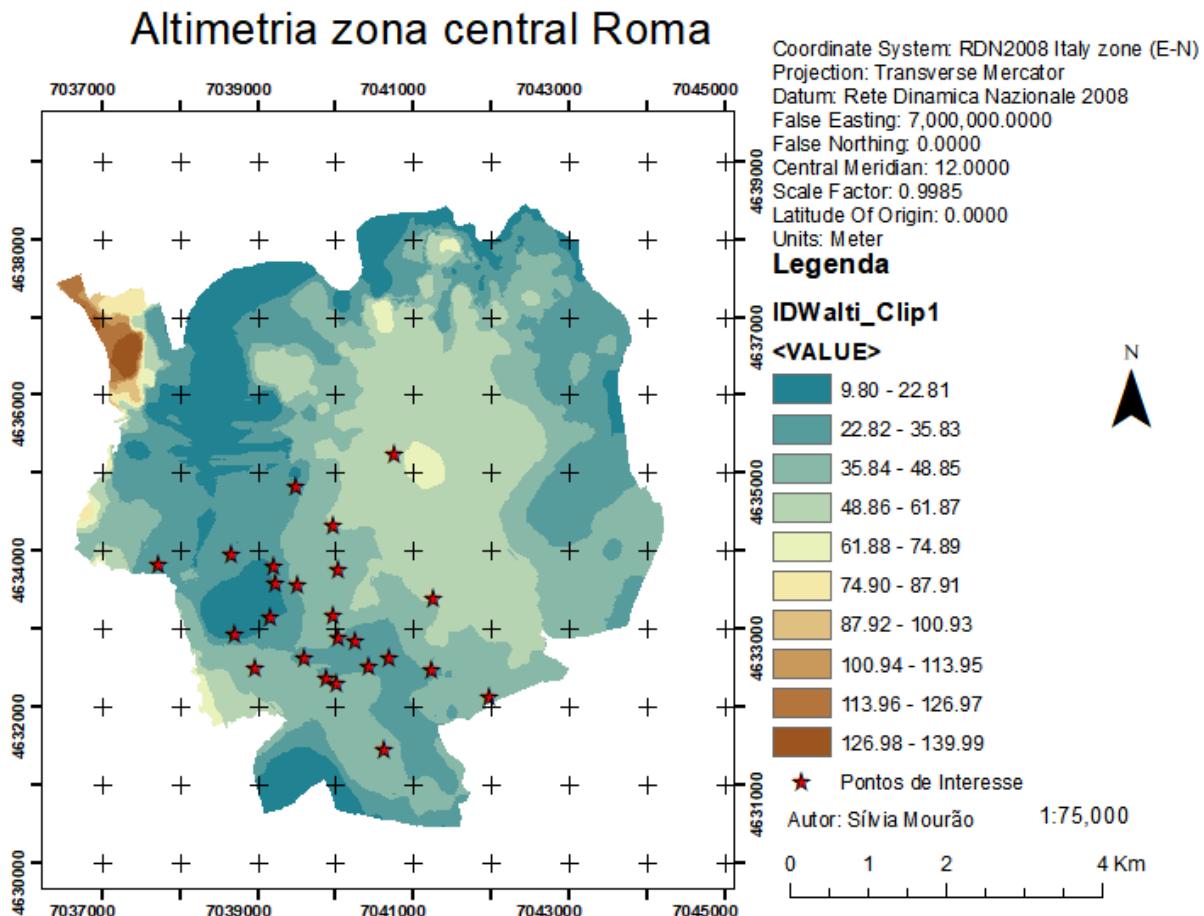


Figura 88 - Estudo da altimetria nas localizações de pontos de interesse turísticos

Por análise de figura 88 verifica-se que a grande maioria dos pontos se encontram nas regiões de 22.82-48.85 o que indica uma diferença altimétrica de cerca de 25 m, no entanto é necessário notar que a interpolação realizada corresponde ao modelo digital de terreno que nem sempre se traduz na localização exata dos pontos, especialmente na zona do rio, que por ter uma diferença de altitude grande esta a influenciar a interpolação da altimetria nessa zona, no entanto o trânsito entre as margens do rio é feito através de pontes que se encontram a uma altitude semelhante á das margens. Um exemplo disso é demonstrado em seguida, correspondendo a figura 89 à localização e detalhe altimétrico interpolado para a zona da ilha no rio Tíber onde se localiza a Basílica de São Bartolomeu, e a figura 90 mostra a imagem real no terreno da ilha.

## Altimetria zona central Roma

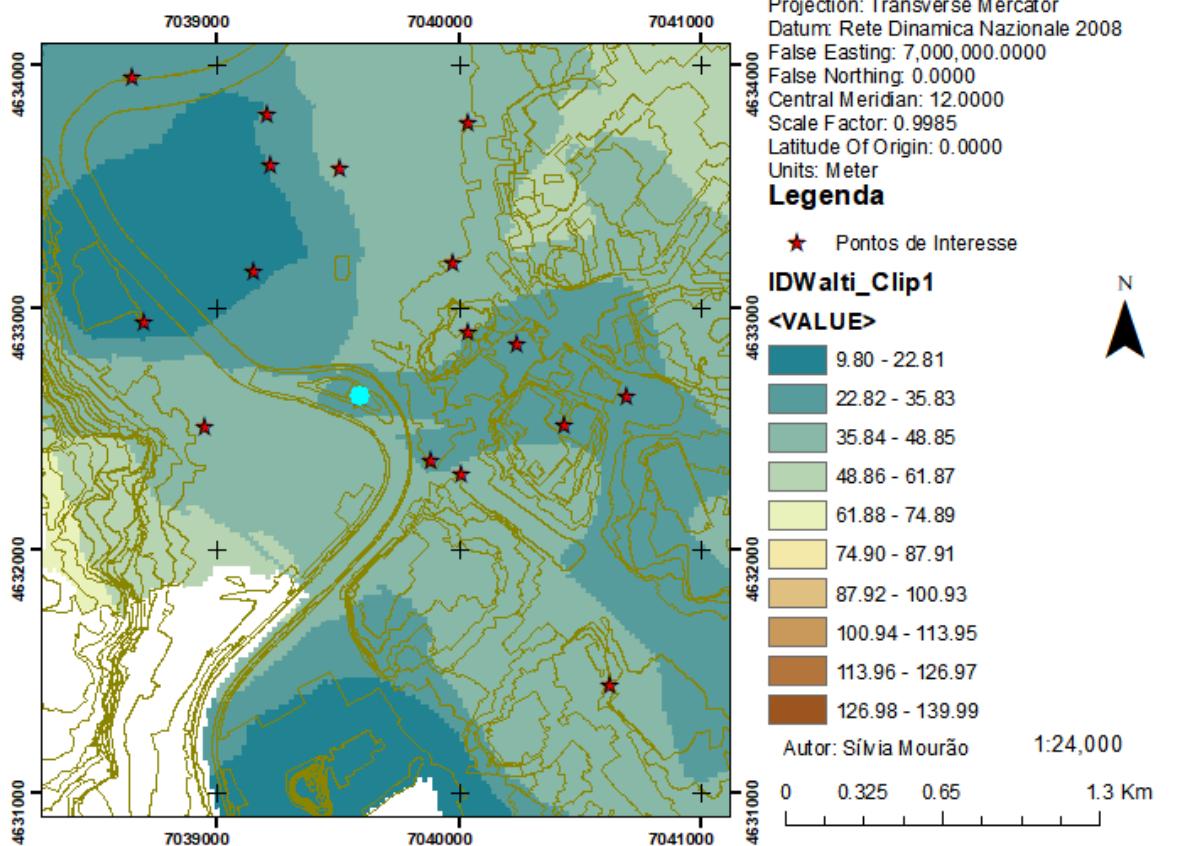


Figura 89 - Detalhe de altimetria na zona da ilha tiberina



Figura 90 - Fotografia da localização na realidade

Assim sendo, a altimetria não vai ser utilizada como medida de impedância para as redes.

## 8. Análise de Redes

A análise de redes baseia-se na teoria de grafos e na topologia. Uma rede é um sistema de elementos interligados, constituída por arcos (linhas) e juncões de ligação (nós) que representam possíveis rotas de uma localização para outra. Através de algoritmos de análise de redes é possível utilizar estas redes para responder a perguntas como qual é o melhor caminho entre um ponto A e um ponto B ou qual é a área de cobertura de um hospital (Catita, 2022).

Neste projeto o foco foi mais uma vez em torno da cobertura WiFi no âmbito do turismo, nomeadamente procurar um caminho ótimo entre pontos turísticos que tenha a maior cobertura WiFi possível, encontrar áreas onde existem lacunas de sinal e perceber qual é o ponto de acesso mais próximo de cada ponto turístico.

### 8.1 Recolha e Tratamento de Dados

Em primeiro lugar foi feita uma recolha dos dados das vias rodoviárias da região de estudo com recurso ao Open Street Map. Essa rede teve de ser alterada, visto que o principal desafio no que toca à análise de redes é a representação de uma rede de arcos-nós. A rede obtida mostra elementos que pertencem num mapa, mas que não são necessários para um grafo utilizado no algoritmo de redes, como faixas de rodagem múltiplas, bifurcações na aproximação a rotundas, etc. Todos estes elementos podem ser modelados para análise de redes num grafo utilizando apenas arcos, nós e uma turn-table. A abordagem utilizada para tornar a rede obtida numa aproximação a um grafo é descrita mais detalhadamente no capítulo seguinte.

Para o conjunto de dados de monumentos, foi utilizado um dataset que contém o top 25 de monumentos mais visitados em Roma. As coordenadas foram obtidas em WGS84 e depois projetadas para a RDN 2008. Analisando inicialmente os dados foram excluídos dois pontos do dataset, que correspondem a zonas muito afastadas da zona central. Dos 23 pontos restantes, um deles foi depois excluído numa fase posterior por se encontrar ainda numa zona mais remota comparativamente aos outros pontos considerados. Para colmatar a eliminação destes três pontos foram acrescentados três pontos turísticos mais próximos dos 22 que permaneceram no dataset: Bocca della Verità, Basilica São Bartolomeu e Circo Massimo. Um outro detalhe a considerar é a existência de 3 pontos que estão

Sílvia Mourão, FC 57541

situados dentro do Vaticano, que embora seja um país diferente está contido nesta região central de Roma. A rede viária na zona do Vaticano estava bastante incompleta e por isso os 3 pontos foram condensados num só e tomou-se esse ponto como origem para a rota.

## 8.2 Criação de Topologia

Tendo primeiramente em conta que o objetivo desta análise é para caminhos pedonais, foram primeiramente eliminados da shapefile todos os caminhos que correspondem a zonas onde apenas podem circular veículos, como autoestradas.

Em seguida foram estabelecidas regras de topologia para o modelo, de forma a que estas pudesse ser analisadas individualmente e gerada uma rede onde não existissem ambiguidades, mas que não será possível aproximar a um grafo puro devido ao processo moroso que é requerido para tal.

As regras de topologia criadas foram as seguintes:

- Não podem existir sobreposições
- Não podem existir dangles – nós que têm apenas um arco ligado
- Não podem existir pseudo-nós – nós que têm apenas dois arcoss ligados
- Não podem existir intersecções ou sobreposições de um arco consigo próprio.

Os pontos de interesse considerados foram importados para o QGIS, onde se criaram círculos de 1.5km em torno de cada monumento, tendo esta shapefile depois sido importada para o ArcMap e utilizada para recortar as estradas apenas na área de interesse para o projeto.

Para facilitar a correção dos erros foi utilizada a barra de ferramentas da topologia e foi sobreposto um basemap do Open Street Map, de forma a que se possa entender por exemplo, quando é que um “dangle” corresponde a uma rua sem saída e quando é que é um erro. Nos casos em que a topologia não precisava de ser corrigida eram acrescentadas exceções às regras. Um exemplo da correção da topologia pode ser visto na figura 91.

## Erros de Topologia - Exemplo

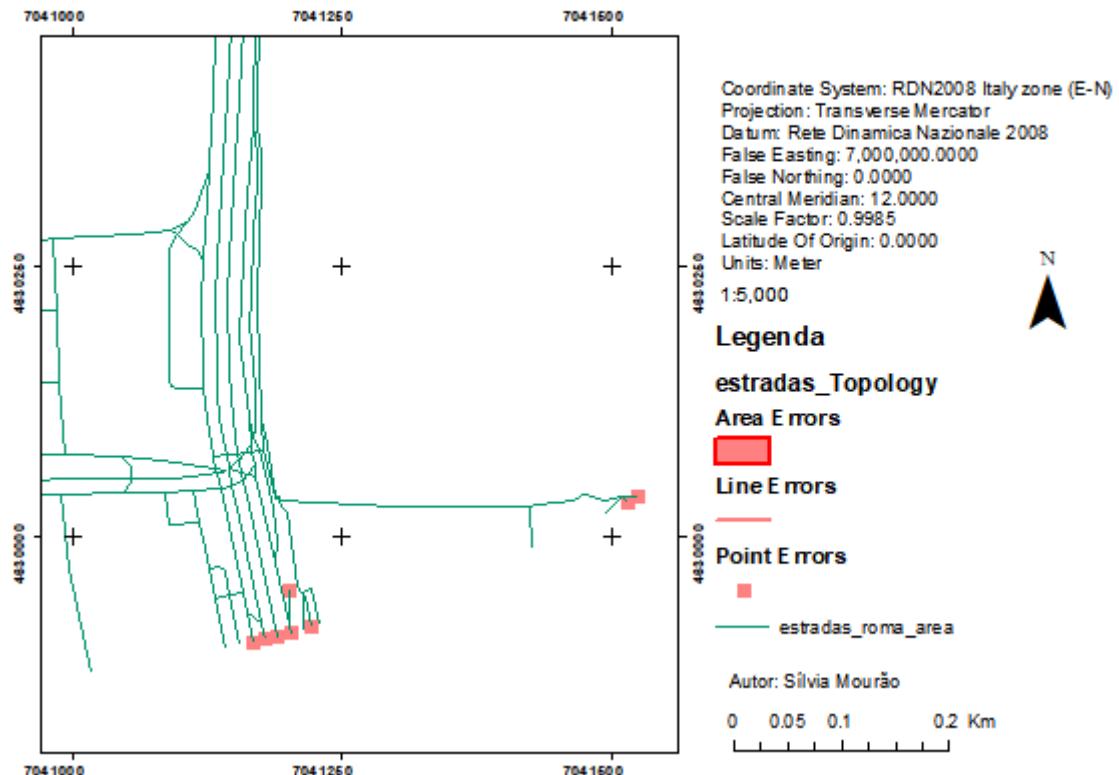


Figura 91 - Correção de erros de topologia

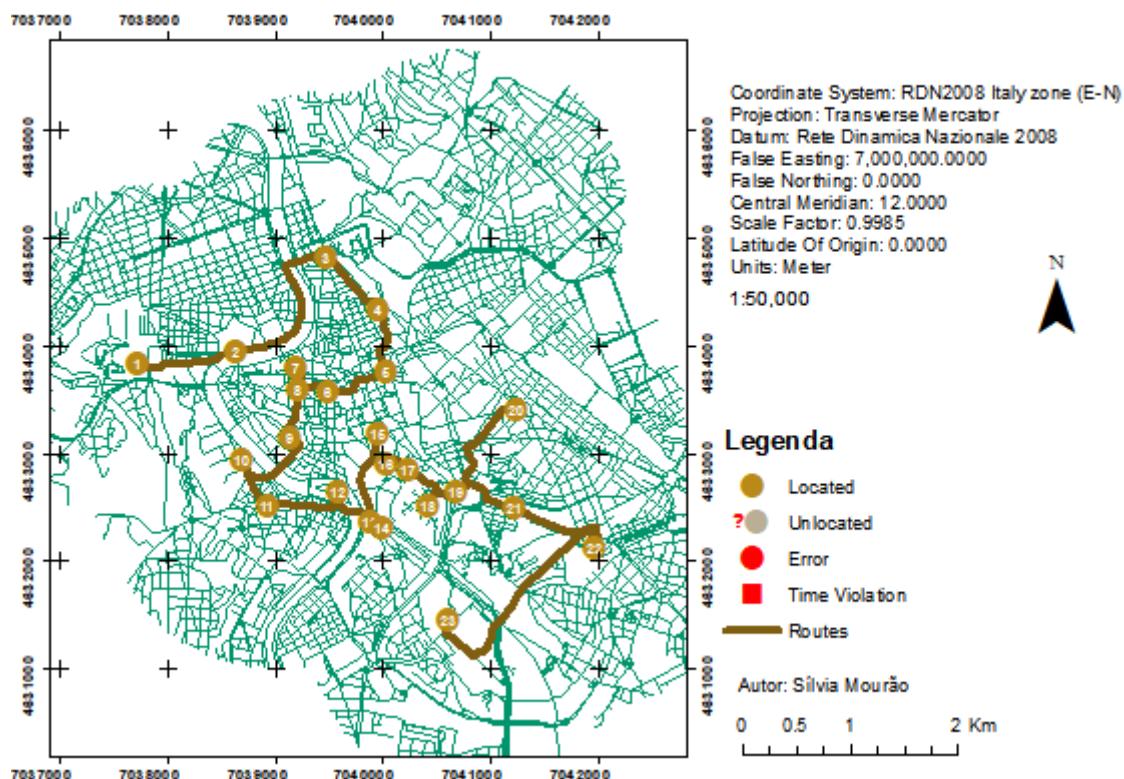
Depois de terem sido corrigidos todos os erros, a rede estava então pronta para ser utilizada nos algoritmos de redes.

### 8.3 Optimal Path

O algoritmo de caminho ótimo tem por base determinar o caminho ótimo numa rede entre uma origem e um destino, conectando um conjunto de pontos com custo mínimo. O custo pode ser definido em função de quaisquer variáveis que se queiram acrescentar à rede, na forma de impedância, sendo o mais simples a utilização apenas do comprimento do arco e podendo modelos mais complexos incluir a velocidade média num troço, a elevação, o tempo médio de percurso, entre muitos outros.

Para este trabalho foi inicialmente definida a rota de pontos turísticos ideal, considerando apenas o comprimento como medida de impedância, que pode ser vista na imagem 92.

## Rota Pontos Turísticos Roma



*Figura 92 - Rota de pontos turísticos utilizando apenas o comprimento como custo*

Foi depois definida uma nova medida de impedância baseada na presença ou não de sinal WiFi durante o caminho, tendo sido definido um custo 0 para uma estrada que está sempre coberta por sinal WiFi, um custo 1 para uma estrada parcialmente coberta e um custo 2 para uma estrada sem cobertura. Utilizando a nova medida de impedância foi calculada novamente a rota ótima entre pontos de interesse, visível na figura 93:

## Rota Pontos Turísticos Roma

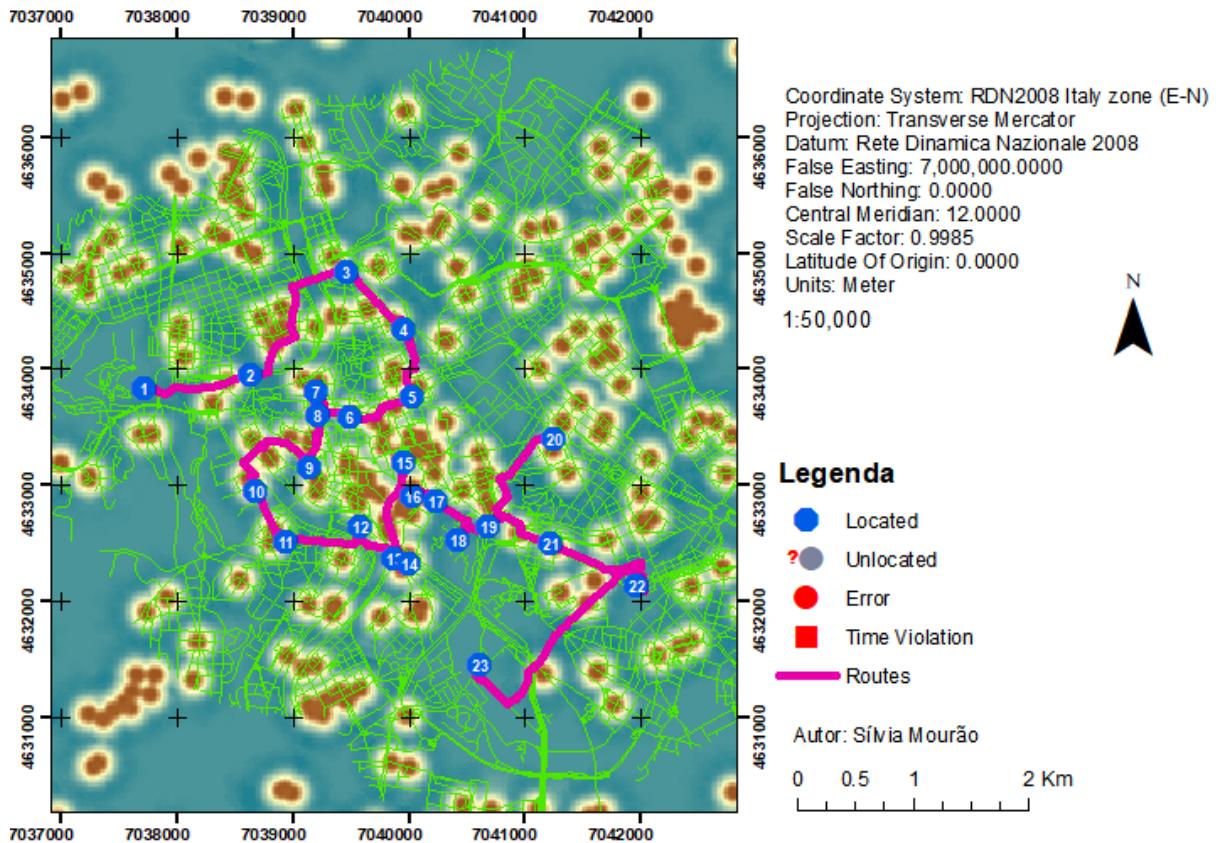


Figura 93 - Rota de pontos turísticos utilizando o sinal WiFi como custo, sobreposto na interpolação de sinal WiFi obtida pela média móvel espacial

Analisando as propriedades das rotas obtidas é possível fazer uma comparação entre os valores obtidos como comprimento total da rota e “contador de WiFi” e comparar as duas soluções:

Attribute	Value	Attribute	Value
ObjectID	3	ObjectID	2
Name	Location 1 - Location 23	Name	Location 1 - Location 23
FirstStopID	1	FirstStopID	1
LastStopID	23	LastStopID	23
StopCount	23	StopCount	23
Total_Wifi	286	Total_Wifi	201
Total_Length	18379.769637	Total_Length	23576.38651

Figura 94 - Comparação das duas soluções da rede

Na figura 94 podemos observar à esquerda os valores obtidos para o caso da primeira rota e à direita da segunda rota. Com base nestes resultados é possível concluir que a rede WiFi pode ainda ser

Sílvia Mourão, FC 57541

melhorada, visto que existe um aumento da distância percorrida entre pontos de interesse de cerca de 5 km, sendo acumulados apenas menos 85 pontos negativos por atravessar zonas sem cobertura.

#### 8.4 - Service Area

A ferramenta da área de serviço define um polígono que agrupa os locais que podem ser alcançados sobre um arco da rede a menos de um determinado custo, tendo neste caso sido utilizada as distâncias de propagação do sinal WiFi de 50, 100 e 150 m.

A área representada na figura corresponde à zona central de Roma vista a uma maior escala, sendo visíveis claras lacunas na cobertura de sinal. A zona com melhor cobertura aparenta ser a zona da margem norte do rio junto à ilha.

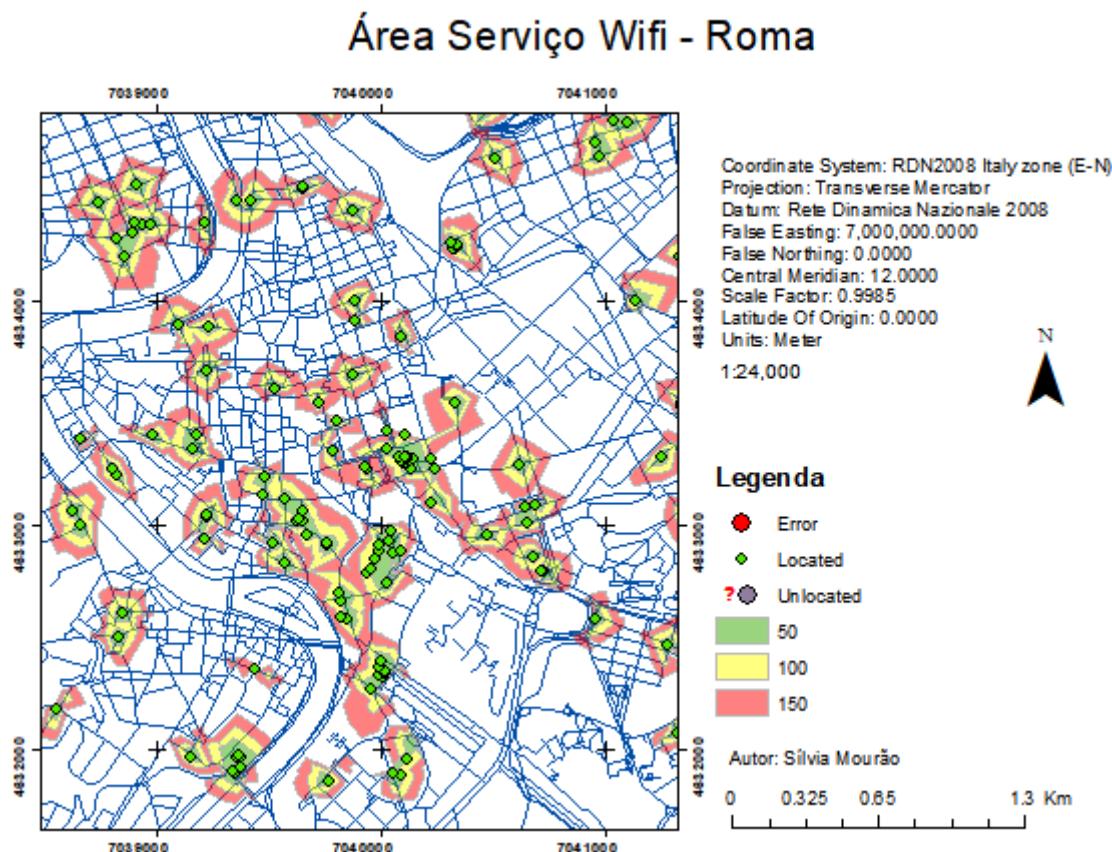


Figura 95 - Área de serviço dos pontos de acesso

#### 8.5 - Closest Facility

Finalmente, utilizou-se o algoritmo de closest facility, que consiste em identificar o trajeto para um ponto mais próximo a partir de um outro tipo diferente de ponto. Para esta fase foram considerados mais uma vez os pontos de acesso WiFi e os pontos de interesse turístico, tendo sido procurado o ponto de acesso mais próximo de cada monumento. Os resultados são visíveis nas figuras 96 e 97.

### Acesso WiFi Mais Próximo - Roma

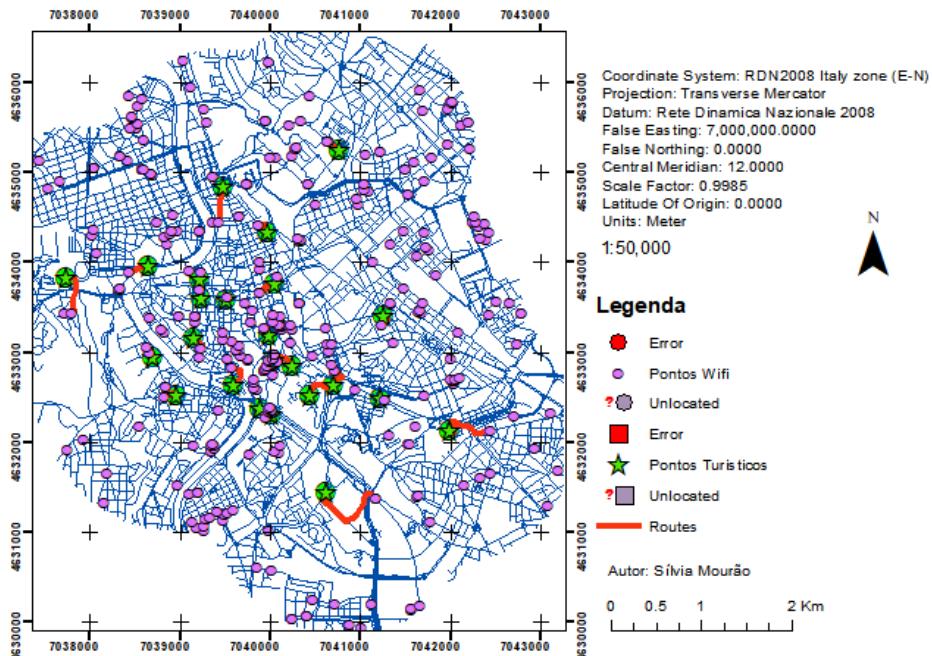


Figura 96 - Pesquisa do ponto de acesso WiFi mais próximo de cada ponto de interesse turístico

### Acesso WiFi Mais Próximo - Roma

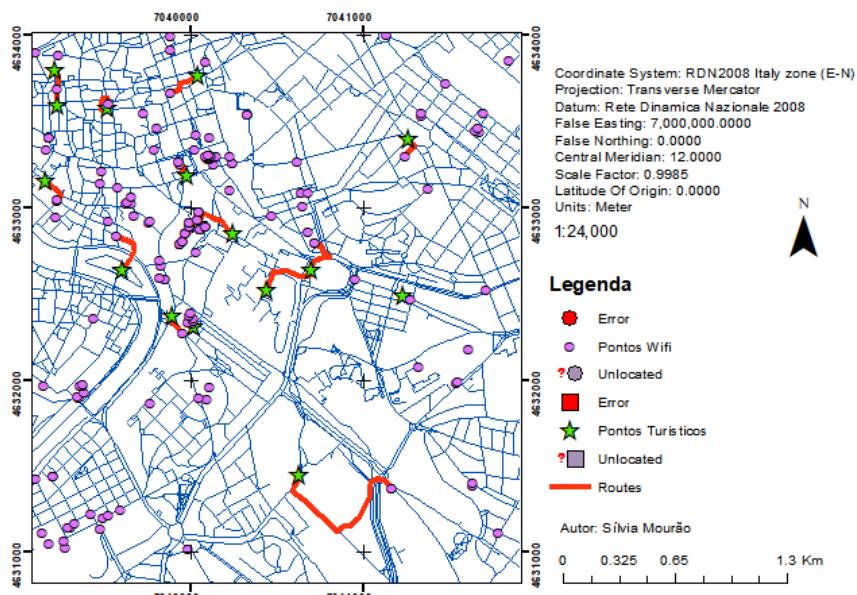


Figura 97 - Detalhe da pesquisa de acesso WiFi mais próximo

Analisando os resultados obtidos através das propriedades das rotas criadas, existe em média uma distância de 255.93 m entre um ponto de acesso e um ponto de interesse turístico, o que é muito superior ao alcance do sinal WiFi.

A maior distância encontrada nesta zona entre pontos é de 987 m, visível na figura 98.

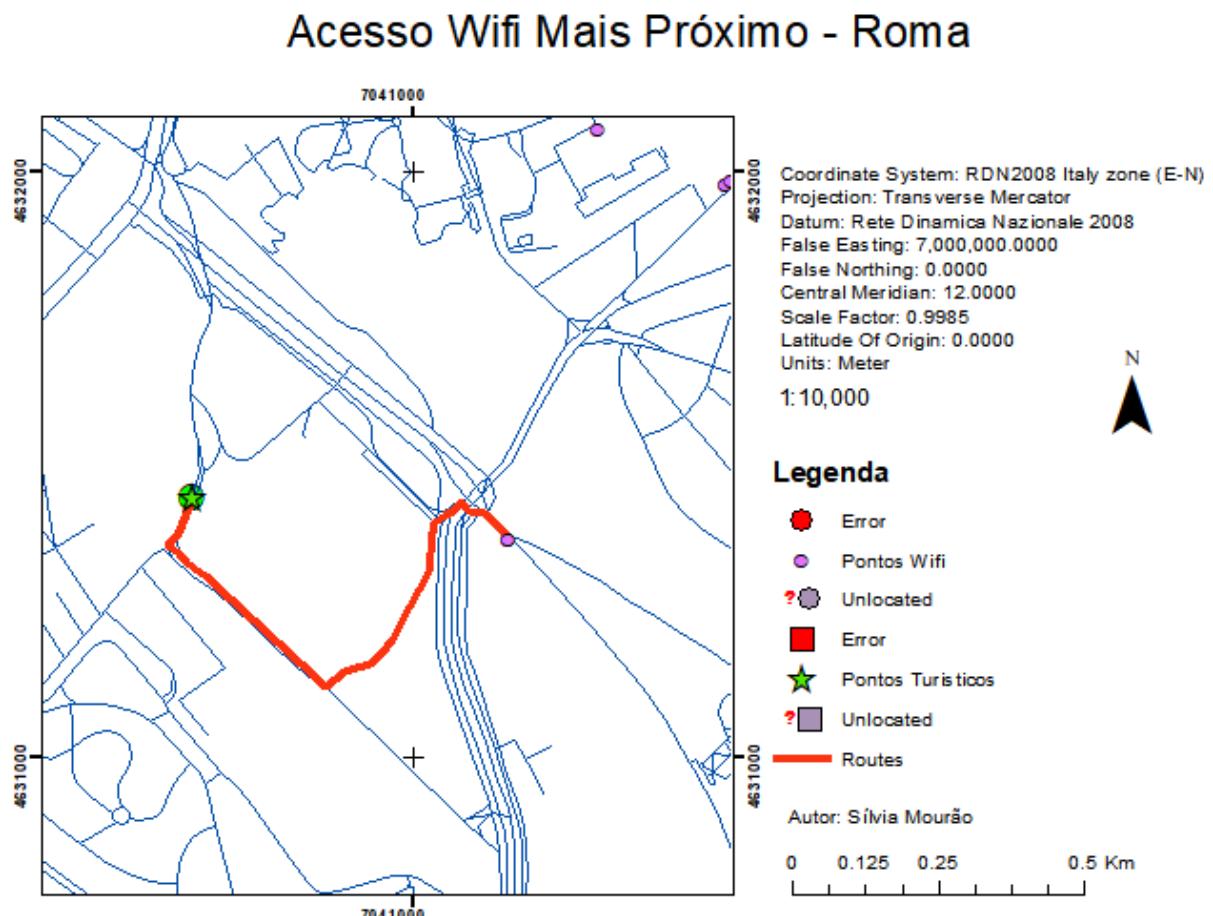


Figura 98 - Distância mais longa a um ponto de acesso

Dos 24 pontos turísticos considerados para este trabalho, apenas 5 pontos se encontram a menos de 100 m de um ponto de acesso e outros 5 pontos encontram-se entre os 100 e os 150m de distância de um ponto de acesso WiFi.

## 9. Conclusões

Este trabalho foi desenvolvido no âmbito da disciplina de Análise Espacial de Dados e permitiu uma melhor compreensão e utilização prática dos métodos estudados em aula para uma aplicação real. Ao longo do trabalho foi bastante clara a forma como estas ferramentas de análise podem permitir a percepção e interpretação de variáveis com distribuição espacial, sendo de particular grande relevância na tomada de decisões, por exemplo, de onde é necessário colocar novos pontos de acesso WiFi ou quais as regiões que já têm uma boa distribuição de sinal e aquelas que precisam ainda de maior investimento.

Na etapa da análise do padrão pontual confirmou-se a existência de um padrão aglomerado para os pontos de acesso WiFi em Itália, sendo que estes se distribuem maioritariamente em torno de cidades com população elevada. Na região da Lazio a distribuição espacial observada concentra-se maioritariamente na zona da cidade de Roma, pelo que embora seja uma das regiões com maior densidade de pontos a sua distribuição é muito irregular.

Na etapa da análise de áreas verificou-se que existe uma autocorrelação espacial positiva na variável em estudo, existindo uma tendência predominante para clusters de valores elevados na zona central, que corresponde à área metropolitana de Roma, e tornando-se numa tendência de clusters de valores baixos à medida que nos aproximamos da periferia da região.

Através dos modelos de regressão foi possível verificar que existem quatro variáveis independentes que conseguem prever com elevada precisão o número de pontos de acesso WiFi numa determinada área. Esta relação foi primeiro conseguida através da regressão linear, mas tendo-se verificado a existência de autocorrelação dos resíduos foi depois conseguida por métodos de regressão espacial, sendo que o melhor resultado foi obtido para o modelo de Spatial Lag.

A partir dos métodos de interpolação foi possível gerar uma superfície que corresponde à intensidade do sinal WiFi e também avaliar a necessidade da utilização do modelo digital do terreno como medida de impedância na análise de redes. Concluiu-se que o modelo que melhor aproximava a distribuição do sinal era a média móvel espacial, sendo o modelo digital do terreno melhor aproximado pelo interpolador do inverso da distância ponderada.

Finalmente, a etapa de análise de redes permitiu avaliar a performance da rede WiFi numa situação real, tendo sido verificadas a distância de pontos turísticos aos acessos WiFi mais próximos, a zona de cobertura do sinal WiFi e também variação de distância entre um caminho ótimo que dá prioridade à cobertura WiFi e um caminho onde apenas nos preocupamos com a distância percorrida.

Em conclusão, a avaliação feita da rede WiFi gratuita é que tem ainda uma cobertura insuficiente, sendo este facto particularmente ilustrado pelas lacunas detetadas durante a fase de análise de redes, mas também a nível da distribuição pouco balançada existente entre as zonas centrais da cidade e a periferia.

## Referências

- Catita, C. (2022). *Aulas de Analise Espacial de Dados*.
- Evennou, F., & Marx, F. (2006). Advanced integration of WiFi and inertial navigation systems for indoor mobile positioning. *Eurasip Journal on Applied Signal Processing*, 2006. <https://doi.org/10.1155/ASP/2006/86706>
- Gao, S., & Prasad, S. (2016). Employing spatial analysis in indoor positioning and tracking using Wi-Fi access points. *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, ISA 2016*, 27–34. <https://doi.org/10.1145/3005422.3005425>
- Grubesic, T., & Murray, A. (2004). Where" matters: Location and Wi-Fi access. *Journal of Urban Technology*, 11(1), 1–28. <https://doi.org/10.1080/1063073042000341952>
- How Far can a WiFi Signal Travel? - Actiontec.com.* (n.d.). Retrieved June 28, 2022, from <https://www.actiontec.com/wifihelp/how-far-can-wifi-signal-travel-2/>
- Li, W., Batty, M., & Goold, M. F. (2019). Real-time GIS for smart cities. *Https://Doi.Org/10.1080/13658816.2019.1673397*, 34(2), 311–324. <https://doi.org/10.1080/13658816.2019.1673397>
- Local Spatial Autocorrelation (1).* (n.d.). Retrieved June 28, 2022, from [https://geodacenter.github.io/workbook/6a\\_local\\_auto/lab6a.html](https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html)
- Project 4: Local Indicators of Spatial Association | GEOG 586: Geographic Information Analysis.* (n.d.). Retrieved June 28, 2022, from <https://www.e-education.psu.edu/geog586/node/673>
- RDN2008 / Italy zone (E-N) - EPSG:7794.* (n.d.). Retrieved June 27, 2022, from <https://epsg.io/7794>
- Regioni italiane.* (n.d.). Retrieved June 27, 2022, from <https://www.tuttitalia.it/regioni/>
- Spatial Analysis Of Geospatial Data: Processing And Use Cases.* (n.d.). Retrieved June 27, 2022, from <https://eos.com/blog/spatial-analysis/>
- Suchocka, M., Kimic, K., Blaszczyk, M., & Kolodyńska, N. (2019). Designing hotspots in the public spaces and public greenery of modern cities – selected issues. *Ecological Questions*, 30(4), 83–91. <https://doi.org/10.12775/EQ.2019.030>
- Valenzano, A., Mana, D., Borean, C., Servetti, A., Valenzano, A. ;, Mana, D. ;, Borean, C. ;, di Torino, P., Jolt, T., & Lab, O. (2016). Mapping WiFi measurements on OpenStreetMap data for Wireless Street Coverage Analysis. *Germany Article*, 16, 2016. <https://doi.org/10.7275/R5G44NHC>
- Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. <https://doi.org/10.1002/0471662682>

Sílvia Mourão, FC 57541

Yaqoob, I., Hashem, I. A. T., Mehmood, Y., Gani, A., Mokhtar, S., & Guizani, S. (2017). Enabling communication technologies for smart cities. *IEEE Communications Magazine*, 55(1), 112–120.  
<https://doi.org/10.1109/MCOM.2017.1600232CM>

Zhang, L., Zhao, L., Wang, Z., & Liu, J. (2017). WiFi networks in metropolises: From access point and user perspectives. *IEEE Communications Magazine*, 55(5), 42–48.  
<https://doi.org/10.1109/MCOM.2017.1600262>