

Análise Espacial de Dados  
**Estudo estatístico usado na análise espacial dos focos de incêndio de Castelo Branco nos anos de 2001 a 2010**

Maria Santos

(mariasantos@me.com)

---

## RESUMO

Com o intuito de proceder a uma análise estatística espacial através de conhecimento de métodos de análise profundos nesta área, procedeu-se à elaboração de uma análise: **ao padrão pontual**, **à distribuição por áreas** e **à regressão linear** dos dados dos focos de incêndio para o distrito de Castelo Branco. Respetivamente, identificou-se que o padrão pontual assumia um padrão do tipo **aglomerado** para a distribuição dos focos de incêndio, que a análise espacial de áreas **aceitava a H0** (na total ausência da autocorrelação espacial, assumindo uma distribuição aleatória na contagem dos nº dos focos existente por área). Na **regressão linear**, constatou-se que o modelo de regressão clássica era suficiente para explicar as variáveis, pela ausência de autocorrelação nos modelos espaciais das variáveis apresentadas.

Noutra análise, relativa à interpolação realizada pelo método determinístico **IDW**, verificou-se uma maior influência, pela representação mais notória da sua superfície contínua, na estimação de valores desconhecidos. Noutro tópico de estudo, referente à **análise de algoritmos em redes**, não foi possível detetar os comportamentos dos algoritmos previstos a analisar, devido à ausência de uma rede bem definida topologicamente (embora computacionalmente tivesse sido realizado o processamento desta em alguns dados).

Deste modo foi possível aprender didaticamente a utilização do *software* nos vários estudos realizados, e, compreender a importância da geostatística, mesmo não tendo sido possível tirar conclusões para todos os estudos.

---

## 1. Introdução

No âmbito da análise espacial, existe o interesse em tentar compreender a distribuição espacial de determinados fenómenos, nos quais se medem as propriedades e relações entre estes. Composta e realizada por três componentes: a **visualização**, a **análise exploratória** e a **modelação** de dados espaciais. [1]

Esta, **análise espacial**, foca-se sobretudo no estudo da **análise estatística de padrões**, por ferramentas, métodos e conceitos abordados posteriormente, tentando deste modo compreender qual a génese que está por detrás de um determinado **padrão observado**.

Após a sua identificação é efetuada uma quantificação através de **indicadores estatísticos** específicos, bem como, uma **validação** que visa explicar o fenómeno em estudo do padrão observado.

Neste projeto são abordados três tópicos: a **análise de um padrão pontual**, a **análise de áreas** e a **regressão espacial** de dados espaciais. [2]

Teve como objetivo principal apresentar uma análise fortemente estatística das ferramentas utilizadas com os diferentes *softwares* (*ArcGIS* e *GeoDa*) na obtenção dos resultados, assim como a aprendizagem didática do mesmo. [3] [4]

Os dados utilizados recaíram nos focos de incêndio de 2001 a 2010 para o distrito de Castelo Branco e na densidade populacional obtida para o mesmo. [5]

## 2. Aquisição dos Dados

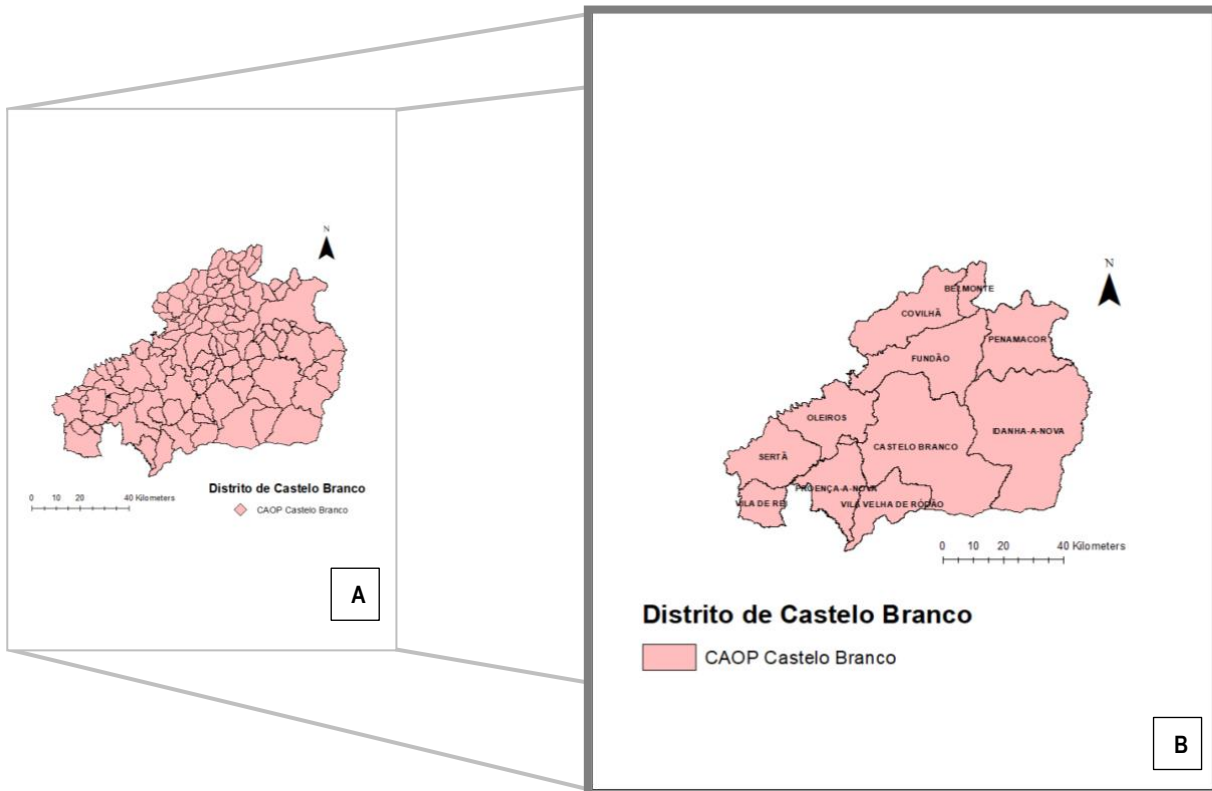
### 2.1 - Dados

**Tabela 1** Dados recolhidos e respetivas plataformas de informação

Nome	Tipo de Dados	Fonte
Carta Administrativa Oficial de Portugal (CAOP) 2017	Shapefile	<a href="http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop/caop_download/">http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop/caop_download/</a>
Focos de Incêndio de 2001 a 2010	Ficheiro Excel	<a href="http://www2.icnf.pt/portal/florestas/dfci/inc/estat-sqif">http://www2.icnf.pt/portal/florestas/dfci/inc/estat-sqif</a>
Densidade Populacional – obtida dos Censos	Ficheiro Excel	<a href="https://www.portugal2020.pt/Portal2020/Media/.../Censos_2011_Pop_Freguesias.xls">https://www.portugal2020.pt/Portal2020/Media/.../Censos_2011_Pop_Freguesias.xls</a>
Nº Crimes total - 2016	Ficheiro Excel	<a href="https://www.pordata.pt/Municipios/Crimes+registados+pe+las+pol%C3%ADcias+total+e+por+algumas+categorias+de+crime-600">https://www.pordata.pt/Municipios/Crimes+registados+pe+las+pol%C3%ADcias+total+e+por+algumas+categorias+de+crime-600</a>
Dados de O <sub>3</sub>	Shapefile	<a href="https://www.eea.europa.eu/data-and-maps/data/interpolated-air-quality-data-1">https://www.eea.europa.eu/data-and-maps/data/interpolated-air-quality-data-1</a>
Rede Viária	Shapefile	<a href="http://download.geofabrik.de/europe.html">http://download.geofabrik.de/europe.html</a>

## 2.2 - Técnicas e ferramentas utilizadas – Tratamento dos dados [6]

A análise estatística desta parte do trabalho é referente à área da região de Castelo Branco. Para tal, e com a

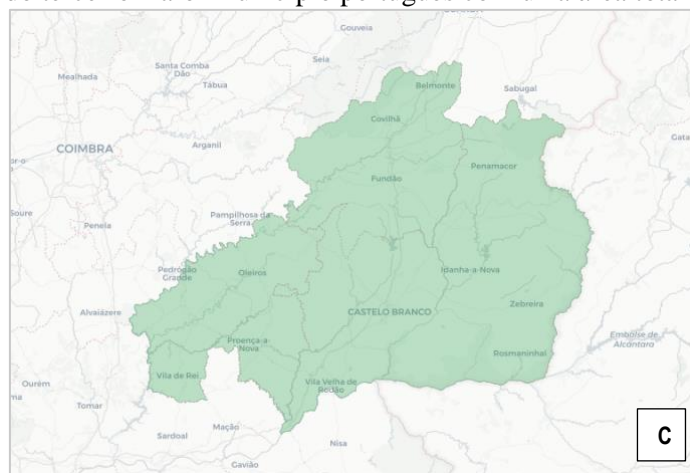


utilização do *ArcGIS*, procedeu-se à identificação desta zona em específico com um *Select by Attribute* ‘Distrito’ = ‘Castelo Branco’ ficando assim o distrito de Castelo Branco dividido com a identificação de todas as freguesias [Figura 1A]. Após este passo, e de modo a ser mais fácil identificar o nome das regiões da área em estudo, procedeu-se ao comando *Dissolve* por concelho através do qual obteve-se uma nova divisão do “Distrito”, mais restritiva [Figura 1B].

As *shapefiles* foram utilizadas com o Sistema de Referência Geográfica *ETRS89 TM06-PT*, Projeção Cartográfica Transversa de *Mercator* e Elipsóide de Referência *GRS80*. No ficheiro dos focos de incêndio definiu-se a mesma projeção através da *toolbox->DataManagement->Define Projection*.

## 2.3 – Identificação e Enquadramento do Distrito

O distrito de Castelo Branco, nome este partilhado também pela sua capital, é constituído por 19 freguesias e 11 concelhos, tratando-se do terceiro maior município português com uma área total de 1438 Km<sup>2</sup>. [Figura 1C]



**Figura 1** Identificação de todas as freguesias do distrito: (A) Geral; (B) Restritivo por Concelhos; (C) Mapa ilustrativo do distrito de Castelo Branco por Concelhos, com enquadramento pelo *basemap* do *software GeoDa*. Imagens (A) e (B) obtidas com recurso ao *software ArcGIS*

### 3. Análise da distribuição – Padrões Pontuais

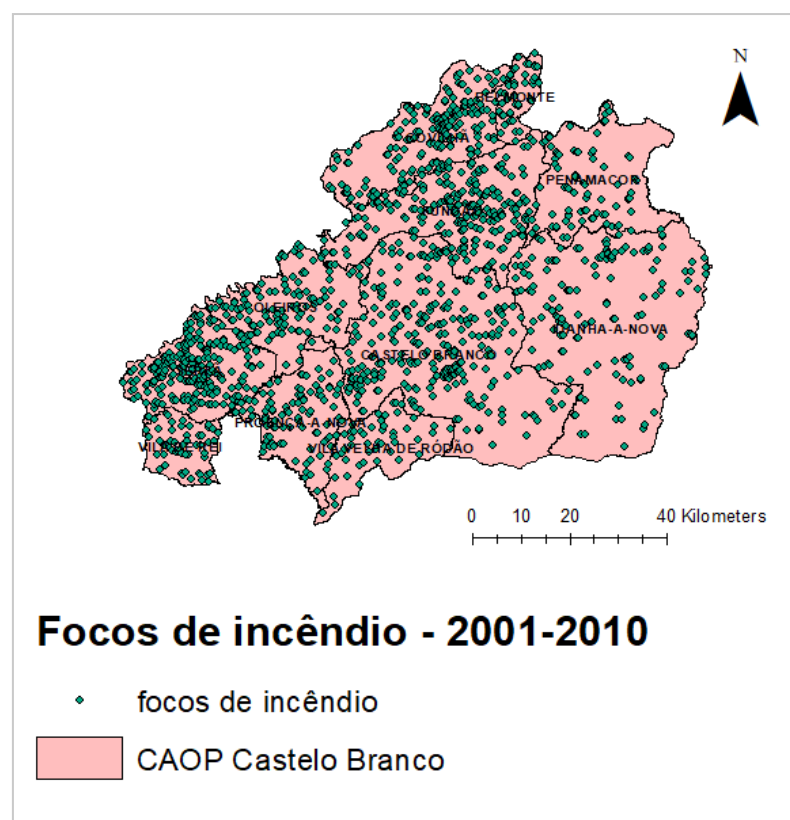
Nas componentes da **análise espacial**, tais como a **visualização** e **análise exploratória** dos dados, pretende-se, respetivamente, determinar e localizar padrões de interesse tendo em conta diversos parâmetros e quantificar padrões com determinados indicadores estatísticos. Na última componente, a **modelação**, visa explicar esses mesmos padrões de interesse.

Ao estudar a distribuição espacial pontual é possível identificar e localizar padrões de interesse, testando-se hipóteses sobre estes com o objetivo de determinar se são **aleatórios**, **aglomerados** ou **regularmente distribuídos**.

#### 3.1. Visualização dos Dados Espaciais

Neste campo a representação espacial dos dados georreferenciados é de vital interesse para a deteção de regularidades ou irregularidades espaciais e/ou associações.

Na **Figura 2** é possível visualizar a distribuição espacial dos focos de incêndio no distrito de Castelo Branco por concelhos.



**Figura 2** Distribuição dos focos de incêndio no distrito de Castelo Branco por concelhos – 2001-10. Imagem obtida através do software ArcGIS.

A utilização de **descritores** de padrões de pontos, como a **frequência** e a **densidade**, atua como parâmetros estatísticos básicos de modo a visualizar a distribuição pontual presente nas vertentes de variação e orientação espacial.

Sendo a **frequência** o número de eventos, isto é, o número de focos presentes na área de estudo, segundo a tabela de atributos foram detetados **5097** eventos nestes 9 anos [**Figura 2**].

$$\lambda = \frac{n}{A}$$

**Equação 1** Fórmula do cálculo da densidade, onde **n** número de eventos e **A** a área da região em estudo

A **densidade** usa a frequência dos eventos sob a área de estudo. Se relacionarmos de acordo com a **equação 1**, estima-se uma densidade de **3.55** focos/km<sup>2</sup> (5097/1438).

Outra das técnicas usadas para o estudo é a da **centrografia** que permite calcular o **centro geométrico** de todos os pontos na concentração dos focos de incêndio e a **distância padrão**, traduzindo o grau de dispersão do conjunto de eventos em redor do centro geométrico. [7]

Esta é uma forma muito básica na análise do padrão pontual, pois o centro geométrico e a distância padrão são estatísticas sumárias muito concisas, escondem informação valiosa sobre o padrão observado.

De todo o modo, a identificação do centro geométrico presente em ambas as imagens é dada pela média dos valores das coordenadas X,Y de cada ponto, garantindo a sua posição na zona centro-norte na cidade de Castelo Branco, bem como a localização da maior concentração de eventos/pontos.

A **distância padrão** revela a concentração dos pontos, através da medição da variância entre a média das distâncias dos pontos ao centro geométrico, expressa em unidades de distância.

Ambas as análises da distância padrão foram efetuadas a 95% e 68% de intervalo de confiança.

Na **Figura 3A** é possível observar-se o **círculo padrão**, que usa a distância como raio e na **Figura 3B** a **elipse padrão**, onde a diferença reside no fato do cálculo dessa distância ser aplicado a cada eixo separadamente delimitando o espaço onde é possível verificar 95% e 68% das observações, respetivamente.

Sendo esta uma técnica um tanto limitada e havendo métodos estatísticos mais poderosos para analisar padrões pontuais, onde se emprega a densidade e a distância como **detetores de padrões de pontos**, é a **análise exploratória** dos dados espaciais que vai permitir determinar um certo tipo de padrão.

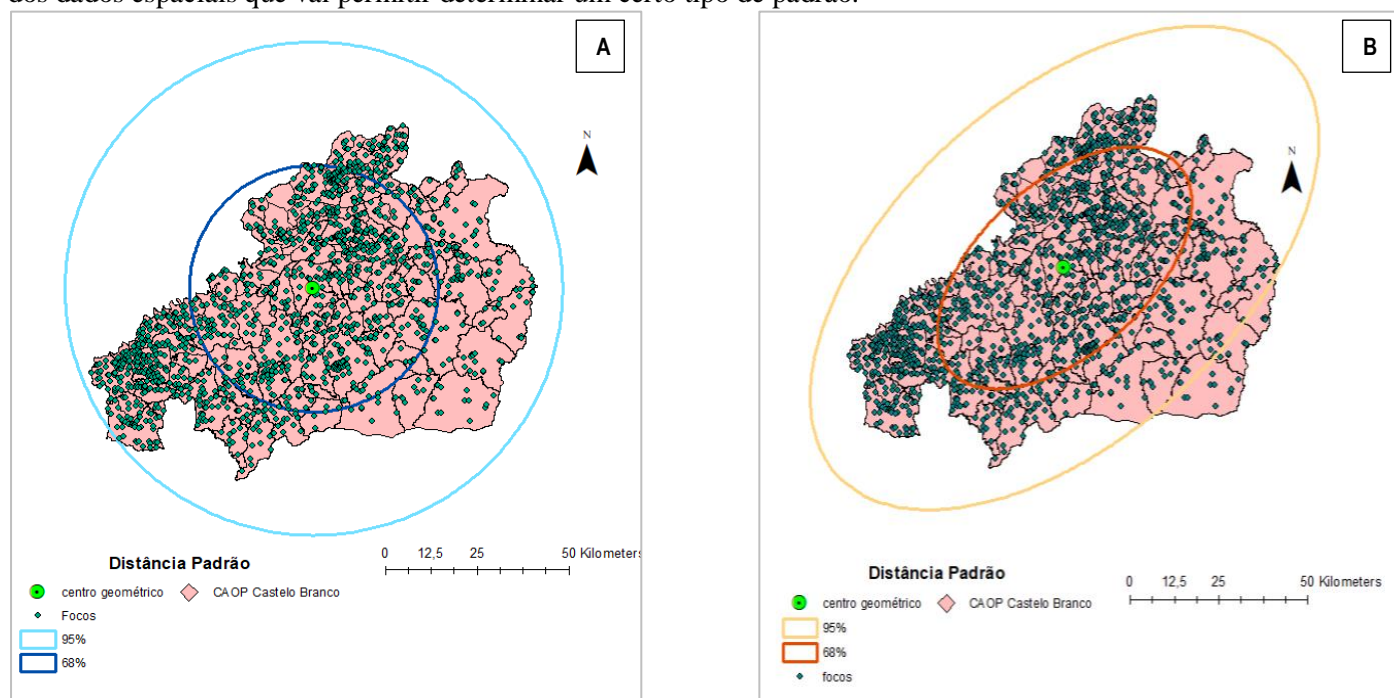


Figura 3 Círculo Padrão (A) e Elipse Padrão (B) da região de Castelo Branco. Imagem obtida pelo software ArcGIS.

### 3.2 Análise Exploratória dos dados

Com o objetivo de determinar se o padrão pontual de focos é do tipo **aleatório**, **aglomerado** ou **regularmente** distribuído, uma das primeiras análises do padrão passou por calcular a variação da densidade dos focos de incêndio na área em estudo. As técnicas baseadas em **densidades** caracterizam assim o padrão em termos de distribuição independentemente do tipo.

Para tal, a utilização de **técnicas globais** e **locais**, permitiram detetar os padrões e medir a distribuição espacial existente nos focos de incêndio.

#### 3.2.1 Análise Exploratória de 1º Ordem – Técnicas Globais

Uma medida básica na medição do padrão da densidade global tem em conta o rácio entre o n° de eventos e a área da região em estudo (como abordado anteriormente no ponto **3.1 Visualização de dados espaciais**)

- **Método de Contagem por Quadrantes** [8]

Esta técnica requer que a área de estudo seja dividida em quadrantes, ou seja, a densidade dos pontos é computada para cada quadrante ao dividir o n° de pontos em cada um pela área dos mesmos. Sendo esta uma técnica muito frágil, pois a escolha do número de quadrantes e até mesmo a forma desses quadrantes (de hexágonos e até mesmo

triângulos), podem influenciar a medida da densidade local, sendo esta uma escolha que deverá ser feita de forma ponderada [Figura 4].

Em *ArcGIS* procedeu-se à criação de uma *fishnet* dividindo a área que contém os focos de incêndio numa grelha retangular, posicionando-os dentro de quadrantes (foi considerada uma grelha de 15 linhas X 8 colunas).

De modo a efetuar a contagem de cada evento no respetivo quadrante obtido procedeu-se à abertura do ficheiro após a criação da *fishnet* no *Quantum GIS*, onde posteriormente se utilizou a ferramenta *count point in polygon*. Após este passo obteve-se a contagem do número de eventos, focos de incêndio, existentes em cada quadrante da grelha.

Em seguida, a tabela de atributos deste mesmo ficheiro foi importada para o *Excel* onde foi realizada a análise estatística tendo em conta o número de eventos em cada polígono.

De acordo com as variáveis apresentadas na **Tabela 2** foi possível calcular o *VTMR* (*variance to mean ratio*), o que posteriormente permitiu determinar o tipo de **padrão pontual** analisado.

Verifica-se que como  $VTMR > 1$  o padrão é do tipo **aglomerado** (existência de *clusters*). De modo a comprovar estatisticamente este resultado, realizou-se um teste de hipótese, que será revisto no ponto mais a frente deste trabalho, **Modelação do padrão pontual**.

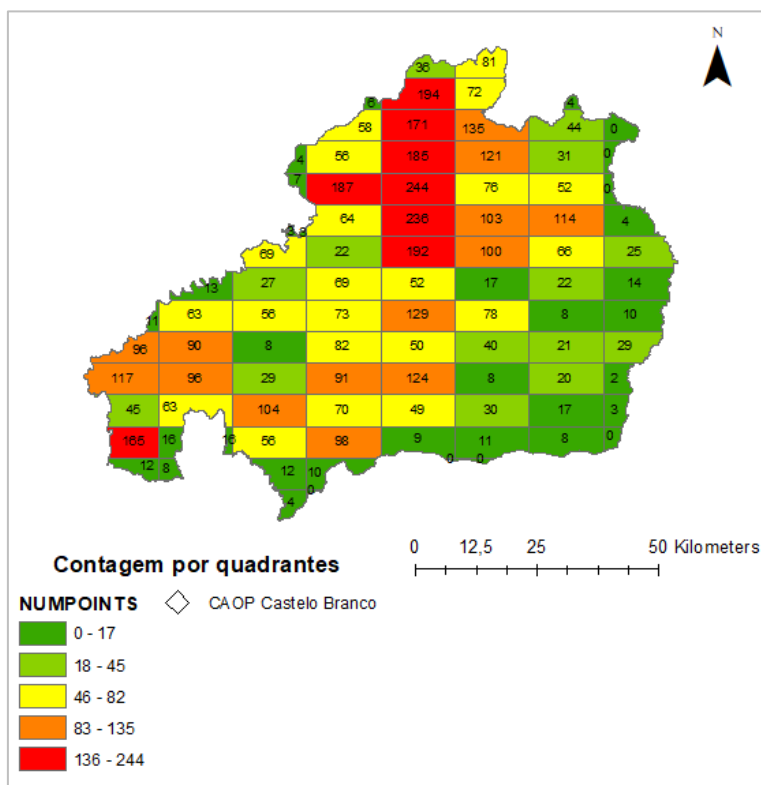


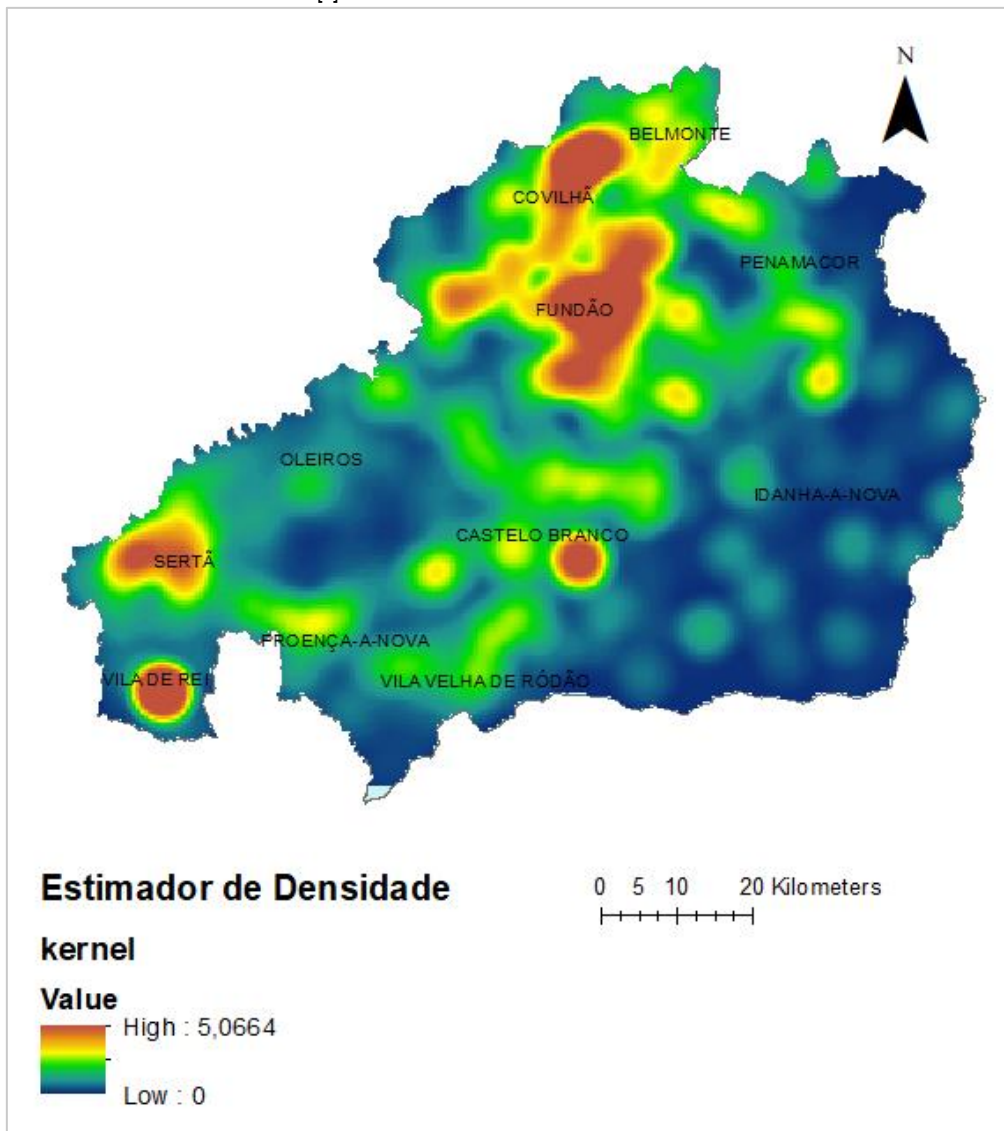
Figura 4 Ilustração do Método de Contagem por Quadrantes. Imagem obtidas pelo software *ArcGIS*

Tabela 2 Dados das variáveis obtidos em *Excel* para realização da análise do método de contagem por quadrante

Variáveis	Grelha 15x8
Nº de Quadrantes:	90
Nº de focos de incêndio:	5097
Variância:	3368
Média:	56.6
VTMR:	59.47



- **Estimador de Densidade** [9]



**Figura 5** Estimador de densidades (*Kernel density estimation*). Imagem obtidas pelo software ArcGIS

A abordagem do **estimador de densidade** (*kernel density estimation*) é uma extensão do método por quadrantes, onde é computada uma densidade localizada por conversão do conjunto dos focos de incêndio numa superfície contínua.

Esta conversão ocorre por emprego de uma função que é proporcional à densidade das amostras por unidade de área, responsável pela contagem de todos os pontos dentro de uma sub-região através de um círculo de influência, que, por sua vez, é então influenciável pela escolha de uma largura de faixa, neste caso a escolha do raio do círculo, de modo a obter o resultado pretendido. [Figura 5]

Como se pode visualizar, este método permite a fácil deteção de *hot spots*, onde a mais **alta densidade de focos** ocorre em Vila de Rei, Sertã, Covilhã, Fundão e Castelo Branco.

A sua realização no ArcGIS teve como *input* a *shapefile* dos focos de incêndio e todas as outras variáveis foram realizadas com as definições *standard* propostas pelo software, exceto a unidade de área em que foi considerada os Km<sup>2</sup>.

### 3.2.2 Análise Exploratória de 2º Ordem – Técnicas Locais

Uma alternativa aos métodos baseados em densidades são os que recorrem à utilização de distâncias para análise de padrões (distância mínima), cujo interesse recai em como os eventos são distribuídos relativamente uns aos outros, evidenciando uma **dependência espacial**. Esta análise é representada por técnicas locais, também conhecidas por efeitos de segunda ordem, tais como o **método vizinho mais próximos – Função G e K-Ripley**.

- **Função G** [10]

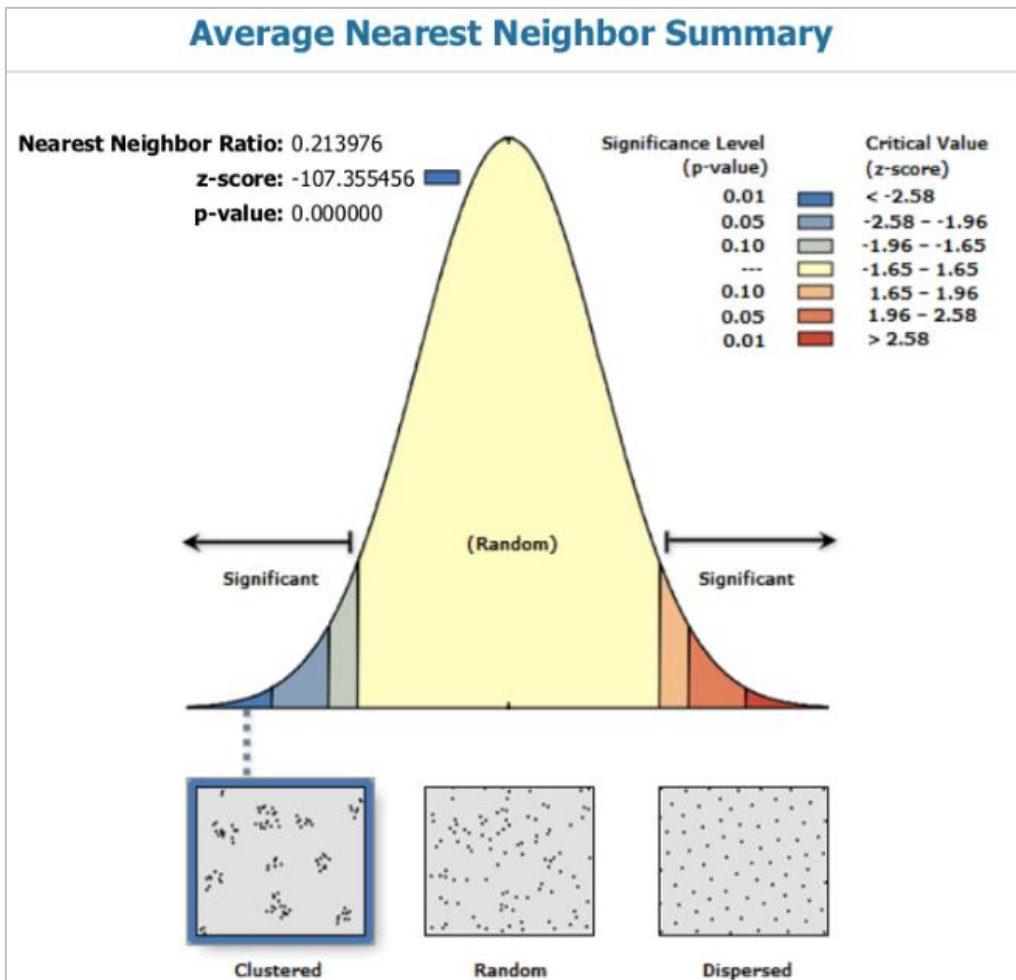


Figura 6 Curva Gaussiana do método vizinho mais próximo com z-score associados. Imagem obtidas pelo software ArcGIS

Esta técnica calcula a média das distâncias de cada ponto na área de estudo ao seu vizinho mais próximo, tendo como limitação o fato de só considerar a distância mínima entre dois eventos.

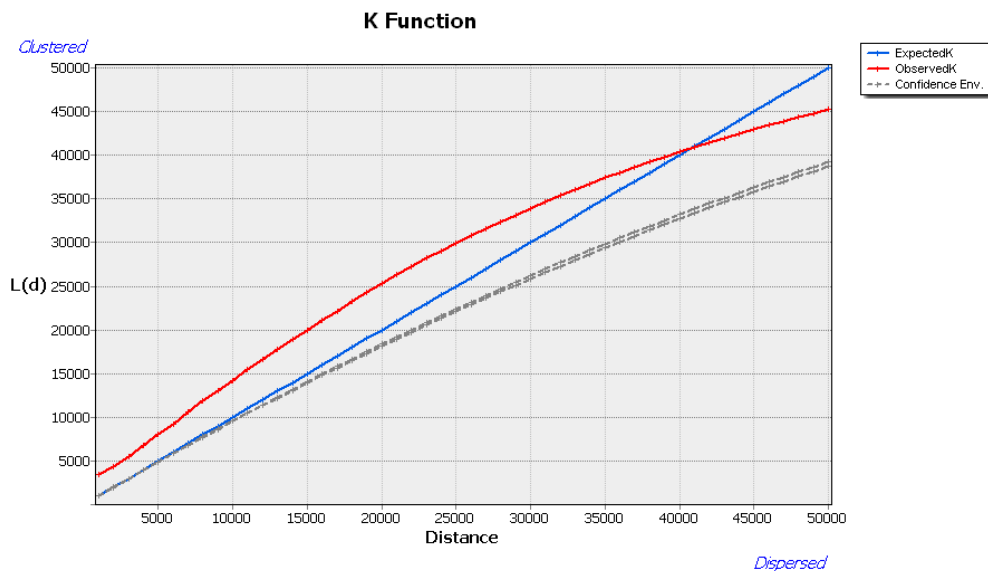
Com recurso ao ArcGIS, foi possível visualizar que o padrão detetado dos focos de incêndio é então do tipo **aglomerado** (*clustered*), concluindo com 99% de confiança e demonstrado pelo nível de significância, que a probabilidade deste padrão ser **regularmente disperso** é de menos 1%. Outra das variáveis usadas para análise é o valor negativo elevado de **z-score**, como medida de análise ao desvio padrão e **p-value** como valor nulo da probabilidade do padrão pontual ser do tipo aleatório. [Figura 6 e 7]

De notar que a **distância média esperada** era de 693.91 metros e que a **distância média observada** foi de 148.5 metros.

Average Nearest Neighbor Summary	
Observed Mean Distance:	148.4800 Meters
Expected Mean Distance:	693.9104 Meters
Nearest Neighbor Ratio:	0.213976
z-score:	-107.355456
p-value:	0.000000

Figura 7 Distância média esperada e observada no método vizinho mais próximo. Resultados obtidos através do software ArcGIS

- **Função *K-Ripley*** [10]



**Figura 8** Representação da linearização da função *K* linearizada - (L(d)). Imagem obtidas pelo software *GeoDa*

Sendo a **função vizinho mais próximo** das mais usadas na análise estatística do padrão pontual, e com especial influência a de *K-Ripley*, que permite uma análise mais precisa da distribuição de pontos através da distância entre **todos os eventos** na área de estudo.

Com um cálculo simples, que consiste em dividir a média da soma do número de focos de incêndio a diferentes distâncias para cada evento pela densidade de eventos por área.

No *ArcGIS* procedeu-se à análise dos focos de incêndio com a *K-Ripley*, onde se computou o envelope de confiança com somente 9 permutações, 50 bandas de distância. As distâncias a avaliar foram definidas, com um incremento de 1000 metros, obtendo-se assim o gráfico ilustrado na **Figura 8** (L(d) é o resultado da linearização da Função *K* para facilitar a interpretação visual do gráfico da mesma).

É possível analisar que quando o valor observado (vermelho) se encontra acima do valor esperado (a azul) encontramos-nos perante um **aglomerado**.

Uma vez terem sido escolhidas 50 bandas de distância para este gráfico é possível observar que a partir dos 40000 metros de distância, o momento em que a linha dos valores observados se cruza com a linha dos valores esperados, apresenta-se um **padrão disperso**.

### 3.3 Modelação Padrão Pontual

Com o intuito de explicar esta identificação de padrões, todo este processo de análise estatística de um padrão pontual para identificação de tipo de padrão observado, necessita de ser **validada**. Para tal, é utilizado um modelo teórico - **CSR** (hipótese de **Aleatoriedade Espacial Total**) para validação dos resultados obtidos previamente pelo uso das técnicas de análise espacial empregues, através da comparação do padrão de pontos observado com os gerados pelo processo aleatório independente conhecido como **CSR**. [11]

Assim, a definição de um modelo teórico de referência **CSR**, consiste num modelo em que a divisão da área de estudo em subáreas, distribuindo os eventos pontuais aleatoriamente, com ocorrências homogêneas e não correlacionadas, com um pressuposto de duas hipóteses:

**H1:** existir **igual probabilidade** de ocorrência de um evento na área de estudo;

**H2:** existir **independência** entre eventos.

A hipótese de aleatoriedade espacial total é realizada por dois métodos. **O 1º método** consiste numa comparação com as distribuições teóricas de referência seguida de um teste de ajustamento, utilizada para o método de contagem por quadrantes e para o do vizinho mais próximo.

**O 2º método** consiste na simulação do processo aleatório com a realização de gráficos com envelopes de simulação e análise dos desvios desse processo - realizado previamente para a **função G** e **K-Ripley**.



- **Método de Contagem por quadrantes**

Como referido anteriormente, procedeu-se à identificação e cálculo das variáveis presentes na **Tabela 2**. A identificação do padrão é então realizada através da interpretação do valor de **VMTR** (>1 o que implica presença de um padrão aglomerado).

Com a realização de um teste de significância estatístico com **chi-quadrado** e tendo em conta o **número de quadrantes**, o **nº de eventos** e o **nº de graus de liberdade**, com a função do *Excel* **chisq.inv**, procedeu-se a identificação dos extremos do intervalo para um nível de confiança de 99%. [**Tabela 3**]

**Tabela 3** Extremos do intervalo para um nível de confiança de 99%

VTMR	$\chi^2$	$\chi^2$ 0.99
		$\chi^2$ 0.01
59.47	5040.36	60.92
		122.94

Como o valor de **chi-quadrado** obtido encontra-se fora do intervalo pretendido, **rejeita-se a hipótese H1** proposta pelo modelo teórico, demonstrando que o padrão é um **aglomerado**.

- **Função G**

Para esta função a identificação do tipo de padrão é dada pelo valor de **z-score**. Como o valor obtido foi de **-107.35**, encontrando-se fora do intervalo **rejeita-se a H2**, com um nível de confiança de 99%; confirmando que o padrão identificado não está na presença de **aleatoriedade espacial completa**, mas sendo assim um aglomerado - valor negativo elevado. [**Tabela 4**]

**Tabela 4** Nível de Significância e valor crítico que retrata a figura 6

z-score	p-value probabilidade	Nível de Confiança
<-2.58 ou >2.58	<0.01	99%

- **Função K-Ripley**

Recorrendo à utilização de envelopes de simulação [**Figura 8**] para o teste da hipótese de aleatoriedade total num determinado nível de confiança, é calculada uma função **K** para uma determinada distribuição pontual aleatória distribuída pelos envelopes de confiança. Este acontecimento é chamado **permutação**. Assim o número de permutações que ocorrem determina o número de vezes que esses pontos são distribuídos aleatoriamente, demonstrando que **quanto maior o número de permutações maior o nível de confiança**.

Esta análise foi realizada com a escolha de **9 permutações**, equivalente a um nível de confiança de 90%. A escolha de um baixo número de permutações recai no tempo moroso de operação do *software* nesta função.

Assim e de acordo com o gráfico, obtém-se com nível de 90% de confiança que o tipo de padrão para os focos de incêndio de Castelo Branco é do tipo **aglomerado**.

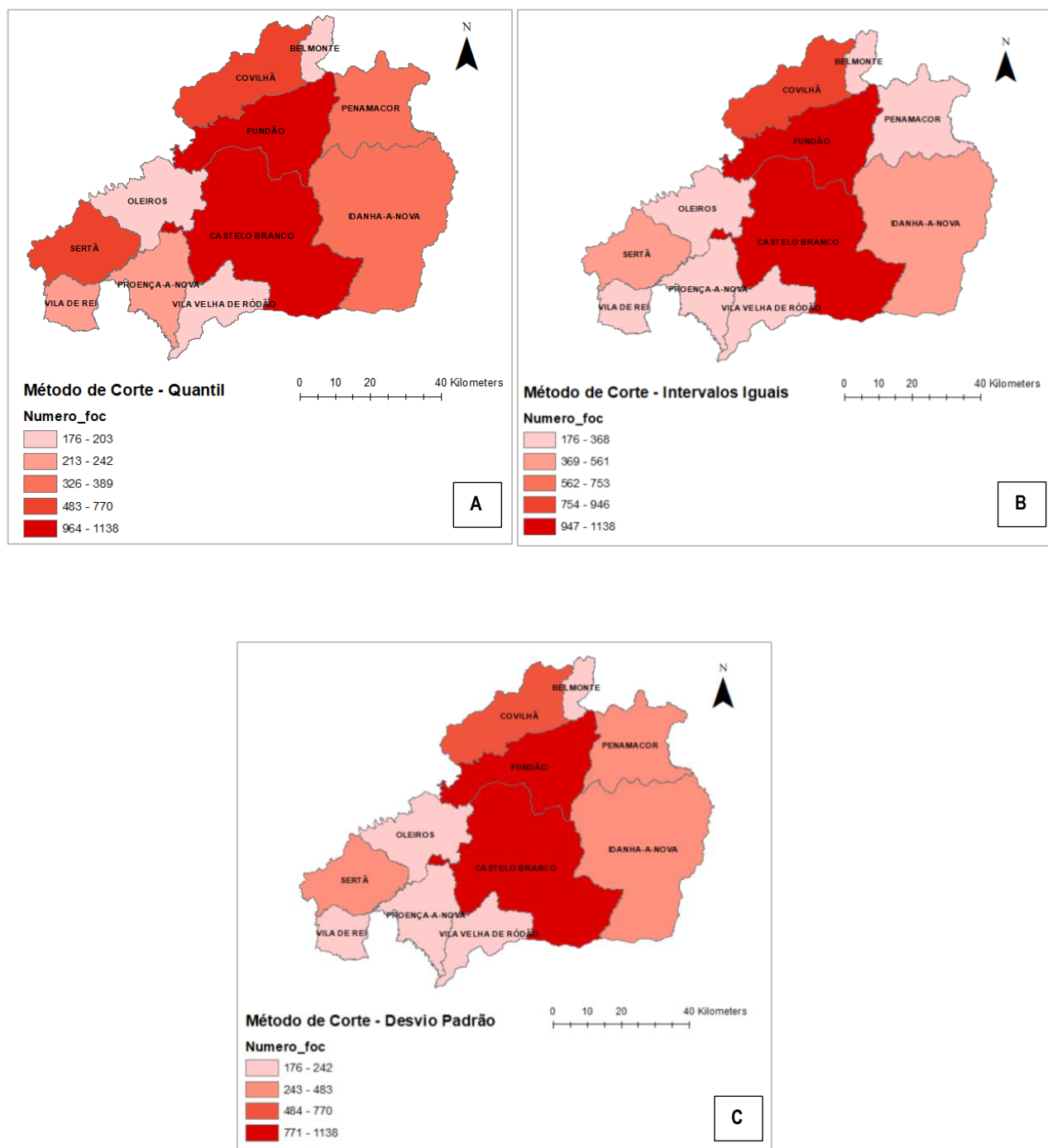
## 4. Análise de Áreas

Tendo a análise de áreas como objetivo verificar a existência de **outliers** no conjunto dos dados em relação aos vizinhos, e assumindo uma análise não estacionária verificada pelo valor constante da **média** (estacionaridade de 1º ordem) e **variância** (estacionaridade de 2º ordem), utilizou-se os dados do número de focos de incêndio por concelho com este mesmo intuito.

Do mesmo modo que o padrão pontual, a análise de áreas assume três componentes: **visualização dos dados espaciais**, a **análise exploratória dos dados** e a **significância estatística**, abordados posteriormente.

#### 4.1 Visualização dos dados espaciais

Ao ter a contagem do número de fogos por concelho efetuada e assumindo os concelhos como áreas vizinhas, procedeu-se à visualização deste efeito pelo **método de planos de corte** usando **quantis**. O mesmo foi efetuado tendo em conta o **desvio padrão** e **intervalos iguais**. A variável do número de focos não foi normalizada e manteve-se a escolha das 5 classes, exceto para o **desvio padrão** que apresentava só 4.



**Figura 9** Perspetivas do número de focos de incêndios segundo diferentes métodos de corte (A) Quantil; (B) Intervalos iguais; (C) Desvio padrão. Imagem obtidas pelo software ArcGIS

É possível observar que a mesma variável, em métodos de corte diferentes, tem ligeiras diferenças visuais na sua representação que embora os valores mais elevados de foco de incêndio nos concelhos de Castelo Branco, Fundão e Covilhã se mantenham dentro da mesma classe nos 3 planos de corte, o Concelho de Sertã, encontra-se em classes diferentes, bem como Penamacor e Idanha a Nova. [Figura 9 A-C]

Isto devido ao fato de no método de **quantil** os eventos são agrupados em quantidades iguais a cada classe, enquanto que no **desvio padrão**, a dispersão é realizada em torno da média.

A desvantagem recai assim mais sobre o conjunto de intervalos iguais que apresenta pouca variação nos valores mais baixos, não representando a realidade dos eventos dos aglomerados previamente observados.

## 4.2 Análise Exploratória dos Dados Areais

Apresenta dois objetivos: **medir a estrutura da correlação espacial** existente e **estimar a magnitude da autocorrelação entre áreas**. Isto é efetuado com indicadores de análise estatística de **1º ordem (globais)**, ao caracterizar a dependência espacial e a correlação de valores no espaço, e de **2º ordem (locais)**, aquando da identificação de agrupamentos e/ou *outliers*.

A aplicação destes métodos passa pela definição de uma matriz de vizinhança, que mostra a ligação entre as subdivisões da área, onde a matriz toma valores 1 para sub-regiões vizinhas e 0 quando tal não ocorre.

Recorrendo ao *software GeoDa*, procedeu-se então à criação de uma matriz *queen type* e simétrica, relacionando a frequência com a distribuição de vizinhos. [Figura 10 A-B]

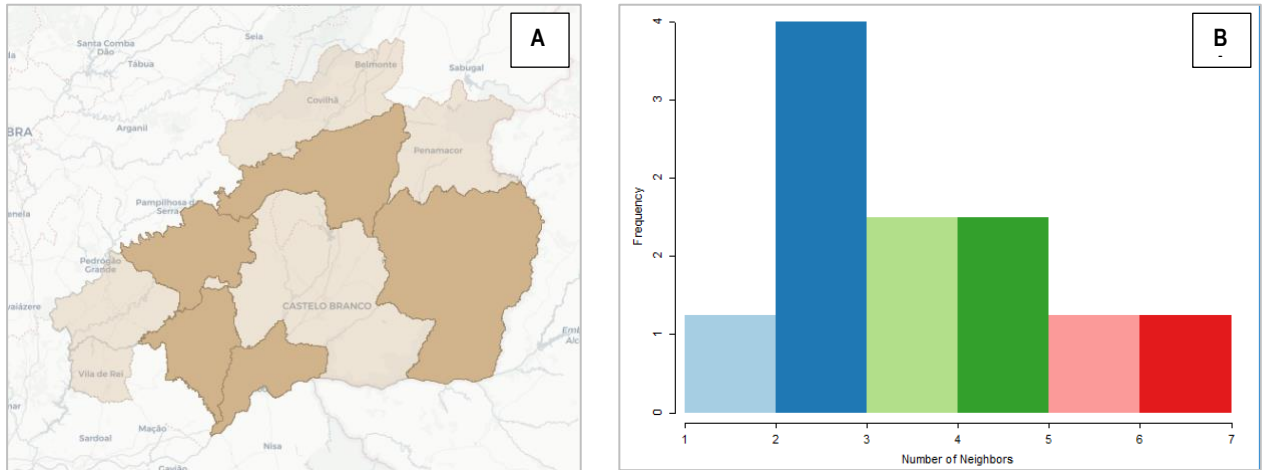


Figura 10 (A) Mapa de vizinhança - demonstrando a vizinhança do concelho de Castelo Branco, com enquadramento através de um *basemap*; (B) Histograma da matriz de vizinhança. Imagens obtidas com recurso ao *software GeoDa*

- **Efeitos de 1º ordem – Índice Global de Moran (I)** [12]

Usado para calcular o grau de **autocorrelação (grau de influência que uma variável tem sobre si mesma)** espacial existente na variável em análise e na sua localização.

De modo a quantificar esse grau de autocorrelação, ocorre:

**Autocorrelação negativa**  $-1 < I < +1$       **Autocorrelação positiva**

Onde 0 assume o valor de aleatoriedade. A autocorrelação **negativa** indica valores dispersos e a **positiva** aglomerados, revelando dependência espacial mais forte.

No mesmo *software* calculou-se o índice global de Moran para a variável dos focos de incêndio, obtendo-se o seguinte gráfico:

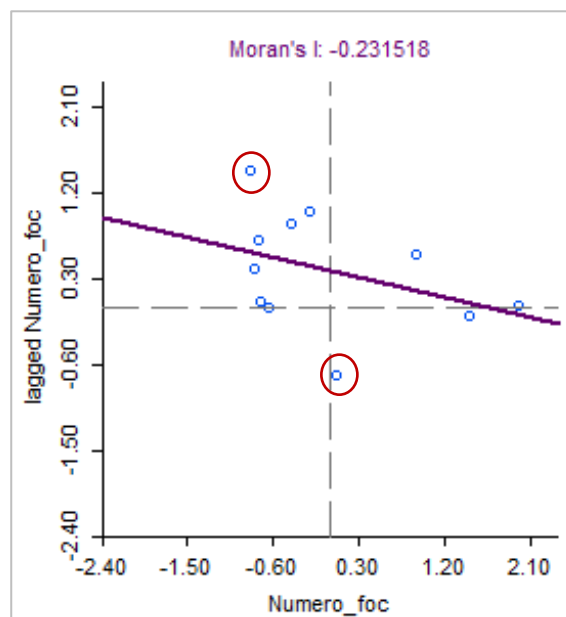


Figura 11 Autocorrelação da variável do número de focos de incêndio – outliers sinalizados a vermelho. Imagem obtidas pelo *software GeoDa*

Deste modo, e uma vez definida a área de análise e a identificação dos vizinhos dos concelhos no conjuntos de dados, é computado o valor médio para cada área, obtendo a **média dos vizinhos** referidos como *lagged Numero\_focos*. Em seguida o *plot* é então realizado pelo *Número\_focos lag vs Número\_focos* para cada concelho. O coeficiente é obtido pelo **declive da reta de regressão linear** que mais se ajusta aos pontos obtidos, apresentando um declive de **-0.23**.

Visualmente observa-se também a presença de **dois outliers**; presentes na localização de Belmonte e Sertã, respetivamente. [Figura 12]

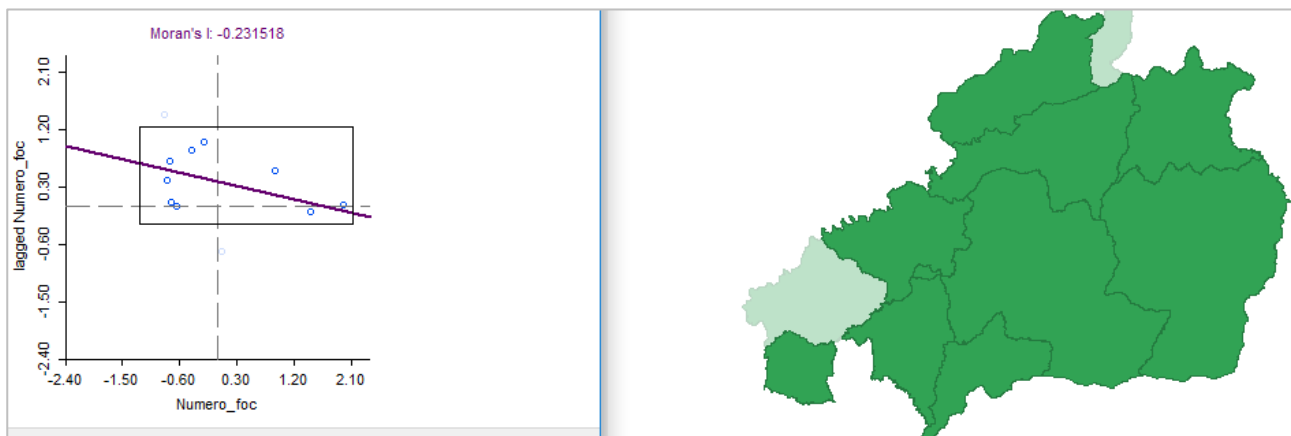


Figura 12 Visualização mais pormenorizada dos outliers. Imagem obtidas pelo software GeoDa

Como o declive da reta é próximo do horizontal, com tendência de autocorrelação negativa, e o **valor do índice** próximo de 0, levanta-se a questão: quão significativo é este índice de *Moran*?

Para tal é necessário realizar a validação estatística, de modo a corroborar se os valores medidos poderão vir a apresentar correlação espacial significativa. A significância de *I* é estimada de duas formas possíveis: por **teste de pseudo-significância**, em que são realizadas técnicas de permutação, ou, por associação do **índice a uma distribuição estatística** aproximada.

Para esta análise foi então realizada o teste de **pseudo-significância (sob a H0)**, para a mesma variável, com a realização de 999 permutações, explicadas previamente. O teste é realizado sob a **Hipotese 0: ausência de autocorrelação espacial**.

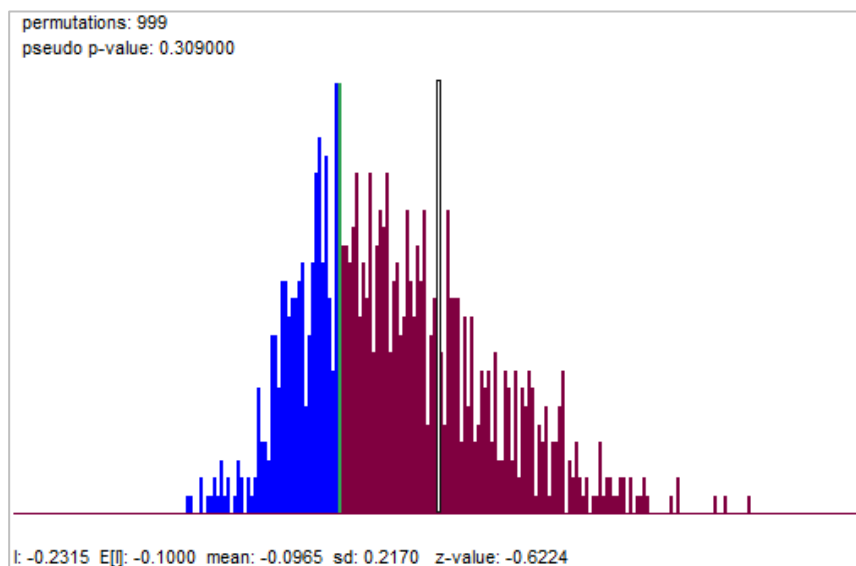


Figura 13 Teste de pseudo-significância com 999 permutações, com nível de confiança de 99.9%. Imagem obtidas pelo software GeoDa

De acordo com a **Tabela 4** como o valor de **z-score** é de **-0.62**, este resultado mantém-se dentro do intervalo, aceita-se a **H0**, e consequentemente confirma-se a tendência de aleatoriedade do nº de focos de incêndio, não revelando **autocorrelação espacial**, pois são **espacialmente independentes**.

Realizando o procedimento no *ArcGIS*, com as opções *standard* selecionadas pelo *software*, obteve-se um **índice de Moran** de **0.20**; com valor de *z-score* de **1.32** ficando este valor dentro do intervalo mesmo para um nível de confiança de 90%; aceitando a hipótese **H0 (padrão de aleatoriedade)**.

Esta diferença no valor de *z-score* poderá ocorrer pela baixa representação da variável utilizada, tendo em conta que se consideraram o número de focos de incêndio somente por 11 concelhos.

- **Efeitos de 2º ordem – Indicadores Locais de Autocorrelação espacial (LISA):** Índice Local de *Moran* e Índice *G*

Usados com o objetivo de identificar o valor de correlação em cada área, neste caso, concelho, comparando o seu valor com os vizinhos adjacentes, ao avaliar a covariância entre estes e ausência ou presença de aglomerados e *outliers* em cada localização.

Os **indicadores LISA**, avaliam assim a hipótese de aleatoriedade espacial, pela identificação local de *hotspots* e de instabilidade locais, por um processo aleatório onde é gerado uma distribuição espacial de referência para análise da significância estatística.

Com a utilização do *software GeoDa*, procedeu-se a utilização destes índices na análise da variável do nº de focos de incêndio. [Figura 14]

- **Índice Local de Moran**

O Índice local de *Moran* permite, muito eficientemente, calcular o **produto dos resíduos** na área selecionada com a **média local dos resíduos** da vizinhança.

Com a utilização deste indicador, tendo em conta a variável do nº de focos de incêndio, foi possível visualizar pelo *LISA Cluster Map* a ocorrência de dois concelhos com características diferentes, apresentados nos “quadrantes” **High-Low e Low-High**. Onde o concelho de referência, apresenta a média local dos resíduos da sua vizinhança com valores não semelhantes, isto é, com **atributo e média abaixo do esperado**. Isto caracteriza assim os concelhos que revelam os índices de **incêndio abaixo da média** para a *High-Low* e os que se encontram **acima da média** em *Low-High*. [Figura 14]

Tal indicador ocorre respetivamente em Sertã, Belmonte e Idanha-a-nova. Para validação dos dados é obtido também o **mapa de significância estatística** onde se comprova por **95% de confiança** que nos respetivos concelhos ocorre **dependência espacial**, embora não seja detetável a presença de aglomerados e/ou *outliers* (**não** evidenciadas estas situações, que por sua vez também não tinham sido identificadas pelo **Índice Global de Moran**).

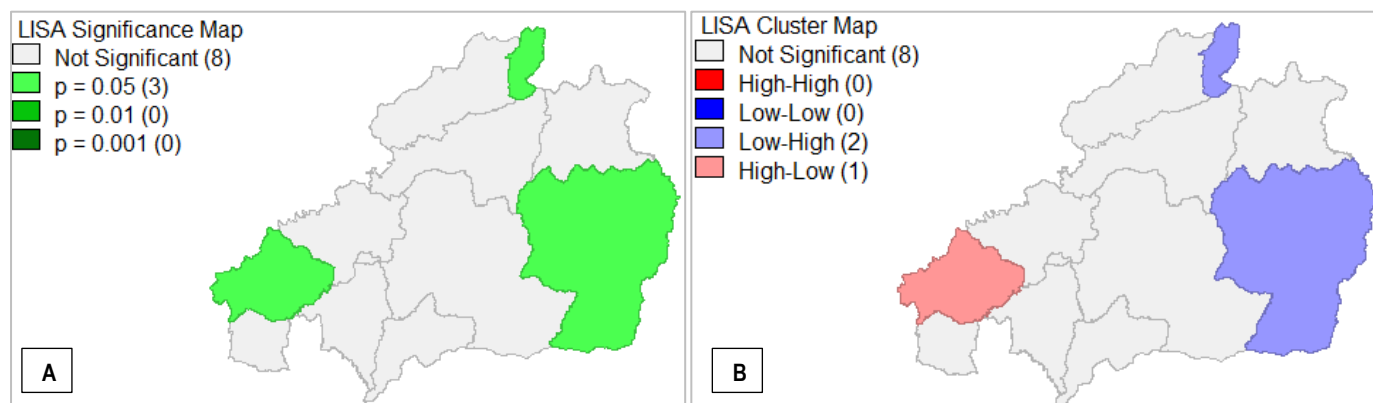


Figura 14 Índice local de *Moran* para a variável do nº de focos de incêndio, no distrito de Castelo Branco. (A) *LISA Significance Map*; (B) *LISA Cluster Map*. Imagem obtidas pelo *software GeoDa*

- **Índice de Geary:  $G_i$  e  $G_i^*$**  [13]

Sendo também um indicador *LISA*, onde a estatística é determinada pela **soma de valores vizinhos** definidos a **partir de uma matriz de vizinhança** – indica um valor para cada área.

Neste método, os valores mais elevados representam a existência de **altos índices de ocorrência** do número de focos de incêndio e valores inferiores indicam **agrupamentos de valores baixos** para a mesma variável, assumindo uma **distribuição normal** para a construção dos intervalos de confiança, onde se assume que:

- **H0**: total ausência de **autocorrelação** local;

A diferença apresentada pelas estatísticas  $G_i$  e  $G_i^*$  recai, respetivamente, na **exclusão do valor  $i$**  da sua soma apresentado pelo **cálculo do índice de aglomeração** de valores elevados e baixos na variável, e, na outra a inclusão do



valor  $i$ , pela **determinação de aglomerados** de valores altos e baixos na variável. O indicador foi assim realizado com o *GeoDa*, com **999** permutações. [Figura 15]

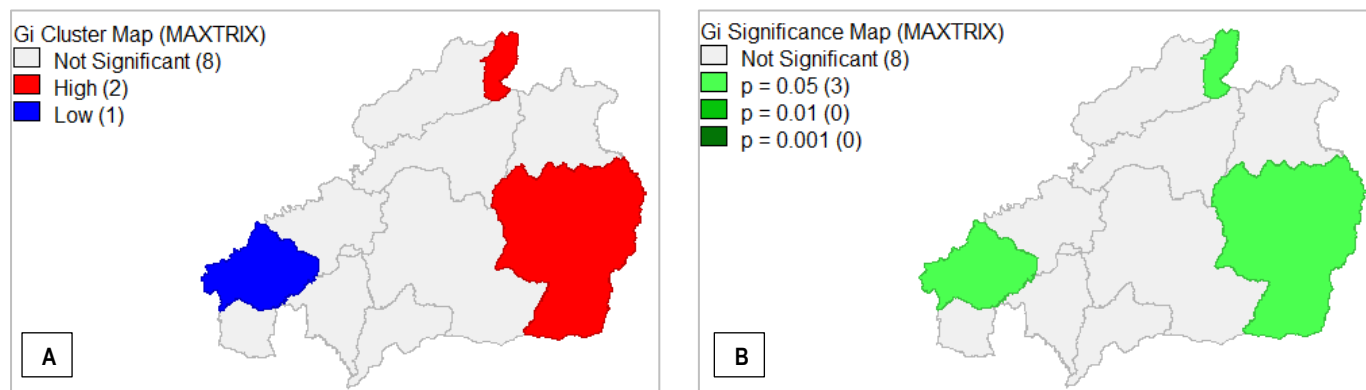


Figura 15 Índice local de Geary para a variável do nº de focos de incêndio, no distrito de Castelo Branco. (A) *Gi Cluster Map (MAXTRIX)*; (B) *Gi Significance Map (MAXTRIX)*; Imagem obtidas pelo software *GeoDa*

É possível observar os concelhos de Belmonte e Idanha a Nova como *hot-spots*, com um **grau/índice de aglomeração** espacial de valores elevados de nº de focos de incêndio e o concelho de Sertã como um *cold-spot* com um **grau/índice de aglomeração espacial de valores High e Low**, definidos pela soma dos valores vizinhos, isto é, o nº de focos de incêndio, adjacentes à área de referência representada.

O índice de  $G_i^*$ , quando calculado, apresentava exatamente os mesmos resultados gráficos representando um **índice de agrupamento de valores High e Low** para os mesmos concelhos.

De notar que embora a identificação com estes indicadores locais de *Moran* e de *Geary* recaiam nos mesmos concelhos, [Figura 14 e 15, respetivamente], estes indicadores medem estatisticamente valores diferentes: **um baseado na variância e outro na soma**.

## 5. Regressão [14]

A análise por **regressão** é normalmente usada de modo a entender, modelar e prever, o comportamento de um conjunto de variáveis que se relacionam entre si, com o intuito de obter resposta a questões, que possam solucionar problemas anteriormente detetados. Para este trabalho seria a procura de soluções e/ou variáveis no âmbito da redução do número de focos de incêndio.

Contudo muitas vezes a construção de um modelo de regressão não é tarefa fácil tendo em conta as variáveis em jogo. Deste modo existem conjuntos de ferramentas de regressão que permitem solucionar este problema, apresentando diagnósticos de forma a se obter um modelo apropriado, isto é, de confiança.

A regressão poderá ser realizada assumindo o seu **método clássico [Equação 2]**, e o seu método espacial, tendo em conta um **conjunto de dados espaciais**, abordados de seguida.

### 5.1 Regressão Linear clássica [15]

A sua objetividade passa por encontrar um **bom ajuste entre os valores estimados pelo modelo e os valores observados**, bem como **determinar quais das variáveis independentes (explicativas) contribuem de forma significativa** para a relação linear. Neste caso, esta determinação revelou-se limitada uma vez que se considerou a variável da densidade populacional, o número de crimes por concelho e o média do declive existente nestes.

Procedeu-se assim à escolha da **variável dependente**, isto é, o número de focos de incêndio (variável que se pretende modelar e prever). De seguida considerou-se os fatores que podem ajudar a explicação desta variável dependente, e dos poucos dados obtidos, estabeleceu-se a densidade da população (número de crimes e declive do terreno para o distrito de Castelo Branco) como **variável independente**.

Muitas vezes é a pesquisa de informação que permite criar um bom modelo de regressão. Neste caso, identificar o tipo de solo e a humidade presente, seriam variáveis significativas, mas por pouco conhecimento do *software* e dificuldade a encontrar esta informação tornou-se complicado fazer a ligação entre estes elementos e a CAOP utilizada. Deste modo, a análise pela utilização da densidade populacional (número de crimes e a média do declive do terreno por concelho) demonstrou-se bastante limitada.

$$Y = \alpha + \beta X$$

Equação 2 Modelo da Regressão Linear clássica

Efetuada esta escolha das variáveis a utilizar procedeu-se então a análise da regressão, afim de aferir se o número de focos de incêndio são influenciados pela densidade populacional existente, pela ocorrência do número de crimes ou até mesmo pelo declive do território. [Figura 16 A-B]

Com o *GeoDa*, procedeu-se à criação de uma **matriz de vizinhança** (*queen matrix*) o que possibilitou a regressão. Obteve-se o ficheiro dos dados estatísticos e o diagrama de dispersão que permitem avaliar o ajustamento das variáveis à reta. Ainda em *ArcGIS*, foi obtido o declive através do *MDT*, onde foi calculada a média pelo concelho. [ANEXO I]

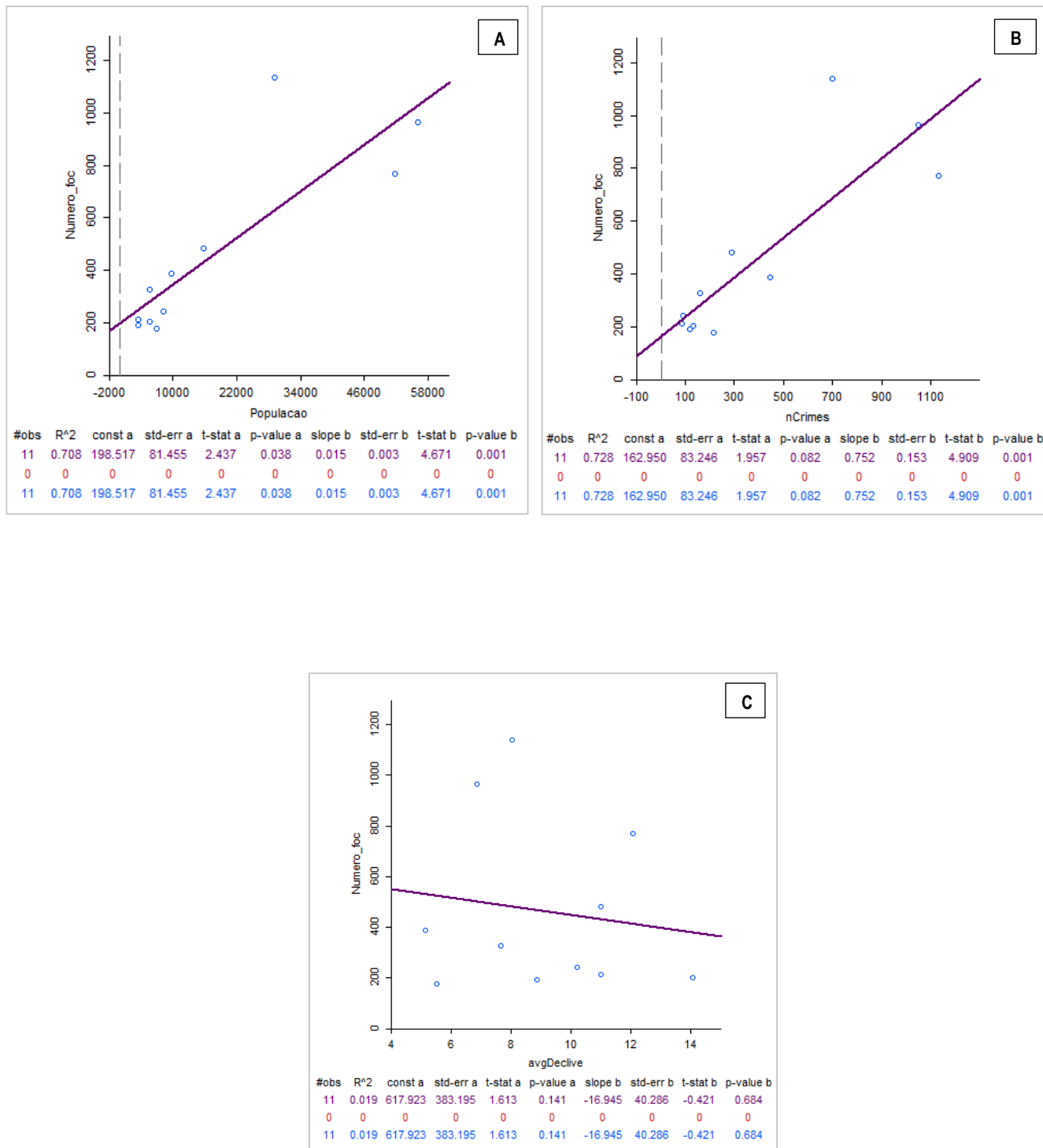


Figura 16 (A) (B) Diagrama de dispersão realizado por *scatter plot*; variável dependente nº de focos de incêndio e as respetivas variáveis independente: (A) população; (B) Nº de Crimes; (C) Média dos declives. Imagem obtidas pelo *software GeoDa*

Tabela 5 Comparação dos modelos de regressão

Modelos de Regressão	Regressão Linear		Regressão Múltipla
Variáveis Independentes	Densidade Populacional	Nº de Crimes	Densidade Populacional/ Nº de Crimes
$R^2$	70.8%	72.8%	73.10%
Log likelihood	-72.45	-72.03	-72.0
AIC	148.91	148.127	150.0

A avaliação dos valores obtidos através do ajuste da regressão linear tem em conta três medidas de elevada importância: o **erro médio quadrático (EQM)**, o **coeficiente de determinação** ( $0 < R^2 < 1$ ) e o **coeficiente de correlação** ( $\rho$ ). [Tabela 5]

Após a visualização do diagrama de dispersão e tendo em conta o **valor do coeficiente de determinação** é de constatar a ausência de correlação entre as variáveis número de focos de incêndio e a média dos declives, não se verificando qualquer relação estatística linear [Figura 16C]. Isto ocorre em parte por se considerar a média como uma medida pouco significativa dos declives por conchelo.

Observando o **valor de  $R^2$** , próximo de zero [Figura 16C], e tendo este como medida da variação da variável dependente que indica que o modelo não se ajusta à amostra, não existe assim uma fundamentação para os valores observados. Por outro lado, esta situação revela novamente a fragilidade e incoerência realizada por esta análise, em parte pelo arranjo/ preparação dos dados e pela escolha frágil da variável. O contrário ocorre quanto mais este valor é próximo de um. As outras variáveis apresentam uma correlação muito mais elevada em cerca de 70%.

A análise de correlação revela assim que a **densidade populacional**, muito surpreendentemente, e o **nº de crimes são variáveis mais poderosas e que apresentam correlação**, o que permite efetuar uma comparação com recurso a uma regressão múltipla, por exemplo.

De modo a prosseguir para uma regressão espacial clássica, considerou-se apenas as variáveis que apresentavam o coeficiente de determinação mais elevado, nomeadamente as variáveis independentes: nº de crimes e densidade populacional [Tabela 5].

## 5.2 Regressão Espacial [16]

Um outro foco de interesse no estudo da regressão clássica é a sua adaptação aos fenómenos espaciais, tal é realizado assumindo a suposição de **hipótese padrão** para um conjunto de dados espaciais **da não correlação das variáveis, isto é, se são independentes e seguem uma distribuição normal**. Contudo, no mundo real, a dependência espacial existe e é demonstrada muitas vezes através dos **resíduos**, uma vez que estes não são **independentes** e encontram-se **correlacionados**.

Para tal, a abordagem a seguir, depois da escolha das variáveis, será o **Modelo de Regressão Espacial Clássica, OLS**, onde se irão mapear os resíduos de forma a visualizar a sua distribuição e medir a **autocorrelação espacial** presente com os índices de *Geary* e *Moran* utilizados anteriormente.

### OLS

O método de seleção do modelo é dado pelos valores de máxima verossimilhança (*log likelihood*) dos diferentes modelos utilizados e *Akaike info Criterion (AIC)*, para além do coeficiente de determinação [Tabela 5].

Enquanto o *AIC*, uma medida da qualidade relativa dos modelos estatísticos, é representada pelo seu **menor valor possível, log likelihood** que determina os valores ideais dos coeficientes estimados, é representada pelo seu **maior valor possível**.

O **modelo de regressão múltipla [Equação 3]** revela assim um maior valor de  $R^2$  e de *log likelihood*, embora todos os valores estejam muito próximos entre si.

Tabela 6 Relatório estatístico da regressão múltipla para ajustamento da equação. Dados obtidos através do recurso ao software GeoDa

Variable	Coefficient	Std. Error	t - Statistic	Probability
Constant	169.089	90.2094	1.87441	0.09774
Nº Crimes	0.558035	0.673863	0.828113	0.43162
População	0.00400382	0.0134714	0.29721	0.77388

$$Y = 169.089 + 0.558X_1 - 0.04X_2 + \varepsilon$$

Equação 3 Modelo de Regressão Múltipla, neste caso para duas variáveis

- **Análise dos resíduos:**

Com a utilização do software GeoDa obteve-se assim os diagnósticos aos resíduos pelos relatórios de ajustamento do modelo de regressão. Nos diagnósticos teve-se em conta: a **condição de multicolinearidade** (demonstra nível de inter-relação entre as variáveis independentes), o **teste de Jarque-Bera** (teste estatístico - *chi quadrado*), a **distribuição normal** aos erros (**H0: distribuição dos resíduos é normal**), o **diagnóstico de heterocedasticidade** (permite detetar variâncias não constantes).

```

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER    13.591477
TEST ON NORMALITY OF ERRORS
TEST          DF          VALUE          PROB
Jarque-Bera   2          4.6741          0.09661
    
```

Figura 17 Resultados do diagnóstico para os resíduos obtidos com o GeoDa no modelo de regressão com as condicionantes (multicolinearidade, teste de Jarque – Bera, distribuição normal aos erros)

De acordo com os diagnósticos obtidos o valor de **multicolinearidade** encontra-se abaixo de 20 como é esperado, e a baixa probabilidade no teste de **Jarque-Bera** indica a distribuição não normal dos erros/resíduos, demonstrando a interferência pela componente espacial. [17]

```

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST          DF          VALUE          PROB
Breusch-Pagan test    2          7.7782          0.02046
Koenker-Bassett test  2          3.3879          0.18379
SPECIFICATION ROBUST TEST
TEST          DF          VALUE          PROB
White          5          8.4827          0.13156
    
```

Figura 18 Relatório dos resultados do diagnóstico para os resíduos obtidos com o GeoDa para a heterocedasticidade

A baixa probabilidade dada pelo diagnóstico de **heterocedasticidade** não é necessariamente uma surpresa, revelando que a variância dos resíduos é afetada pela dependência espacial dos dados, **rejeitando** assim a **hipótese padrão** referida no ponto 5.2.

Após a obtenção dos valores dos *OLS residuals* na tabela de atributos, procedeu-se à sua análise *versus* as respetivas variáveis independentes consideradas: nº da População e nº de crimes. [Figura 19 A e B]

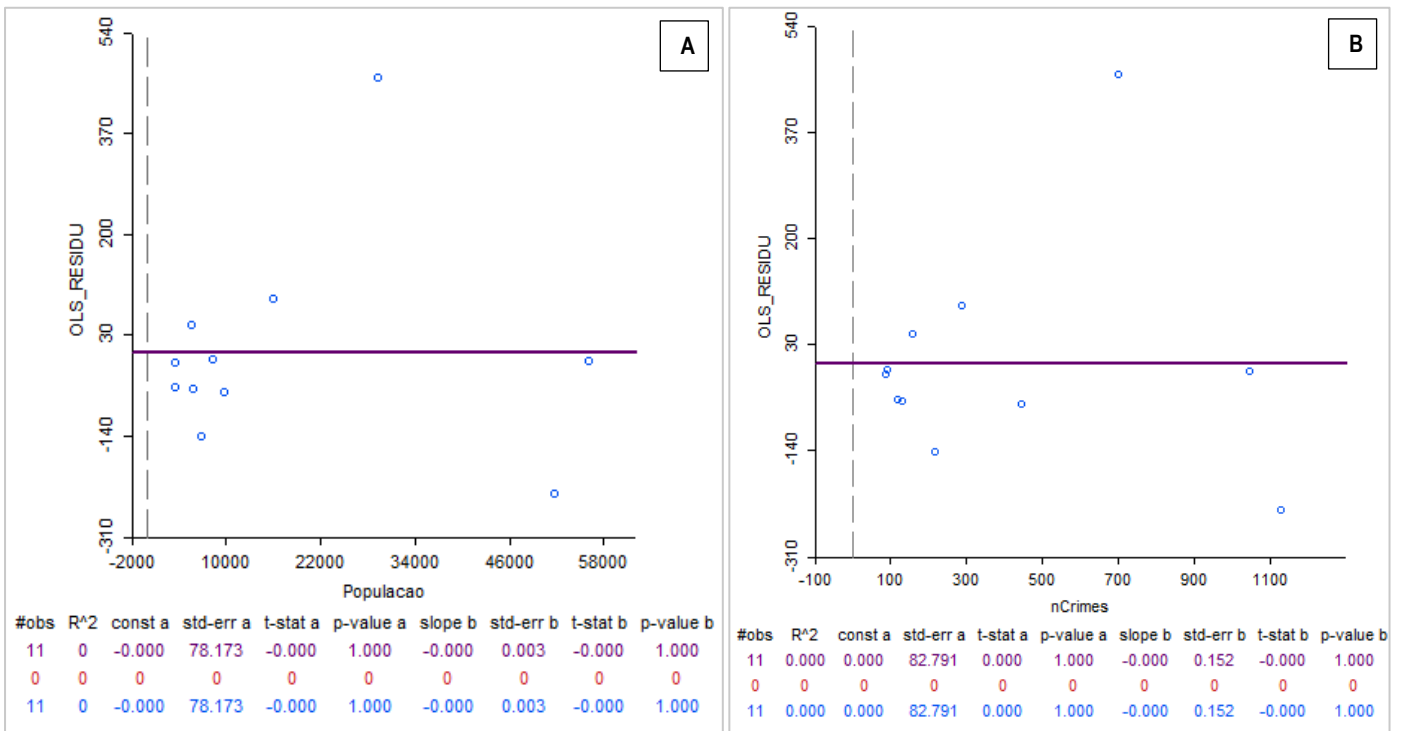


Figura 19 Diagrama de dispersão dos Resíduos vs variável independente: (A) População; (B) número de crimes. Imagens obtidas através do software GeoDa

Embora se observe a ausência de correlação (independência) dos valores dos resíduos, estes não seguem uma **distribuição normal com média “0” e variância constante**, o que revela que a componente espacial interfere, sendo assim necessário realizar a análise de **autocorrelação dos resíduos**.

- **Análise Autocorrelação dos resíduos**

Com o índice de *Moran* local procedeu-se então aos diagnósticos dos resíduos, de modo a verificar a existência de autocorrelação.

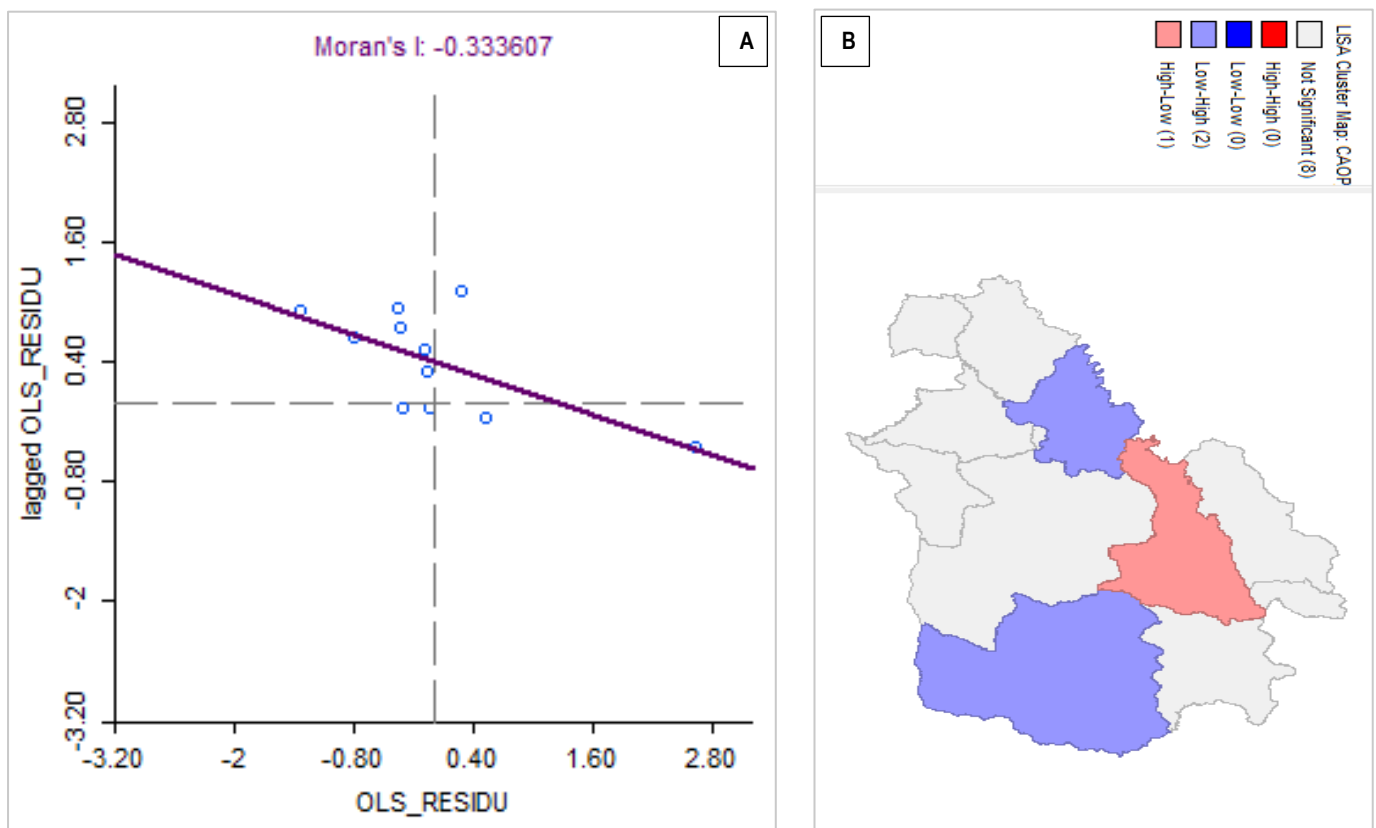


Figura 20 (A) Índice local de Moran aplicado aos resíduos e (B) CAOP . Imagens obtidas através do software GeoDa



Como o valor do índice de *Moran* é de -0.33 (próximo de 0) revela-se assim a ausência de autocorrelação, e consequentemente a inexistência de dependência espacial entre as variáveis propostas na regressão múltipla.

Deste modo, e após a autocorrelação obtida entre as variáveis, é possível realizar modelos de regressão **globais**, utilizando um **único parâmetro que captura a estrutura de correlação espacial**, e modelos **locais**, onde os **parâmetros variam continuamente no espaço**:

- **Modelos Espaciais Globais – *Spatial Lag Models e Spatial Error Models*** - Neste caso, a autocorrelação espacial é atribuída, respetivamente, à variável dependente Y e no outro, atribuída ao erro. [18]

- **Modelos Espaciais Locais — Modelos de regressão com regimes espaciais ou modelos com efeitos espaciais contínuos - *GWR (Geographically Weighted Regression)*** - Onde as variações são modeladas de forma discreta e nos outros de forma contínua. No caso do *GWR* é ajustado o modelo a cada área, ponderando todas as observações em função da distância desta área aos seus vizinhos. [19]

Devido à ausência de autocorrelação obtida pelas variáveis, não houve modo de realizar o método de seleção de máxima verossimilhança, e, ao contrário do referido, não foi possível proceder, nem com **modelos espaciais globais**, nem com **modelos espaciais locais**.

No entanto, como nota de curiosidade/aprendizagem e sabendo que o efeito esperado não corrobora com a análise pretendida, nem apresenta significado estatístico válido, procedeu-se à realização de outros modelos espaciais no *software*.

Na aplicação dos modelos para *GWR*, ***Spatial LAG*** e ***ERROR*** usou-se a *CAOP* por concelhos, onde a variável dependente continuou a ser o número de fogos, testando como variáveis explanatórias, o nº de crimes e a população, comparando com o **Modelo OLS** previamente criado. Para os modelos **Globais** utilizou-se o *software GeoDA* obtendo os relatórios de diagnóstico com os respetivos dados [Tabela 7]; e para visualização realizou-se o *GWR* e o *OLS* em *ArcGIS*, pois eram as únicas ferramentas acessíveis. [Figura 21]

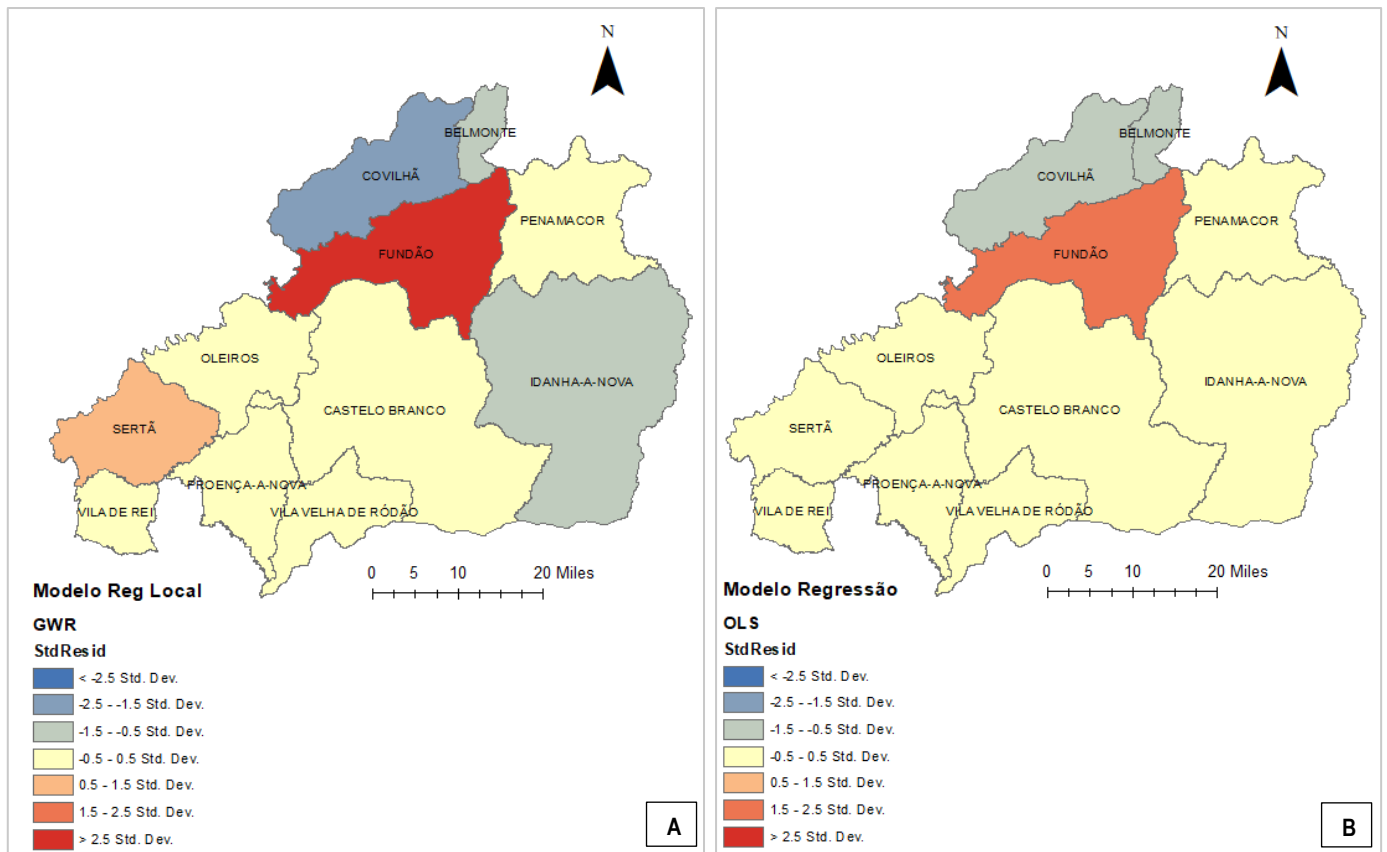
Tabela 7 Comparação do Modelo de Regressão Clássico com os Modelos espaciais de efeito global

Modelos de Regressão	R <sup>2</sup>	Log likelihood	AIC
OLS	73.10%	-72.00	150.00
<i>Spatial Lag</i>	81.63%	-70.33	148.66
<i>Spatial Error</i>	86.76%	-69.72	145.45

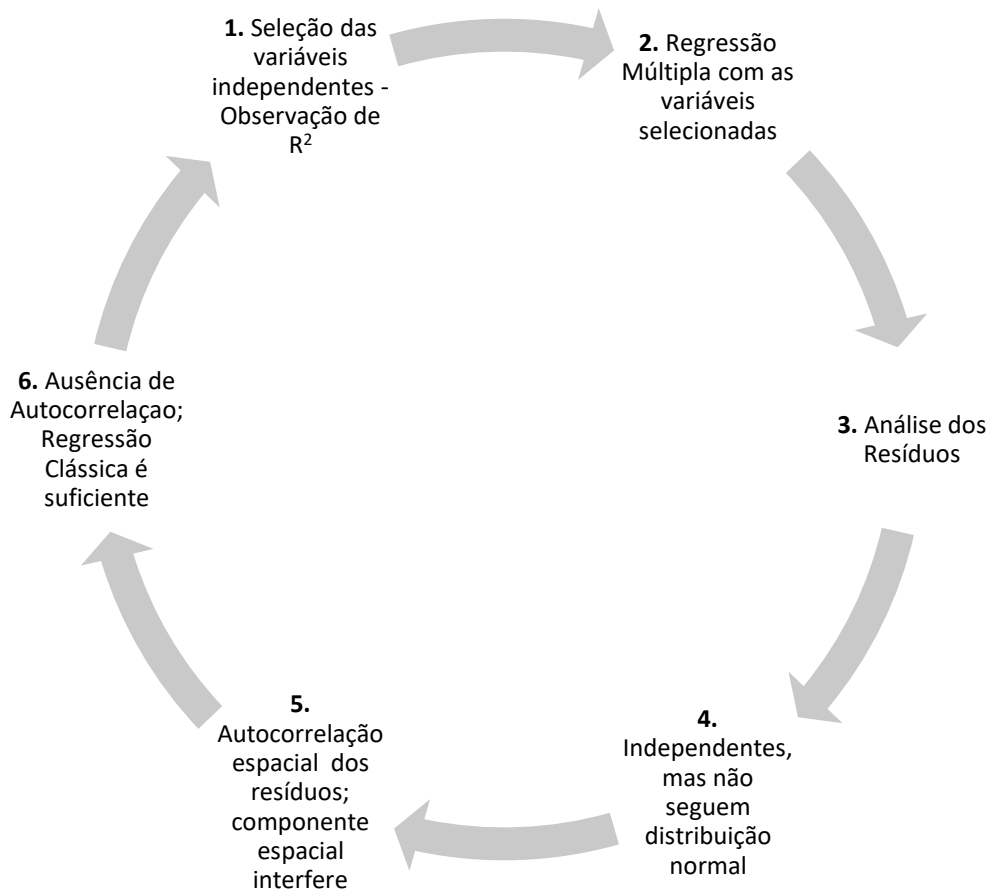
Comparativamente pelo **maior valor** de coeficiente de determinação e **pelo menor** de *AIC*, os modelos de regressão espaciais globais apresentam melhores resultados que o *OLS*. Embora a escolha do melhor modelo de regressão assenta-se no *Spatial Error*, uma vez que não foi detetada qualquer autocorrelação esta escolha não apresenta validade estatística, servindo só para complemento de informação e aprendizagem.

Após a realização das regressões no *ArcMap* é possível visualizar a estatística da distribuição espacial dos resíduos, onde o *GWS* ajusta o modelo a cada ponto, ponderando todas as observações em função da distância a partir desse mesmo ponto. Assim sendo a maioria dos concelhos estão dentro da média, exceto o Fundão e a Covilhã que se afastam, um com valores positivos e outro com valores negativos.

Conclui-se assim que a incoerência recai na escolha frágil das variáveis, que não explicam, nem influenciam diretamente os focos de incêndio. De modo a complementar este estudo deveriam ter sido escolhidas outras variáveis, mais propícias, ou em último caso assumir que a independência das variáveis não é simplesmente linear, ou até mesmo aceitar que a correlação não implica causalidade até os resultados serem confrontados com a realidade, particularmente no caso da variável independente da densidade de populacional em função do número de focos de incêndio.



**Figura 21** Exemplo do modelo realizado para GWR (A) e OLS (B). O resultado não apresenta qualquer validade estatística existente nas variáveis utilizadas; GWR é um modelo espacial com efeito local e o OLS o modelo de regressão clássica, sem qualquer motivo de comparação. Imagens obtidas através do software ArcGIS



**Figura 22** Resumo dos passos da regressão efetuada às variáveis previamente escolhidas

## 6. Interpolação

Para a realização do estudo da interpolação consideraram-se dados de componentes de análise atmosférica obtidas do site da *EEA (European Environment Agency)*[20], cujo objetivo **foi estimar os valores desconhecidos de pontos da amostra através de valores conhecidos da região**, onde, com o uso destes dados pontuais, obtidos de estações permanentes na medição dos valores atmosféricos de  $O_3$ , foi possível por interpolação espacial, representar superfícies contínuas com o objetivo de observar padrões espaciais. Os dados são referentes aos valores de concentração de  $O_3$  do ano de 2014. [Figura 23]

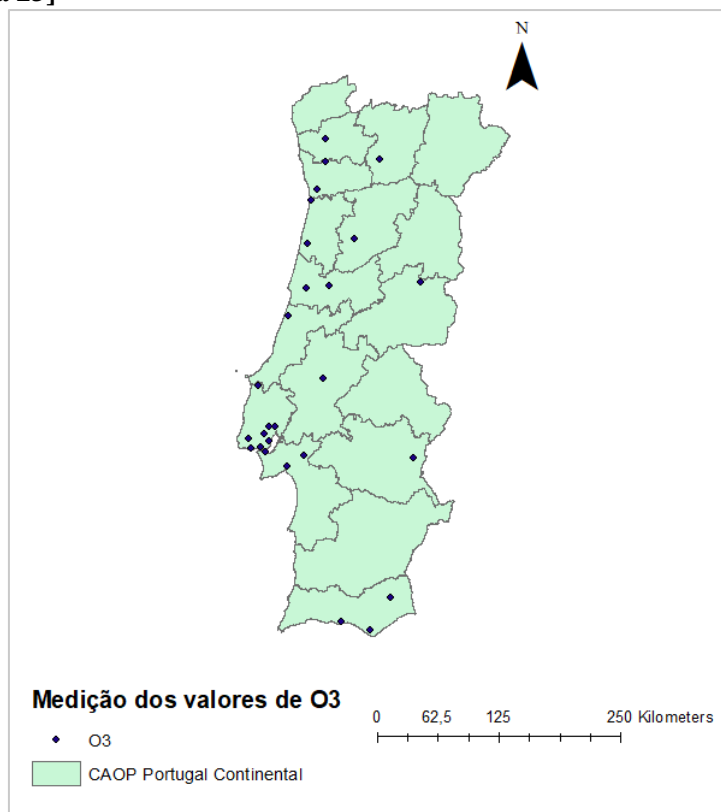


Figura 23 Medição e localização dos valores de  $O_3$ , obtidas do site da *EEA*. Imagens obtidas através do software *ArcGIS*

De modo a interpolar estes valores, teve-se em conta a sua **extensão espacial** por métodos **locais**, considerando um **conjunto local de pontos vizinhos** (capturar a variação), um tipo de **ajustamento exato** (estimar um valor idêntico ao valor observado), um **modelo determinístico** (superfície é interpolada a partir de um conceito de proximidade, sem estimativa dos erros dos valores estimados - *IDW* [21] e *Thiessen Polygons* [22]) e um **modelo probabilístico** (autocorrelação espacial com vista a modelar a variabilidade existente - *kriging* [23]).

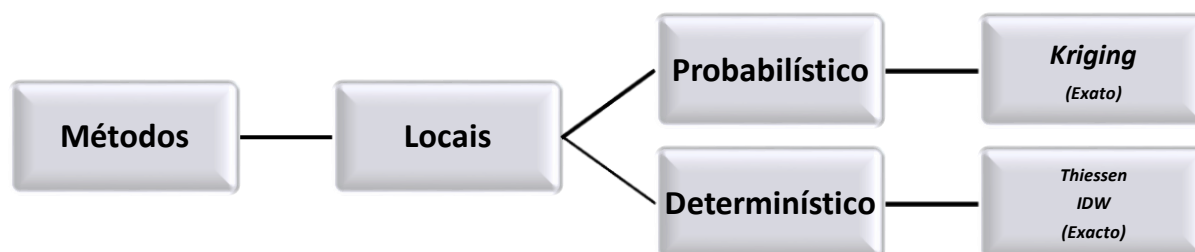


Figura 24 Fluxograma dos Métodos de interpolação usados

Com recurso ao *Geostatistical Analyst* [24] procedeu-se à interpolação dos dois métodos, um probabilístico, o *kriging*, com propriedades de não tendiocidade e outro determinístico, *IDW*. A interpolação por polígonos de *Thiessen* efetuou-se pela *toolbox* do software.

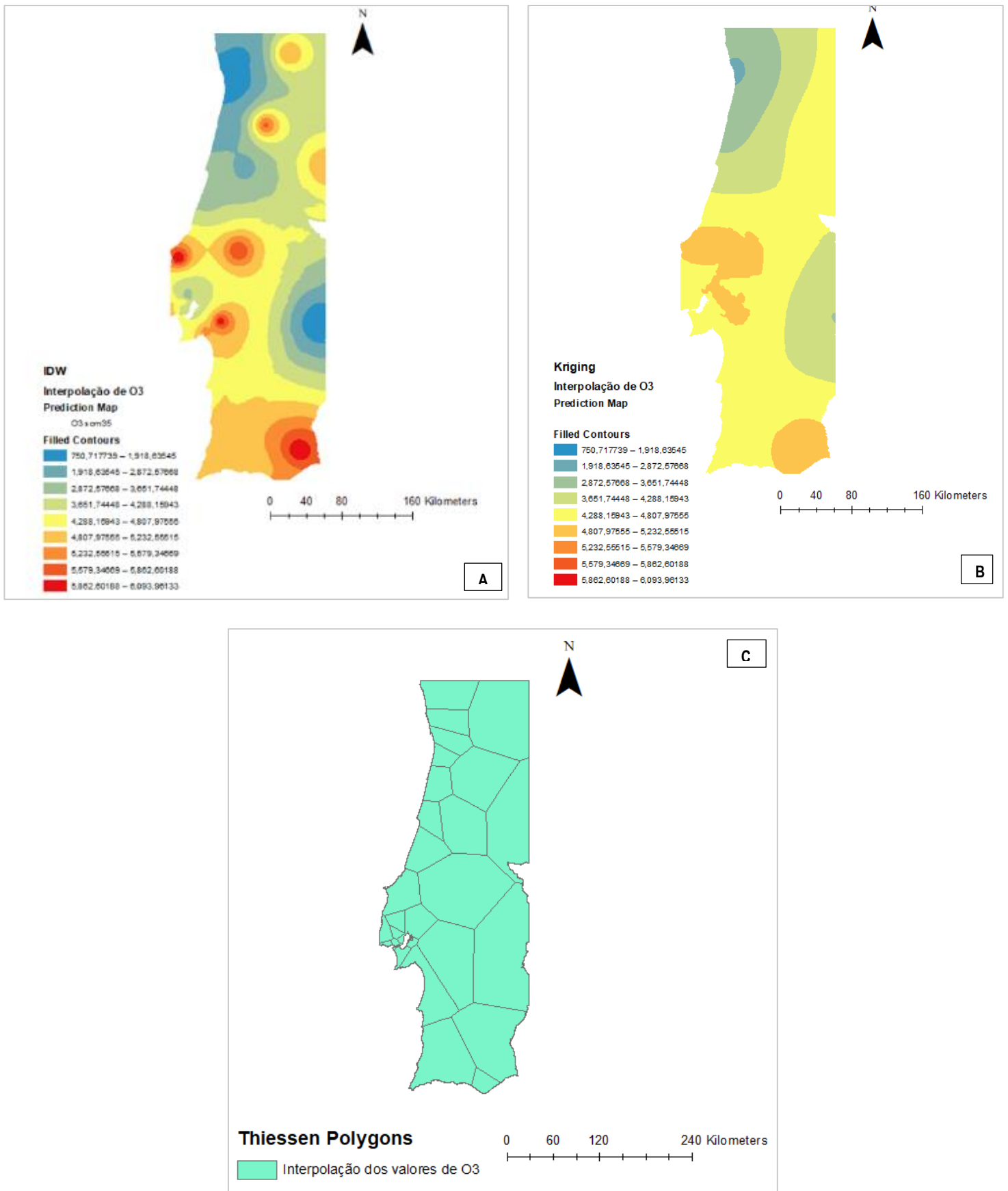


Figura 25 Métodos de interpolação realizados com o *Geostatistical Analyst*: (A) Método determinístico *IDW*; (B) Método probabilístico *Kriging Ordinary*; (C) Método determinístico *Polígonos de Thiessen*. Imagens obtidas através do software *ArcGIS*

A escolha dos valores a interpolar referentes somente a um ano e ao valor de *O<sub>3</sub> som35* tem em vista a análise e comparação pelos diferentes métodos de interpolação, não propriamente avaliar a evolução deste poluente atmosférico.

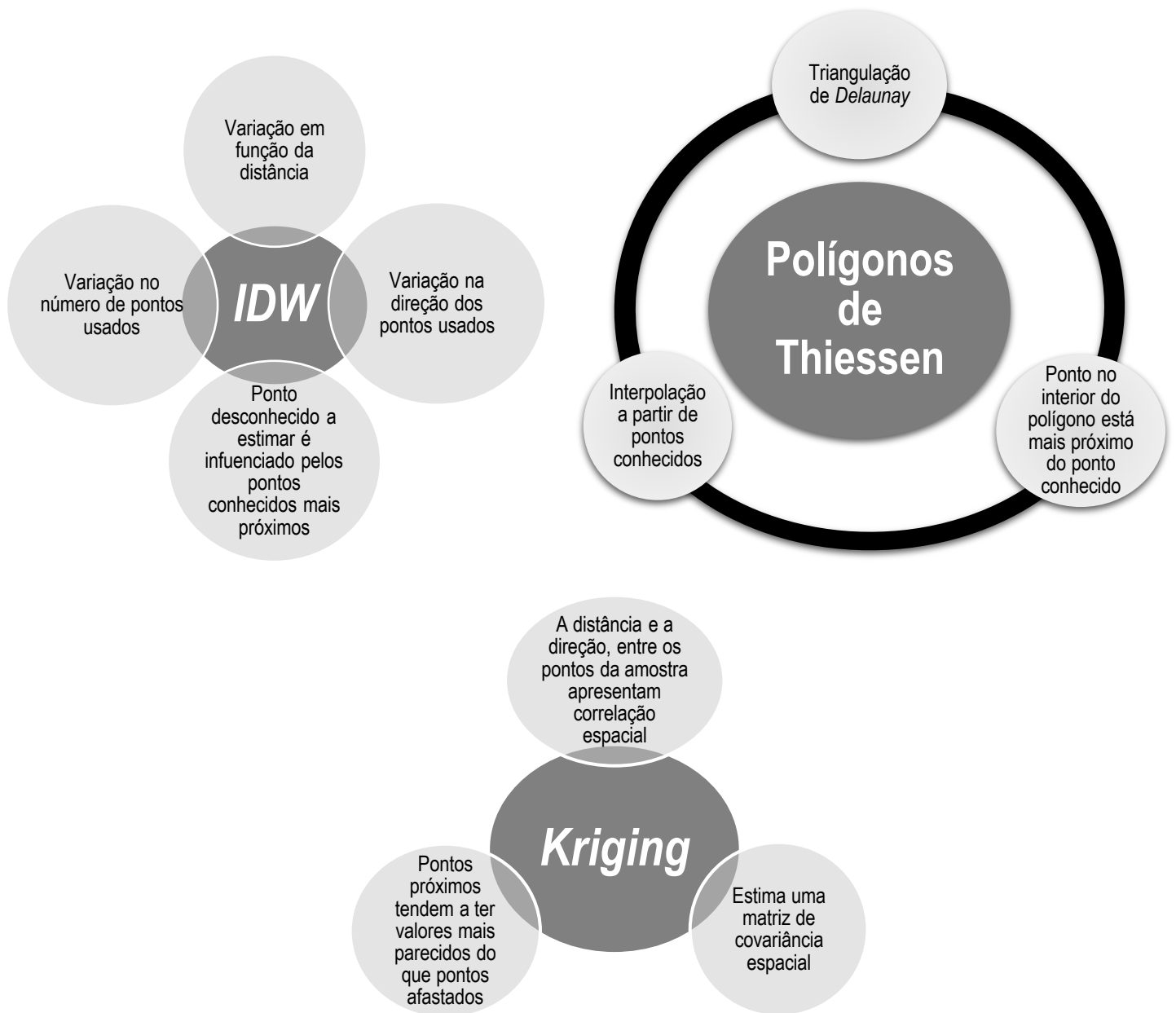


Figura 26 Características dos métodos de interpolação utilizados

Comparativamente, entre o método de interpolação de *kriging* e *IDW*, embora tivessem sido efetuados com o mesmo número de classes (*filled contours*), com o mesmo número de pontos e distância, mantendo as definições *standard* do *software*, é de se notar os diferentes contrastes visuais na superfície contínua.

O *IDW* demonstra realmente que os pontos com a distância mais próxima uns dos outros, revelam uma maior influência na interpolação, sendo facilmente possível avaliá-los na superfície contínua nos tons de cor mais quente [Figura 25A].

O *Kriging Ordinary*, assume um conceito de média ponderada e proximidade, onde não há tendência a determinar a interpolação entre os dados, tal é observável numa interpolação mais discreta, nunca assumindo valores máximos [Figura 25B].

Os Polígonos de *Thiessen* oferecem uma cobertura de pontos em polígonos de toda a área, tendo em conta que qualquer valor estimado dentro do polígono está mais próximo do ponto conhecido dentro do mesmo, demonstrando os vários polígonos diferenciais. [Figura 26C]



## 7. Redes

São usadas, de modo a revelar direção e movimento entre localizações pontuais, sendo representadas por **nós** (localizações e intersecções) e por **arcos** (ligação entre os mesmos).

Encontram-se responsáveis pela organização de todos os elementos geométricos numa estrutura vetorial, sendo caracterizadas por um grafo, com uma topologia evidente de **conetividade** e **diretividade**. Utiliza-se uma matriz de adjacência, onde cada arco, normalmente, tem associado um peso, contribuindo de diferentes maneiras à resposta da rede a determinados algoritmos. [25]



Figura 27 Algoritmos existentes no *Network Analyst* [26]

Neste último capítulo, e de modo a proceder à utilização destes algoritmos, na análise de redes, recorreu-se ao tratamento dos dados da rede viária do distrito de Castelo branco fornecida pelo *Open Street Map* [27]. Aos dados das estações da corporação de bombeiros, para o mesmo distrito, as coordenadas obtidas X,Y em *.txt* foram adicionadas com “*add X Y data*” e convertidas do sistema de projeção de *Hayford Gauss* para *ETRS 1989 PT-TM06*.

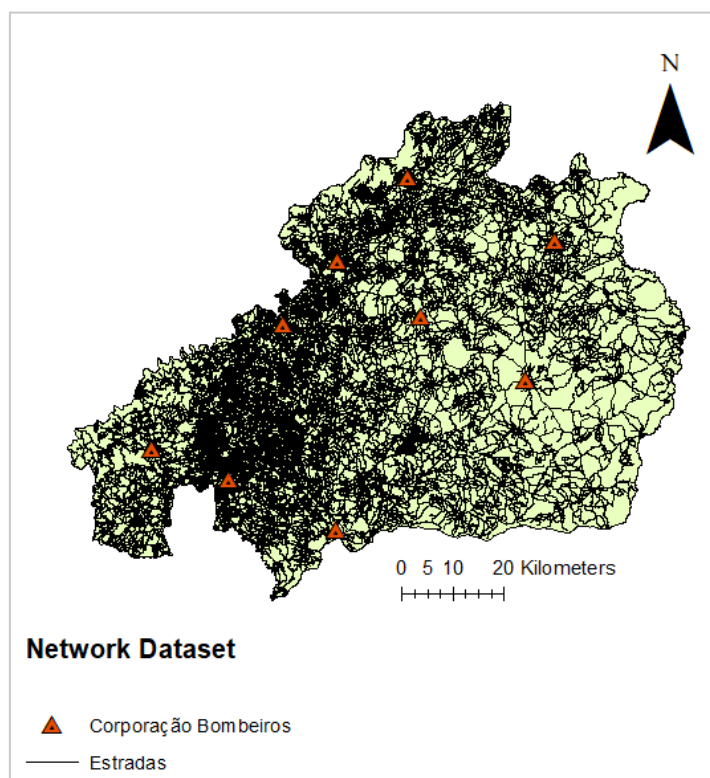


Figura 28 Visualização da rede viária (topologicamente não tratada) e da localização das corporações de Bombeiros. Imagens obtidas através do *software ArcGIS*

Após a ativação *Network Analyst* [Figura 29] criou-se uma *New File geodatabase* onde foi criada uma *feature dataset* e dentro dessa um *Network Dataset*, com os dados previamente mencionados.

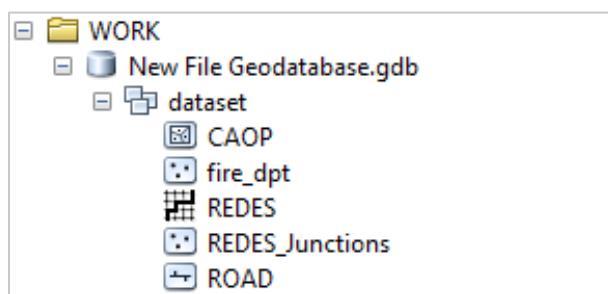


Figura 29 Criação do *Network Dataset*

Após a criação da *Network Dataset* foi possível verificar todas as *Junctions* existentes (é de se notar a caótica *junction* causada pela ausência de conectividade). [Figura 30]

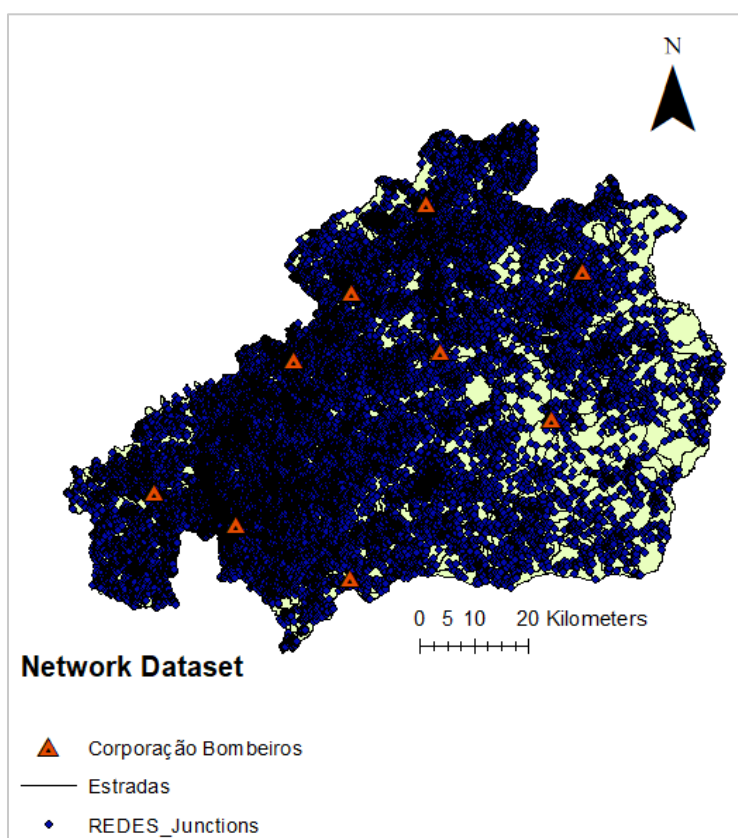


Figura 30 Visualização das *Junctions* após a criação do *Network Dataset*. Imagens obtidas através do software ArcGIS

Como a rede não se encontrava topologicamente tratada (por este processo ser moroso e estar fora do âmbito da análise da disciplina), tornou-se inviável as análises dos algoritmos, bem como os valores de impedância a serem adaptados.

Embora a ideia original fosse analisar o padrão previamente estudado dos focos de incêndio pela localização das corporações de bombeiros e obter respostas a partir dos algoritmos, tendo em conta a rede viária adjacente a estas localizações (por exemplo: qual o *closest facility* de um determinado distrito; ou sabendo que a área de incidência do maior número de fogos é na zona sudoeste, qual o *optimal path* entre a corporação de bombeiros mais próxima e determinado ponto em questão), infelizmente, esta ideia, não foi possível ser concretizada.

De qualquer modo, em seguida, serão mostradas algumas imagens que demonstram a tentativa de visualização dos algoritmos *Route*, *Service Area* e *Closest Facility*. [Figura 31]

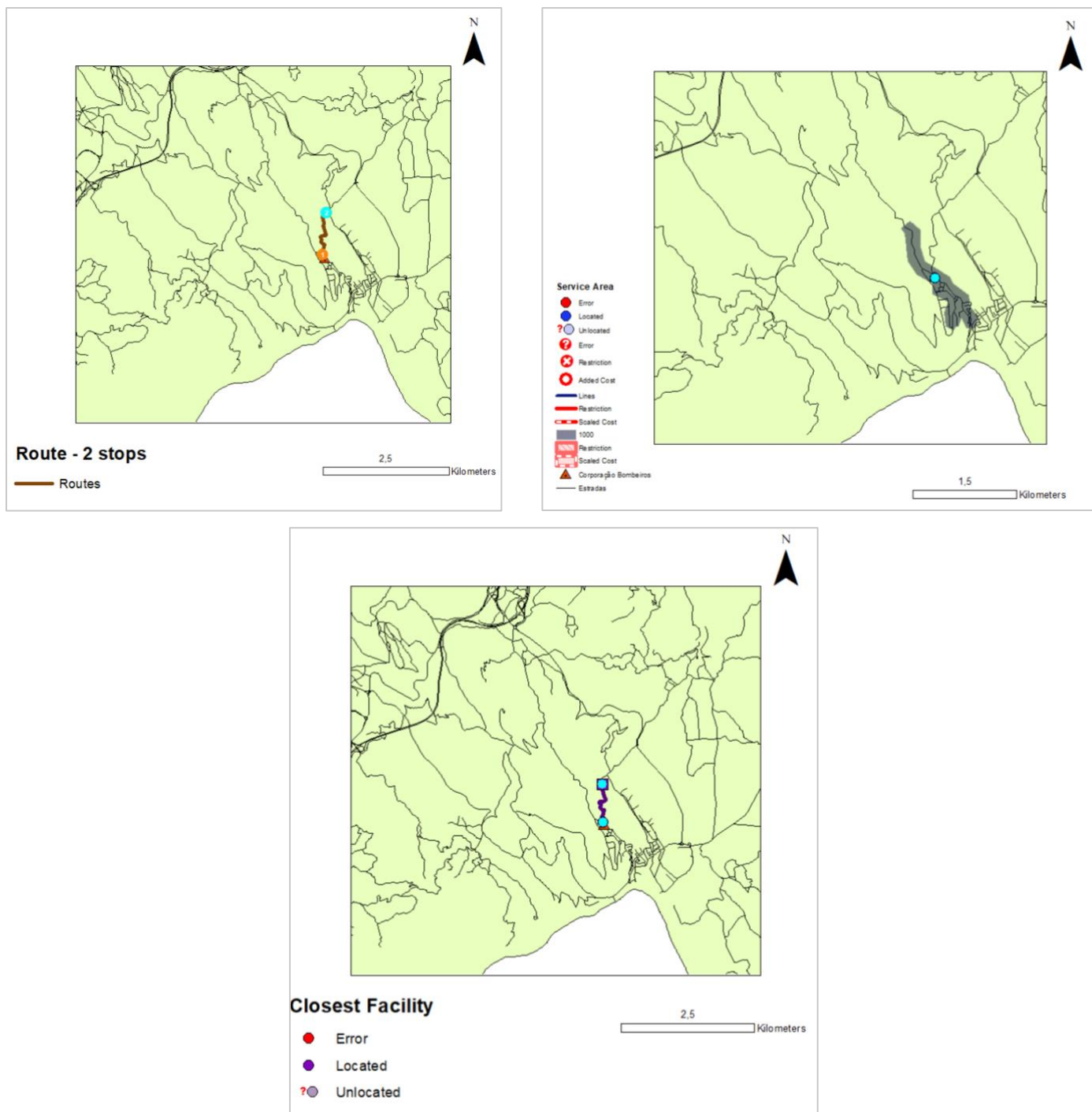


Figura 31 Exemplificação de imagens obtidas no processo de criação de (A) Route, (B) Service Area e (C) Closest Facility. Imagens obtidas através do software ArcGIS

Devido à falta de conectividade dos nós na rede viária, não é possível que os algoritmos corram de modo a demonstrar os efeitos pretendidos, e consequentemente, não sendo viáveis em qualquer análise *à posteriori*.

De qualquer modo procedeu-se à criação de uma *route*, com duas paragens [Figura 31A], calculou-se a área de serviço para uma determinada *facility* (Corporação de Bombeiros), procedendo-se ao cálculo da *closest facilities* [Figura 31B], e, assumindo um incidente na rede, a corporação de bombeiros mais próxima identificada, através do cálculo da *route* permitida [Figura 31C].

## 8. Conclusão

A tarefa de obtenção de diferentes dados com características suficientes para serem profundamente analisados por métodos estatísticos e com a utilização de ferramentas previamente mencionados não foi alcançada conforme inicialmente se pretendia. No entanto, de modo a prosseguir com o estudo geoestatístico da análise espacial de dados, utilizaram-se os dados fornecidos pela docente e revelados ao longo deste projeto.

Nesta análise utilizaram-se os *softwares* GeoDa e ArcGIS.

Na **análise do padrão pontual** foi possível, através de diferentes técnicas observar-se um padrão de tipo **aglomerado** pelas técnicas que envolviam o parâmetro da **densidade da distribuição dos focos**, bem como pela confirmação dos **testes de significância** estatística por **indicadores globais e locais**.

Em relação à **análise de áreas**, após a sua visualização por métodos de corte, que teve como pressuposto a criação de uma matriz de vizinhança, onde por efeitos de 1º ordem confirma-se a tendência de **aleatoriedade** do nº de focos de incêndio, não revelando **autocorrelação espacial existente (isto é aceita-se H<sub>0</sub>)**, pois são **espacialmente independentes o nº de focos por área**. Relativamente aos efeitos de 2º ordem, foi possível identificar os *clusters* existentes na área de referência por interpretação dos valores vizinhos.

Ambas as estatísticas, *G* e *Moran*, supõem a **normalidade** nas suas distribuições na construção de intervalos de confiança para a **hipótese nula de total ausência de autocorrelação local**. No entanto, medem coisas fundamentalmente diferentes, uma a **variância** e a outra a **soma**, como referido.

Na **regressão**, não foi possível obter resultados conclusivos pela variação dos modelos existentes, para além que de acordo com os dados apresentados a regressão clássica era suficiente para explicar as variáveis. Embora houvesse independência das variáveis, a não distribuição normal apresentada, levou à necessidade de analisar a autocorrelação, que acabou por não ter sido detetada pelos modelos espaciais. E que embora a correlação obtida pela densidade populacional, seja elevada, esta não implica um efeito de causalidade.

A **interpolação** é um método muito eficaz e representativo na criação da superfície contínua por valores conhecidos, sendo o *IDW*, um método diferenciativo para os dados usados.

A **aplicação das redes** foi dificultada pela ausência de uma rede topologicamente bem elaborada. Não tendo sido possível obter resultados, procedeu-se à prática dos algoritmos mencionados, de modo a demonstrar uma possível aplicação aos dados, pretendendo elucidar uma futura resposta no combate a incêndios, tendo em conta a localização das estações e a conectividade da rede viária.

Embora o principal objetivo deste projeto fosse a aprendizagem dos conceitos das ferramentas estatísticas no estudo dos diferentes tópicos, foi possível realmente em alguns casos tirar conclusões colaborativas destas, e perceber a forte influência e importância da geoestatística em vários estudos como resposta numa análise comparativa a partir de dados reais.

## Referências Bibliográficas

- [1] Fischer M. M, Getis A. (2018) *Handbook of applied spatial analysis : software tools, methods and applications*. 6th Edition, Springer. New York;
- [2] Kawamoto, M. T. (2012) *Análise de técnicas de distribuição espacial com padrões pontuais e aplicação a dados de acidentes de trânsito e a dados da dengue de Rio Claro-SP*. Tese de Mestrado. Universidade Estadual Paulista Julio de Mesquita Filho, Instituto de Biociências de Botucatu, São Paulo;
- [3] Luc Anselin - *Spatial Structures in the Social Sciences*. [Consult. 01 Maio 2018] Disponível em <https://s4.ad.brown.edu/Resources/Tutorial/Modul2/GeoDa2.pdf>;
- [4] ESRI – *Technical Support*. [Consult. 01 Maio 2018] Disponível em <https://support.esri.com/en/knowledge-base>;
- [5] Grupo CRISE (1998-2000) – *Cruzamento de dados (Cooperações encontradas)*. [Consult. 01 Maio 2018] Disponível em [http://scrf.igeo.pt/asp/cb\\_cr.asp?x=253831&y=316986](http://scrf.igeo.pt/asp/cb_cr.asp?x=253831&y=316986);
- [6] Francisco José Reyes Peralta (2010) - *Análisis Espacial con Datos Raster en ArcGIS Desktop 9.2*. [Consult. 05 Maio 2018] Disponível em <https://drive.google.com/file/d/0Bz2JARLkKYt-akhsRE13MnRBa0E/view>;
- [7] Correia, J. D., Silva, N.C.C., Oliveira, L. C., Ferreira, L. F. (1996) *Centro Geodésico e Centróide: Uma Abordagem Conceitual*. Cartografia e Cadastro, Nº5 Consult. Maio 2018] Disponível em [https://ww2.ibge.gov.br/home/geociencias/geodesia/artigos/1996\\_Centro\\_Geodesico\\_Centroide.pdf](https://ww2.ibge.gov.br/home/geociencias/geodesia/artigos/1996_Centro_Geodesico_Centroide.pdf);
- [8] Manuel Gimond (2017) - *Intro to GIS and Spatial Analysis: Chapter 11.2.2.1 Quadrat Density* [Consult. Maio 2018] Disponível em <https://mqimond.github.io/Spatial/hypothesis-testing.html>;
- [9] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. 1st Edition Chapman and Hall. London ISBN 978-0521460866;
- [10] Anselin, L. - *Point Pattern Analysis Nearest Neighbor Statistics* [Consult. 06 Maio 2018] Disponível em [https://spatial.uchicago.edu/sites/spatial-data.uchicago.edu/files/9\\_points\\_2\\_r.pdf](https://spatial.uchicago.edu/sites/spatial-data.uchicago.edu/files/9_points_2_r.pdf) ;
- [11] Manuel Gimond (2017) - *Intro to GIS and Spatial Analysis: Chapter 12. Hypothesis testing*. [Consult. Maio 2018] Disponível em <https://mqimond.github.io/Spatial/hypothesis-testing.html>;
- [12] Stephanie (2016) - *Moran's I: Definition, Examples. USA: statisticshowto* [Consult.06 Maio 2018] Disponível em <http://www.statisticshowto.com/morans-i/>;
- [13] M.Sawada (2009) - *Global Spatial Autocorrelation Indices - Moran's I, Geary's C and the General Cross-Product Statistic*. [Consult. 06 Maio 2018] Disponível em <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>;

- [14] Hazewinkel, M. (2012) - *Regression analysis, Encyclopedia of Mathematics*. [Consult. Maio 2018] Disponível em [http://www.encyclopediaofmath.org/index.php?title=Regression\\_analysis&oldid=28558](http://www.encyclopediaofmath.org/index.php?title=Regression_analysis&oldid=28558) ISBN 978-1-55608-010-4;
- [15] Dr. Iain Pardoe (2018) - *Simple Linear Regression*. [Consult. 10 Maio 2018] Disponível em <https://onlinecourses.science.psu.edu/stat501/node/250/>;
- [16] Eugene Brusilovskiy (2009) - *A Brief Introduction to Spatial Regression. USA: Business Intelligence Solutions* [Consult. 10 Maio 2018] Disponível em <http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Regression.php>;
- [17] Jarque, C. M., and A. K. Bera. A (1987) Test for Normality of Observations and Regression Residuals. *International Statistical Review*. **55**: 163-172;
- [18] Darmofal, D. (2015) *Spatial Lag and Spatial Error Models. In Spatial Analysis for the Social Sciences (Analytical Methods for Social Research)*. 1st Edition Cambridge: Cambridge University Press;
- [19] Martin Charlton (1997-2017) - *Geographically Weighted Regression A Tutorial on using GWR in ArcGIS 9.3*. [Consult. Maio 2018] Disponível em [https://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr\\_arcgis/GWR\\_Tutorial.pdf](https://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr_arcgis/GWR_Tutorial.pdf);
- [20] Agência Europeia do Ambiente (EEA) (2018) – *Data and maps. EU* [Consult. Maio 2018] Disponível em <https://www.eea.europa.eu/data-and-maps>;
- [21] GISGeography (2018) - *Inverse Distance Weighting (IDW) Interpolation*. [Consult. Maio 2018] Disponível em <https://gisgeography.com/inverse-distance-weighting-idw-interpolation/>;
- [22] ESRI (2017) - *Create Thiessen Polygons*. [Consult. Maio 2018] Disponível em <http://pro.arcgis.com/en/pro-app/tool-reference/analysis/create-thiessen-polygons.htm>;
- [23] GISGeography (2018) - *Kriging Interpolation – The Prediction Is Strong in this One*. [Consult. Maio 2018] Disponível em <https://gisgeography.com/kriging-interpolation-prediction/>;
- [24] ESRI - *Introduction to the ArcGIS Geostatistical Analyst Tutorial*. [Consult. Junho 2018] Disponível em <http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/introduction-to-the-arcgis-geostatistical-analyst-tutorial.htm>;
- [25] ESRI (1995-2012) - *What is Network Analyst?* [Consult. Junho 2018] Disponível em <http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//004700000001000000>;
- [26] ESRI (1995-2010) - *Network Analyst Tutorial*. [Consult. Junho 2018] Disponível em <http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/network-analyst-tutorial.pdf>;
- [27] Open Street Map (2018) - *OpenStreetMap powers map data on thousands of web sites, mobile apps, and hardware devices*. [Consult. Junho 2018] Disponível em <https://www.openstreetmap.org/#map=7/39.602/-7.839>;
- [28] Cristina Maria Sousa Catita, Faculdade de Ciências da Universidade de Lisboa (2018) - *Análise Espacial de Informação Geográfica*. [Consult. Maio- Junho 2018] Disponível em <https://fenix.ciencias.ulisboa.pt/courses/aeigeo-284554468266479/sumarios?p=5>;



## 9. Anexos

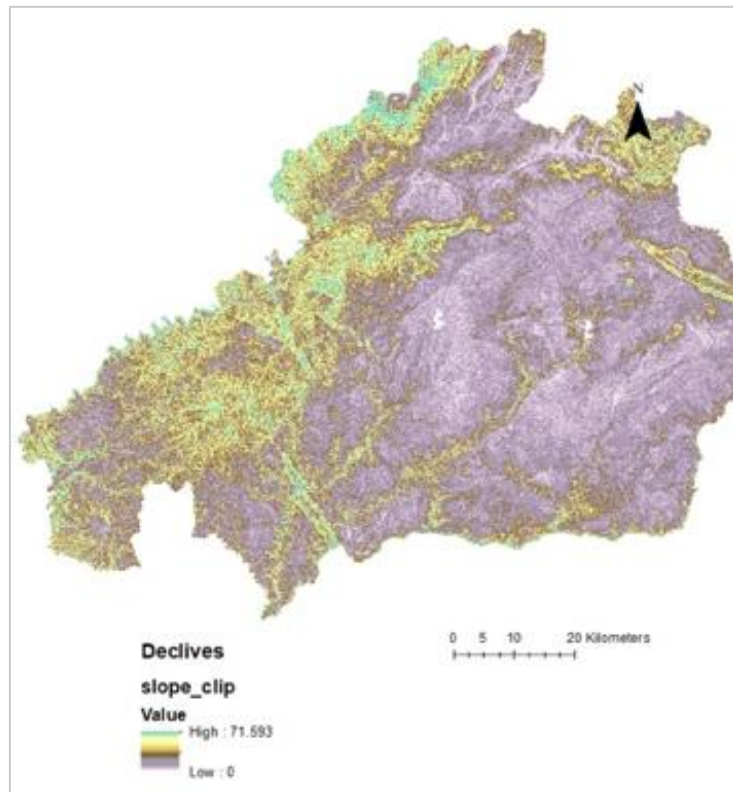


Figura 32 Declive do terreno - Castelo Branco. Imagens obtidas através do software ArcGIS

A média dos declives por concelho foi obtida pelo cálculo do *Slope* através do *DTM*. Neste, aplicou-se a ferramenta *zonal statistics*, calculando a média do declive, onde de seguida efetuou-se o *JOIN* com a *CAOP* utilizada. Os dados da população foram obtidos para cada freguesia, com a opção *sumarize* da tabela de atributos obteve-se os dados da população para cada concelho (efetuou-se o mesmo procedimento para a variável número de crimes).

De seguida importou-se a *CAOP* com os respetivos atributos das variáveis independentes, juntamente com a variável independente para o software *GeoDa* de modo a efetuar a análise de regressão.

```

-----
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set          : Caop_2
Spatial Weight    : CAOP_atr
Dependent Variable : Numero_foc  Number of Observations: 11
Mean dependent var : 463.181818  Number of Variables   : 3
S.D. dependent var : 324.851524  Degrees of Freedom    : 8
Lag coeff. (Lambda) : -0.940917

R-squared        : 0.867644  R-squared (BUSE)      : -
Sq. Correlation  : -          Log likelihood         : -69.722812
Sigma-square     : 13967.4   Akaike info criterion : 145.446
S.E of regression : 118.184   Schwarz criterion     : 146.639
-----

```

Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	228.5	46.8727	4.87491	0.00000
Populacao	-0.00798186	0.00726111	-1.09926	0.27165
nCrimes	1.04806	0.309019	3.39157	0.00070
LAMBDA	-0.940917	0.250498	-3.75618	0.00017

Figura 33 Output do Spatial Error Model. Relatório obtido através do software GeoDa



```

-----
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set           : Caop_2
Spatial Weight     : CAOP_atr
Dependent Variable : Numero_foc  Number of Observations: 11
Mean dependent var : 463.182   Number of Variables   : 4
S.D. dependent var : 324.852   Degrees of Freedom    : 7
Lag coeff. (Rho)  : -0.498579

R-squared          : 0.816305   Log likelihood         : -70.3312
Sq. Correlation    : -          Akaike info criterion  : 148.662
Sigma-square       : 19385     Schwarz criterion      : 150.254
S.E of regression  : 139.23
-----

```

Figura 34 Output do Spatial Lag Model - Relatório obtido através do software GeoDa

```

-----
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set           : Caop_2
Dependent Variable : Numero_foc  Number of Observations: 11
Mean dependent var : 463.182   Number of Variables   : 3
S.D. dependent var : 324.852   Degrees of Freedom    : 8

R-squared          : 0.731016   F-statistic           : 10.8708
Adjusted R-squared : 0.663770   Prob(F-statistic)    : 0.00523488
Sum squared residual: 312240   Log likelihood        : -72.0033
Sigma-square       : 39030.1   Akaike info criterion : 150.007
S.E. of regression : 197.56   Schwarz criterion     : 151.2
Sigma-square ML    : 28385.5
S.E of regression ML: 168.48
-----

```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	169.089	90.2094	1.87441	0.09774
Populacao	0.00400382	0.0134714	0.29721	0.77388
nCrimes	0.558035	0.673863	0.828113	0.43162

Figura 35 Output do OLS - .Relatório obtido através do software GeoDa