

ESTUDO E ANÁLISE DA DISTRIBUIÇÃO ESPACIAL DOS INCÊNDIOS NO DISTRITO DE ÉVORA

Carlos Godinho

(fc48511@alunos.fc.ul.pt)

RESUMO

No âmbito da disciplina de Análise Espacial de Dados do Mestrado em Engenharia Geoespacial, foi proposto a execução de um trabalho prático que envolvesse os conhecimentos adquiridos nas aulas teóricas e práticas, de forma a explorar as diferentes técnicas de análise espacial em três domínios diferentes. Nomeadamente na análise da distribuição espacial de incêndios no distrito de Évora (padrão pontual), na análise de autocorrelação e regressão espacial dos incêndios e por fim compreender se a distribuição espacial dos focos de incêndio poderá ser ou não justificada com a densidade populacional, a taxa de analfabetismo e a taxa de desemprego. Pretende-se com a elaboração deste relatório, transmitir de uma forma sintética e simplificada o trabalho desenvolvido, explicando faseadamente a metodologia adotada e analisando resultados.

1. Introdução

A compreensão da distribuição espacial de dados desperta interrogações em diversas áreas do conhecimento. No que respeita à análise espacial, esta constitui uma ferramenta útil em eventos passados de modo a ser possível prever ocorrências futuras.

No presente trabalho, realizado no âmbito da unidade curricular de Análise Espacial de Dados, serão abordadas diferentes ferramentas que permitem analisar a forma como os eventos se relacionam, espacialmente ou entre si. Com este trabalho, pretende-se garantir a coerência das análises e dos resultados, sendo relevante sobretudo a exploração das ferramentas de análise espacial. Assim, pretende-se compreender a distribuição dos incêndios no distrito de Évora, ocorridos entre 2011 e 2015, tendo em conta a sua localização espacial e os fenómenos ocorridos no espaço. Para tal, o processo de análise espacial envolve diversas etapas que deverão respeitar uma sequência lógica de procedimentos.

Este projeto visa aplicar técnicas de diferentes tipos de análises possíveis. Deste modo, o trabalho divide-se em cinco etapas principais, onde na primeira etapa é realizada a análise da distribuição espacial dos focos de incêndio em Évora no período de 2011 a 2015 utilizando ferramentas de análise de distribuição pontual, densidade pontual e identificação de clusters.

Na segunda etapa, através de ferramentas de análise de áreas, autocorrelação espacial e testes de significância estatística, são analisadas as variáveis focos de incêndio, densidade populacional, taxa de analfabetismo e taxa de desemprego no distrito de Évora.

A terceira parte corresponde ao uso de técnicas de regressão linear e regressão espacial de modo a explicar eventuais associações não só entre as variáveis em estudo, mas também espaciais. Pretende-se definir um modelo de regressão explicativo para a ocorrência de focos de incêndio.

Na quarta etapa aplicaram-se diferentes métodos de interpolação e validaram-se através de um conjunto de nove vértices geodésicos da rede geodésica nacional com elevada precisão altimétrica.

Na última etapa efetuou-se uma análise de redes na freguesia de Reguengos de Monsaraz através da aplicação de três algoritmos, nomeadamente best route, closest facility e service area.

Neste estudo foram utilizadas ferramentas de autocorrelação espacial¹ e aplicaram-se testes de significância estatística aos indicadores de associação espacial observados. Tentou-se definir um modelo de regressão explicativo para a ocorrência de focos de incêndio, mas verificou-se que a correlação que esta variável apresentava não o permitia, relativamente às restantes variáveis.

¹ Índice Global de Moran e Local Indicators of Spatial Autocorrelation (LISA).

Ao longo do projeto foram utilizados operadores disponibilizados por alguns sistemas de informação geográfica². Todos os dados utilizados nestes softwares foram uniformemente transformados para o sistema de coordenadas PT-TM06/ETRS89 e foram adquiridos em diversas fontes de informação de acordo com a Tabela 1.

DADOS	FONTE
CAOP Continente: Carta Administrativa Oficial de Portugal Continental - 2015	http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop/caop_download/carta_administrativa_oficial_de_portugal__versao_2016/
Focos de Incêndio – Portugal Continental	http://www.icnf.pt/portal/florestas/dfci/inc/estat-sgif
Densidade populacional – Municípios	http://www.pordata.pt/Municipios/Densidade+populacional+segundo+os+Censos-591
Taxa de Analfabetismo – Municípios	http://www.pordata.pt/Municipios/Taxa+de+analfabetismo+segundo+os+Censos+total+e+por+sexo-721
Taxa de Desemprego - Municípios	https://www.pordata.pt/Municipios/Taxa+de+desemprego+segundo+os+Censos+total+e+por+sexo+(percentagem)-405

Tabela 1 - Dados e fontes

O distrito de Évora³ pertence à sub-região do Alentejo Central. Limita a norte com o distrito de Santarém e com o distrito de Portalegre, a leste com Espanha, a sul com o distrito de Beja e a oeste com o distrito de Setúbal. Nas figuras 1 e 2 está ilustrado o enquadramento do distrito e dos concelhos constituintes.

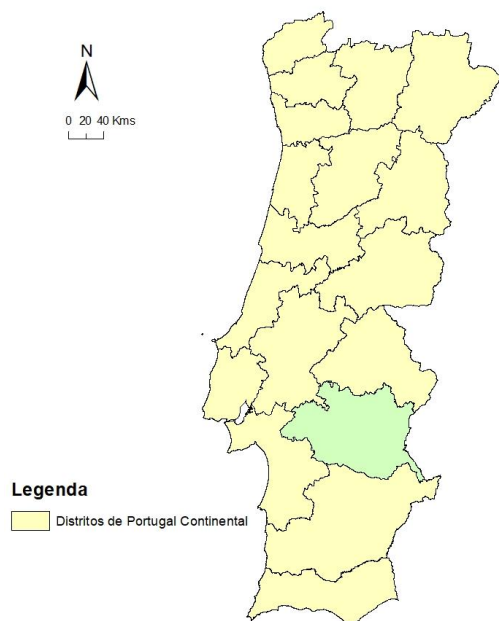


Figura 1 - Concelhos inseridos no distrito de Évora

Figura 2 - Enquadramento administrativo de Évora em Portugal Continental

² ARCGIS® 10.6.0, GeoDaTM 1.8.16.4 e QGIS 3.0.0.

³ Tem uma área de 7 393 km² sendo o 2.º maior distrito português. A sede do distrito é a cidade com o mesmo nome e é formado por catorze municípios.

2. Análise da Distribuição Espacial de Focos de incêndios

O objetivo na análise espacial de padrões de pontos é verificar se existe uma tendência nos eventos observados exibindo um padrão sistemático, como regularidade, agrupamentos ou aleatoriedade. Primeiramente faz-se uma visualização dos dados com o intuito de identificar e localizar padrões espaciais de interesse. De seguida, efetua-se uma análise exploratória dos dados para quantificar padrões com indicadores estatísticos e, por fim, proceder-se à modelação que permite detetar padrões espaciais de interesse.

Após descarregar os dados nacionais de incêndios em formato Excel, procedeu-se à aplicação de filtros para selecionar o distrito pretendido (Évora) assim como eliminar os falsos alarmes e as queimadas, considerando assim apenas os incêndios florestais e agrícolas para o estudo. Procedeu-se à união de todos os dados da serie temporal a estudar: ano 2011 a 2015. Verificou-se que o sistema de coordenadas dos dados era Datum Lisboa HayfordGauss IGeoE. No ArcGis importaram-se todos os dados, transformando-os para o sistema de coordenadas ETRS89 e exportaram-se para formato shapefiles.

Após importar os dados verificou-se a existência de alguns outliers que se teve de corrigir, conforme ilustra a figura 3.

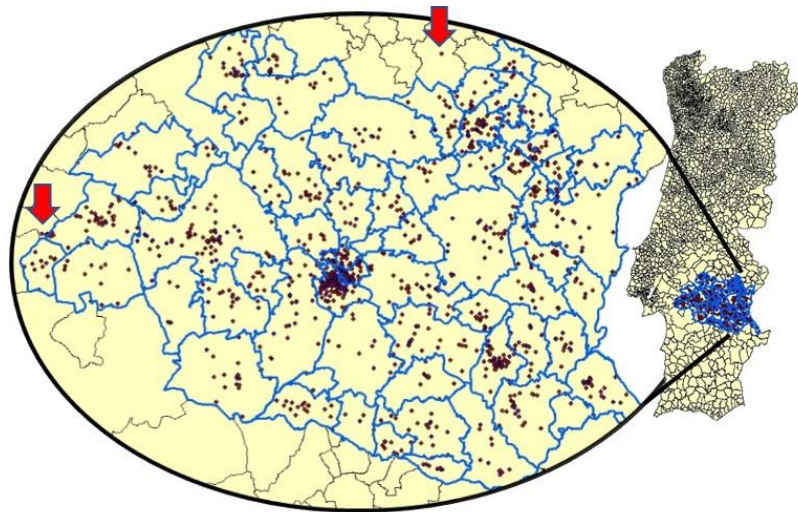


Figura 3 - Outliers de incêndios no distrito de Évora

2.1. Visualização do padrão pontual

Durante esta etapa procurou-se verificar se os eventos observados apresentavam algum tipo de padrão sistemático e quais os fatores que poderiam determinar esse padrão espacial.

Na análise de padrões pontuais, os eventos de interesse são identificados como pontos localizados no espaço, tendo como objeto de estudo a sua própria localização. No entanto, para além da localização geográfica dos eventos, é fundamental conhecerem-se outros parâmetros estatísticos básicos para a caracterização dos dados espaciais, tais como a frequência⁴ e a densidade⁵.

⁴ Obtiveram-se 1591 focos de incêndio no distrito de Évora na serie temporal em estudo.

⁵ Obteu-se 0,21 focos de incêndio km².



Figura 4 - Focos de incêndio (2011-2015) no distrito de Évora

Para melhor entender a distribuição espacial dos dados, determinou-se o centro geométrico e o centro médio dos incêndios, a elipse e o círculo padrão que delimitam o espaço no qual está concentrado cerca 68% das observações.

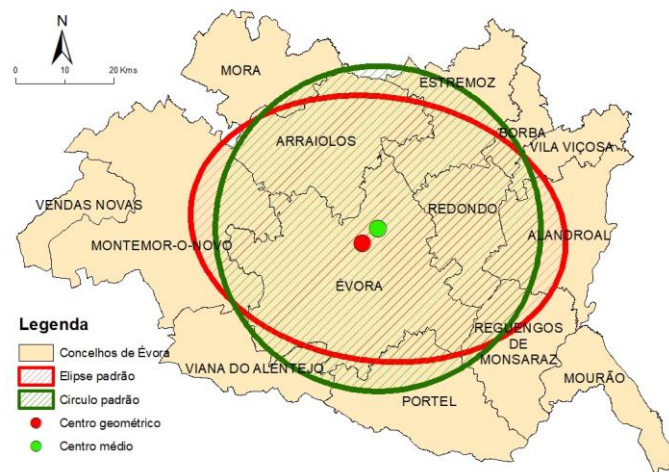


Figura 5 - Centro geométrico, centro médio, elipse e círculo padrão

É possível verificar na Figura 5 que o centro geométrico, que identifica a localização de maior concentração de pontos, localiza-se no concelho de Évora. Facto que não é estranho, visto ser um concelho com maior número de focos de incêndio. Quanto ao grau de dispersão dos focos, observa-se que a elipse padrão não é muito pronunciada, manifestando uma configuração ligeiramente direcionada ao longo do eixo Oeste-Este, evidenciando assim uma diferença entre o semi-eixo menor e maior.

Apesar destas operações permitirem uma análise visual imediata e conclusiva, é necessário estudar esta distribuição quanto ao seu padrão e identificar zonas de maiores aglomerados pontuais.

3. Análise Exploratória de Dados e Modelação do Processo Pontual

Nesta etapa efetuou-se uma análise rigorosa de eventuais padrões espaciais, tentando verificar associações espaciais entre as variáveis que caracterizam os fenómenos estudados. Os métodos desenvolvidos permitiram explorar as propriedades de 1ª e 2ª ordem do padrão pontual.

A análise estatística dos padrões de dados pontuais requer um modelo de referência para verificar a significância dos métodos exploratórios realizados anteriormente. Inicialmente, é necessário definir o modelo de referência e de seguida verificar se o padrão pontual se ajusta ao modelo. Esta verificação

pode ser feita comparando distribuições teóricas de frequência ou através de gráficos com envelopes de simulação. Normalmente, o modelo teórico a ser adotado é o da aleatoriedade espacial completa, pressupondo-se que existe igual probabilidade de ocorrência de um evento em qualquer parte da região de estudo (hipótese 1) e que existe independência na posição dos eventos (hipótese 2). Assim, para testar a hipótese de aleatoriedade espacial completa é necessário comparar os valores das distribuições de referência com os valores estatísticos obtidos pelos métodos executados.

3.1. Técnicas de análise Exploratória de 1ª Ordem (Globais)

Para explorar as propriedades de 1ª ordem e efetuar o reconhecimento e medição do padrão de uma distribuição de pontos utilizou-se o método da contagem por quadrantes e o estimador de densidade Kernel.

O método da contagem por quadrantes consiste numa grelha estruturada na área em estudo procedendo-se depois à simples contagem de eventos em cada um dos quadrantes. Esta técnica é a mais simples de aplicar, embora tenha mais limitações pois está dependente da dimensão da célula escolhida. Quadrantes grandes poderão produzir uma fraca descrição do padrão, mas quadrantes demasiado pequenos poderão não conter eventos.

Para efetuar o método de contagem por quadrantes foi necessário, em primeiro lugar, criar a grelha cujos limites correspondiam aos limites dos dados, sendo definido um espaçamento entre quadrantes da grelha, de 7500 metros, como sendo o que melhor descrevia o padrão pontual na área de estudo, obtendo-se desta forma uma média de 9 focos de incêndio por quadrante. A Figura 6 apresenta os resultados da contagem.

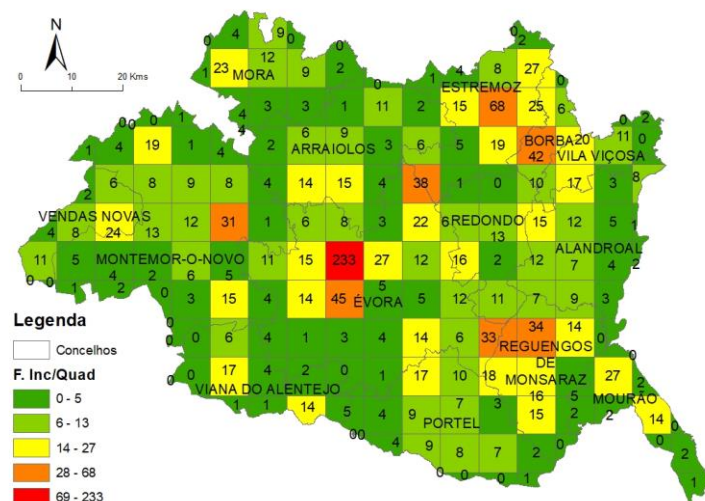


Figura 6 - Focos de incêndios por quadrante

De forma expetável, observa-se na Figura 6 que a maior concentração de pontos (ou seja, focos de incêndio) encontra-se no concelho de Évora e no concelho de Estremoz, existindo claramente a presença de um padrão fortemente aglomerado.

Neste método é calculada a média de eventos por quadrante, a variância associada e ainda a estatística VTMR⁶. Se o valor de VTMR for inferior a 1, o padrão é regularmente disperso, ou seja, os eventos estão distribuídos quase regularmente no espaço não contendo um padrão de aglomerados. No caso de VTMR ser superior a 1, o padrão é de aglomerados ou clusters, ou seja, existência de conjuntos de dados agrupados e áreas extensas sem pontos. Para o caso da VTMR ser igual a 1, o padrão apresenta uma distribuição aleatória, não tendendo nem para aglomerado nem para a dispersão. No caso estudado, o VTRM (42,3) é claramente superior a 1, revelando que apresenta um padrão de aglomerados ou clusters.

⁶ Variance To Mean Ratio é a medida da divisão da variância pela média, sendo assim uma medida normalizada em torno de 1.

Para podermos afirmar que estamos perante uma amostra deste tipo, com um certo nível de confiança é necessário efetuar o teste do qui-quadrado. Para este teste testamos duas hipóteses:

- H0: Padrão de pontos distribuído aleatoriamente
- H1: Padrão de pontos distribuído em aglomerados

A variável de teste é calculada multiplicando o número de graus de liberdade⁷ pelo VTRM.

Mean	9,22
Var	389,81
VTRM	42,30
m (nº de quadrantes)	172,00
qui ²	7233,45
alpha	0,02
P	0,98
qui ² left	130,94
qui ² right	216,94

Tabela 2 - Resultados do teste do qui-quadrado

Para um grau de confiança de 98% e nível de significância de 0.02 o valor do χ^2_{left} e do χ^2_{right} é de 130,94 e 216,94 respetivamente, e a variável de teste de 7233,45. Uma vez que o valor da variável de teste é muito superior ao valor do χ^2_{right} , rejeitamos a hipótese nula e podemos afirmar que estamos perante um padrão de pontos em aglomerados.

O estimador de densidade Kernel, consiste em converter o conjunto dos eventos pontuais numa superfície continua que represente a densidade. Esta função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse. Neste caso, a visualização do padrão de pontos permite a deteção de hotspots, ou seja, regiões de influência, verificando se o processo é ou não estacionário a partir das variações locais de intensidade. No Kernel a densidade é determinada em todos os locais da região de estudo e não somente nos locais onde existem eventos.

A melhor representação da densidade do padrão pontual foi obtida com uma largura de banda de 10000 metros. Denota-se que a densidade⁸ de focos de incêndio, é nitidamente alta nos concelhos de Évora e Estremoz e constata-se pequenos hotspots situados no centro dos concelhos de Reguengos de Monsaraz e Borba. Verifica-se que a densidade não é constante na área de estudo, graças à existência de variações locais de intensidade, sendo este processo consequentemente não estacionário de 1ª ordem.

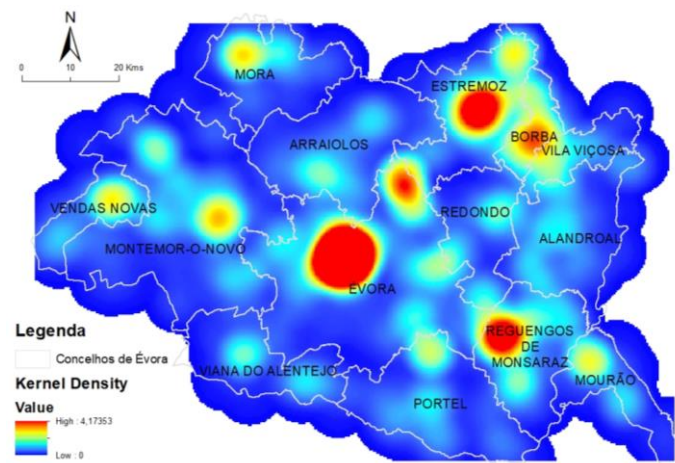


Figura 7 - Kernel Estimation Density dos focos de incêndios

De forma consistente com o método anterior, verifica-se que a zona de hotspots é essencialmente a noroeste e centro do distrito.

⁷ Número total de quadrantes menos 1.

⁸ Número eventos a dividir por unidade de área.

3.2. Técnicas de análise Exploratória de 2ª Ordem (Locais)

Para explorar as propriedades de 2ª ordem utilizaram-se os métodos de distância ao vizinho mais próximo função G e função K, através da medição de distâncias entre pontos. Estes métodos assumem estacionariedade sobre pequenas escalas na região de trabalho, significando que a média e a variância do processo são independentes da localização e constantes na região.

3.2.1. Função G

A função G assenta na medição de distância entre cada evento e o seu vizinho mais próximo, quantificando a forma como estão distribuídos espacialmente no distrito de Évora. Deste modo, se a média de todas as distâncias for inferior a valores de uma distribuição aleatória hipotética, estamos perante um aglomerado, se estiver acima estamos perante uma dispersão pontual. Era expectável de acordo com a análise já efetuada, que o padrão de distribuição pontual dos focos de incêndio fosse de aglomerado como mostra a Figura 8.

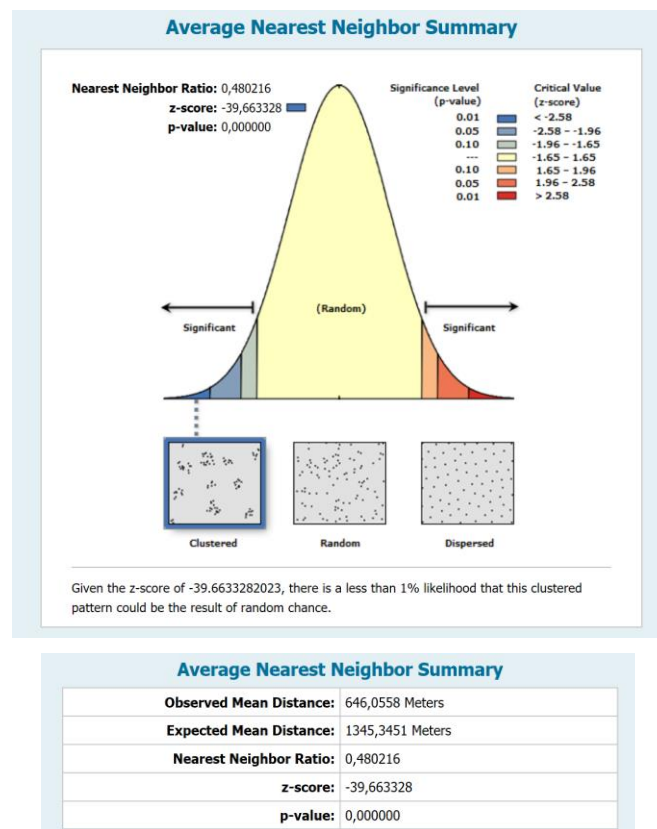


Figura 8 - Resultados da operação do vizinho mais próximo (Função G)

Nesta operação, a distância média esperada era de 1345 metros e a distância observada foi de 646 metros. O resultado da Função G indica que o fenómeno em estudo se trata de um padrão de aglomerado, com menos de 1% de probabilidade deste padrão ser disperso. O z-score é uma medida de desvios padrão, estando relacionado com o p-value que representa a probabilidade do padrão pontual observado ser resultado de um processo aleatório. A hipótese nula equivale a afirmar que o padrão pontual apresenta aleatoriedade espacial completa. Se o valor para o z-score se encontrasse entre os valores -1.65 e 1.65, significava a aceitação da hipótese nula com 90% de confiança. Visto que o valor z-score foi de aproximadamente -39 desvios padrão, rejeita-se a hipótese nula, afirmando-se com 99% de confiança que não se está perante a presença de aleatoriedade espacial completa, mas sim de aglomeração.

Este método apenas avalia a distância mínima entre dois eventos, podendo ser pouco vantajoso em padrões bastante aglomerados cujas distâncias entre eventos sejam muito pequenas relativamente a outras.

3.2.2. Função K

A função K permite uma análise mais precisa da distribuição de pontos, especialmente quando se lida com múltiplas distâncias e escalas espaciais. Para obter os melhores resultados da função K deve-se definir corretamente o número de distâncias a avaliar, a distância de partida e os incrementos. O objetivo é encontrar a distância a partir da qual deixa de existir um padrão de aglomerado entre os pontos e passa a existir um padrão disperso.

Neste caso foram feitas várias tentativas, verificando-se que os melhores resultados obtiveram-se usando 15 distâncias, começando com uma distância de 10 quilômetros em incrementos de 10 quilômetros também (além disso foram efetuadas 10 permutações que produziram envelopes de simulação, analisados num próximo capítulo). Na Figura 9 é possível observar o resultado da aplicação da função K.

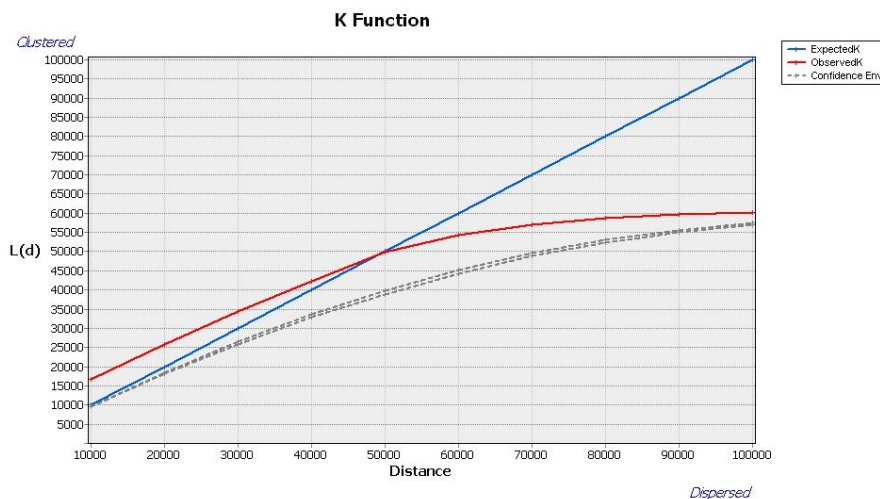


Figura 9 - Resultado da aplicação da Função K (distâncias em metros)

Nas distâncias em que a linha vermelha⁹ se encontra acima da linha azul¹⁰ significa que existe um padrão de aglomerado. É possível observar que apenas a partir de cerca de 50 quilômetros é que o padrão pontual tende para a dispersão. Isto significa que existe coerência entre os dois métodos de análise de eventos de segunda ordem, e que os focos de incêndio apresentam um padrão de aglomerados.

Na análise da condição de aleatoriedade espacial completa, é necessário obter valores para a função K com envelopes de simulação. Estes envelopes permitem concluir, para um dado nível de confiança, se existe aleatoriedade espacial completa ou não. O envelope de confiança distribui pontos de forma aleatória na área de estudo, denominado de permutação, calculando a função K para esta distribuição. O número de permutações determina o número de vezes que os pontos são distribuídos aleatoriamente pela área de estudo. Geralmente, as permutações podem ser associadas a níveis de confiança, como por exemplo, 9 permutações equivalem a 90% de confiança, 99 permutações equivalem a 99% de confiança e 999 permutações equivalem a 99.9% de confiança. Devido á morosidade deste processo, apenas testámos a função K com 9 permutações e concluiu-se que o padrão de distribuição pontual em estudo é de aglomerado com 90% de confiança.

4. Análise de Áreas e Autocorrelação Espacial

A segunda fase do trabalho consiste na análise dos dados por áreas para analisar a existência de correlação entre estas, e detetar possíveis outliers nos dados. Assim, é possível determinar se os processos espaciais na região de estudo são ou não estacionários de 1ª ou de 2ª ordem.

⁹ Valor observado.

¹⁰ Valor esperado

Primeiramente, será feita uma visualização dos dados espaciais, seguindo-se uma análise exploratória dos mesmos e verificação da significância estatística.

4.1. Visualização dos dados espaciais

Nesta análise o padrão pontual foi agregado em unidades territoriais correspondentes aos concelhos no distrito de Évora. As variáveis escolhidas são o número de focos de incêndio (2011-2015), a densidade populacional (2011), a taxa de analfabetismo (2011) e a taxa de desemprego (2011).

Considerando os 14 concelhos constituintes, classificaram-se os mapas das distribuições espaciais das variáveis em cinco classes de quantis.



Figura 10 - Distribuição espacial do número de focos de incêndios nos municípios do distrito de Évora

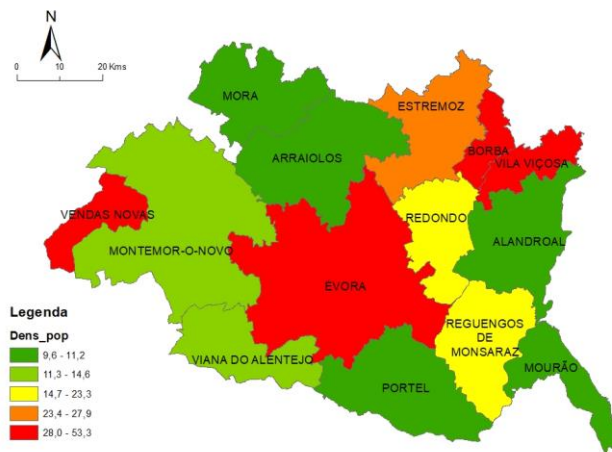


Figura 11 - Distribuição espacial da densidade populacional nos municípios do distrito de Évora - Habitantes / Superfície em Km²



Figura 12 - Distribuição espacial da taxa de analfabetismo nos municípios do distrito de Évora - (População residente com 10 e mais anos que não sabe ler nem escrever / População residente com 10 e mais anos) * 100

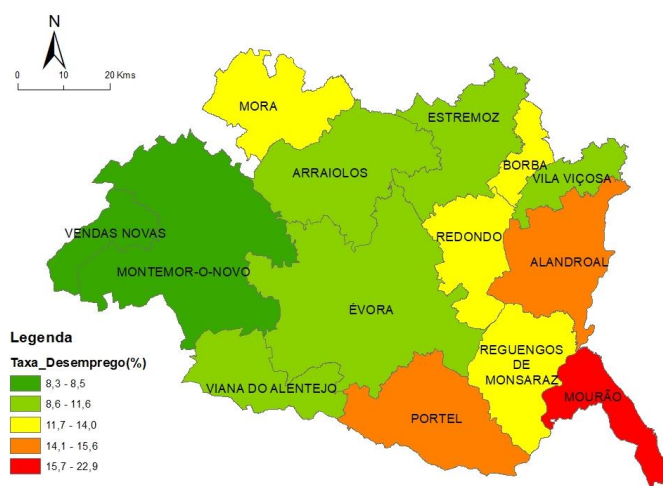


Figura 13 - Distribuição espacial da taxa de desemprego nos municípios do distrito de Évora - (População desempregada / População ativa) * 100

No mapa da densidade populacional os concelhos que apresentam os valores mais elevados são Borba e Vendas Novas. A variável taxa de analfabetismo apresenta uma distribuição espacial dispersa, verificando-se os valores mais elevados nos concelhos de Mora, Viana do Alentejo, Borba e Alandroal. Quanto à variável taxa de desemprego, verificamos uma distribuição espacial com os valores mais elevados na parte este do distrito, destacando-se o município de Mourão com o valor mais alto.

4.2. Análise exploratória dos dados

Nesta etapa, pretende-se medir a estrutura da correlação espacial e estimar a magnitude da autocorrelação espacial entre municípios. As estatísticas podem ser divididas em globais, interessando a forma como estão correlacionados os valores no espaço, ou locais, cujo o objetivo é identificar clusters e outliers, através de valores específicos para cada área.

4.2.1. Índices Globais de Autocorrelação Espacial

Para as estatísticas de autocorrelação espacial global, consideraram-se todos os incêndios em cada município, tendo por objetivo caracterizar a dependência espacial mostrando como os valores estão correlacionados no espaço e fornecendo uma medida geral da associação espacial.

O Índice Global de Moran mede a autocorrelação espacial, comparando o valor de uma variável com os valores de outras variáveis em quaisquer localizações. Este índice mede a autocorrelação espacial baseado simultaneamente nos valores das variáveis e nas suas localizações. O Índice de Moran considera a covariância entre pares e sendo um indicador global, determina o nível de interdependência espacial entre todos os polígonos em estudo.

Valores próximos de 1 indicam aglomerados e a autocorrelação espacial é positiva, o que significa que valores altos tendem a estar na vizinhança de valores altos e valores baixos tendem a estar na vizinhança de valores baixos. Se o coeficiente de correlação é nulo, significa aleatoriedade e se a autocorrelação espacial for negativa, o valor do índice indica que valores altos estão rodeados de valores baixos e valores baixos estão rodeados de valores altos.

O gráfico de Moran compara os valores normalizados do atributo numa área com a média dos valores normalizados dos seus vizinhos. Como se verifica na Figura 14 a orientação do gráfico é no primeiro e segundo quadrante, logo a autocorrelação é positiva, ou seja, localizações com vizinhos de valores semelhantes (clusters). Se a orientação fosse no terceiro e quarto quadrante (Q3 e Q4) então a autocorrelação seria negativa, isto é, localizações com vizinhos com valores distintos e padrão dispersivo.

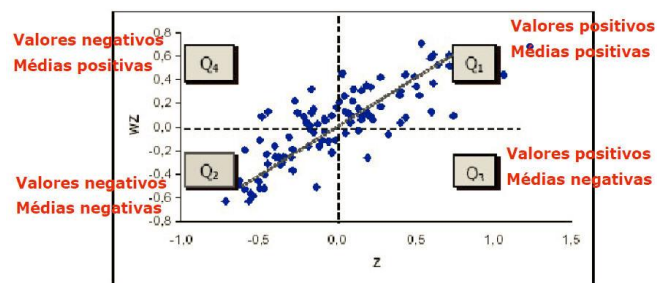


Figura 14 - Gráfico explicativo de Moran

De forma a quantificar o grau de autocorrelação espacial existente em cada uma das variáveis de estudo, foi calculado o índice global de Moran. O método de relação espacial entre concelhos foi a contiguidade de 1ª ordem para criação da matriz de proximidade espacial ou matriz de vizinhança.

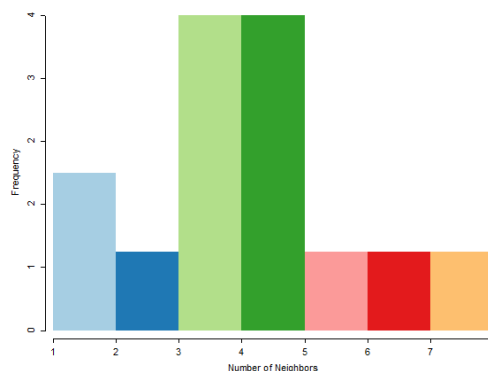


Figura 15 - Matriz de vizinhança

Tendo por base a Tabela 3, analisando os valores obtidos do índice global de Moran é possível verificar que a nível global as variáveis focos de incêndios e densidade populacional apresenta uma distribuição espacial aleatória (ausência de autocorrelação espacial) pelo que a variável é espacialmente independente. O índice global de Moran relativo ao poder de compra, à taxa de analfabetismo e à taxa de desemprego no sector primário indica autocorrelação espacial repartida entre a aleatoriedade e a autocorrelação espacial negativa (com valores próximos de -1), indicador de padrão dispersão, onde áreas vizinhas são menos similares do que as mais distantes.

Variáveis	Número de focos de incêndios	Densidade populacional	Taxa de analfabetismo	Taxa de desemprego
Índice de Moran Global	-0.068	-0.013	-0.208	0.336

Tabela 3 – Índice de Moran global das variáveis em estudo

De seguida são apresentados os diagramas de dispersão de Moran, para uma análise detalhada de cada uma destas variáveis de estudo.

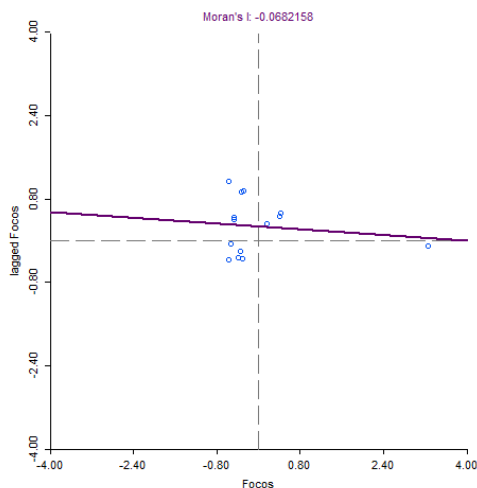


Figura 16 - Diagrama de dispersão de Moran para a variável número de focos de incêndios

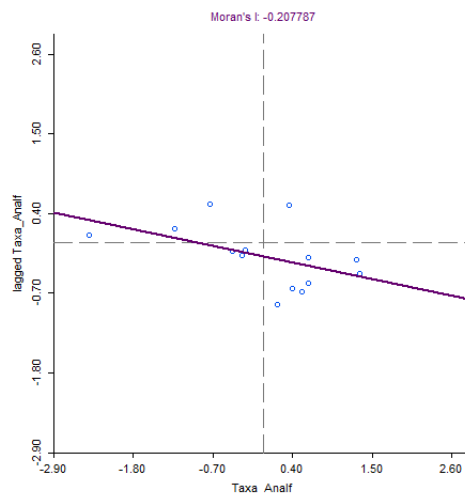


Figura 18 - Diagrama de dispersão de Moran para a variável taxa de analfabetismo

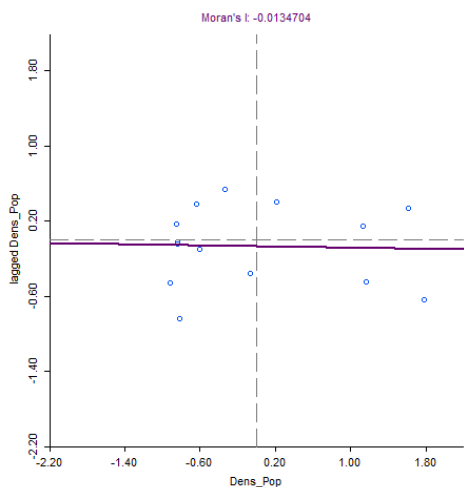


Figura 17 - Diagrama de dispersão de Moran para a variável densidade populacional

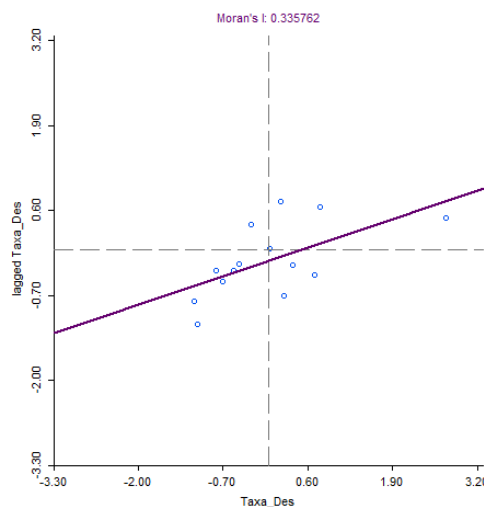


Figura 19 - Diagrama de dispersão de Moran para a variável taxa de desemprego

O índice global de Moran da variável focos de incêndios e densidade populacional, Figura 16 e 17 respetivamente é quase zero, revelando a já referida ausência de autocorrelação espacial (aleatoriedade espacial), observa-se pelo respetivo diagrama de dispersão que se trata de uma aleatoriedade absoluta, dada a nuvem de pontos se encontrar dispersa pelos quatro quadrantes do diagrama.

A variável de analfabetismo apresenta uma tendência para associação espacial negativa (quadrantes 3 e 4), sinal de que estes concelhos têm vizinhos com valores distintos.

No diagrama da Figura 19, observa-se a tendência global para uma associação espacial positiva (quadrantes 1 e 2) na variável taxa de desemprego, sinónimo de um padrão espacial com concelhos vizinhos mais similares que os mais distantes.

Depois de se obter o índice global de Moran é fundamental conhecer a sua validade estatística. Para tal procedeu-se à aplicação de testes de pseudosignificância estatística pela geração de 999 permutações, para cada variável em análise. A análise da significância estatística efetuou-se segundo os critérios da distribuição normal¹¹.

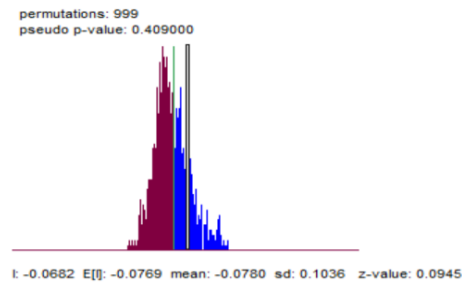


Figura 20 - Teste de Pseudo-Significância (999 permutações) número de focos de incêndios

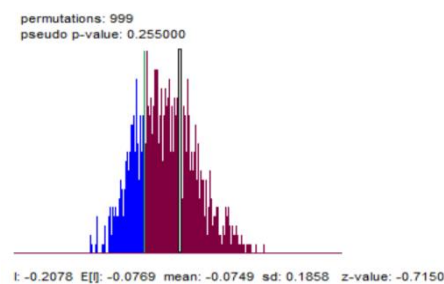


Figura 21 - Teste de Pseudo-Significância (999 permutações) Taxa de Analfabetismo

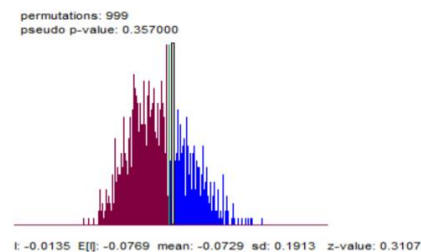


Figura 22 - Teste de Pseudo-Significância (999 permutações) Densidade Populacional

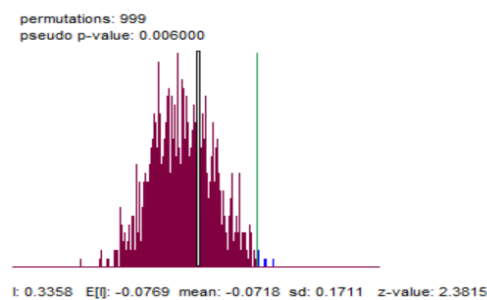


Figura 23 - Teste de Pseudo-Significância (999 permutações) Taxa de Desemprego

A variável taxa de desemprego é a única variável a apresentar distintamente um resultado com significância estatística. Esta significância, permite validar a conclusão de que, para o índice global de Moran, a taxa de desemprego apresenta uma tendência para autocorrelação espacial positiva (com 99% de certeza), ou seja um padrão aglomerado. As restantes variáveis, em consequência do reduzido desvio padrão nos valores z-score apresentam resultados correspondentes ao intervalo da distribuição normal sem significância estatística.

¹¹ Critérios do nível de significância, nível crítico e Z score.

4.2.2. Índices Locais de Autocorrelação Espacial

Para as estatísticas de autocorrelação espacial local, também são considerados todos os incêndios em cada município, e avaliam a covariância entre um determinado polígono e uma certa vizinhança definida em função de uma distância. O objetivo é produzir um valor específico para cada área, permitindo a identificação de aglomerados (clusters), de valores extremos (outliers) e a existência de vários regimes espaciais, que não são identificados pelos indicadores globais.

O Índice Local de Moran consiste num produto do resíduo no polígono de referência com a média local dos resíduos dos seus vizinhos adjacentes, em que os valores altos e positivos apontam para a presença de cluster e valores baixos indicam um regime espacial de desigualdade ou zonas de transição. Os resultados seguintes foram obtidos após 99 permutações.

O LISA cluster map indica quais os quadrantes do mapa de dispersão de Moran a que pertence cada município. Da análise do mapa, verifica-se que os municípios a vermelho e a azul são os que contribuem significativamente para a autocorrelação positiva (Q1 e Q2) pois são municípios cujos vizinhos têm valores similares (altos próximos de altos e baixos próximos de baixos).

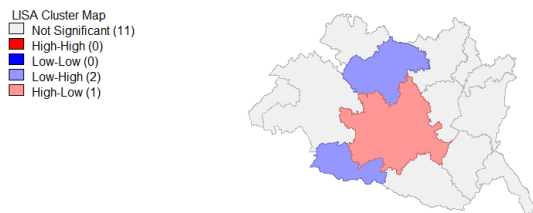


Figura 24 - LISA Cluster Map da variável número de focos de incêndio

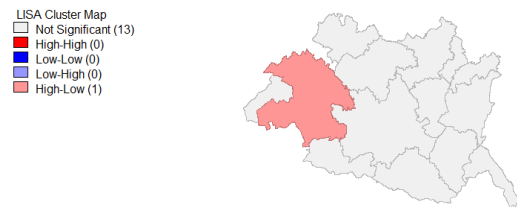


Figura 26 - LISA Cluster Map da variável taxa de analfabetismo

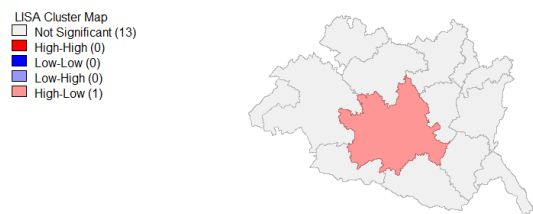


Figura 25 - LISA Cluster Map da variável densidade populacional

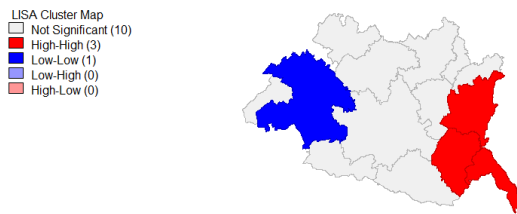


Figura 27 - LISA Cluster Map da variável taxa de desemprego

O LISA significance map mostra o nível de significância estatística com que cada município pode ser classificado, no sentido de contribuir para o valor global de autocorrelação. O método usado para as simulações é o método de Monte Carlo. Da análise do mapa, verifica-se que os municípios a verde-escuro são os que contribuem mais significativamente para a autocorrelação, diminuindo essa contribuição com a diminuição da cor e não tendo qualquer significado estatístico os municípios a branco.

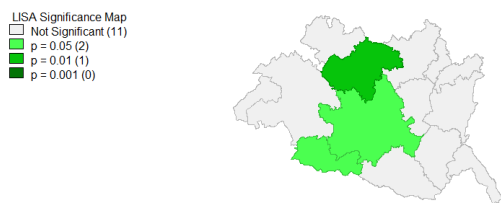


Figura 28 - LISA Significance Map da variável número de focos de incêndio

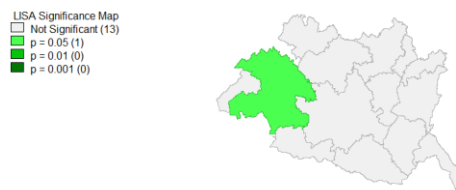


Figura 30 - LISA Significance Map da variável taxa de analfabetismo

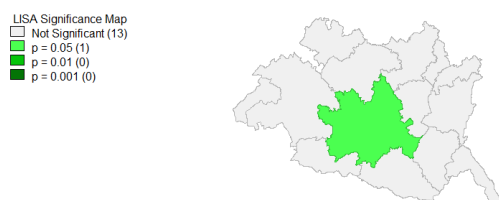


Figura 29 - LISA Significance Map da variável densidade populacional

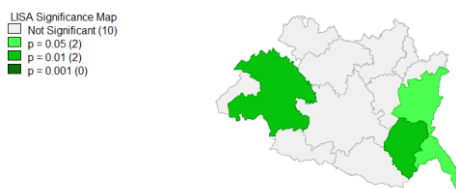


Figura 31 - LISA Significance Map da variável taxa de desemprego

5. Regressão Espacial

Nesta etapa, pretende-se usar técnicas de regressão espacial de modo a explicar eventuais associações espaciais. Enquanto a correlação mede o grau de relacionamento entre duas variáveis ou mais, a regressão fornece uma equação que descreve o relacionamento matemático entre elas.

5.1. Regressão Clássica (OLS)

Um modelo de regressão é uma ferramenta estatística que utiliza o relacionamento existente entre duas ou mais variáveis de maneira que uma delas possa ser descrita, ou o seu valor estimado, em relação às outras variáveis.

O primeiro passo da regressão espacial é analisar os resíduos do modelo. Para verificar a existência de correlação espacial deve-se mapear os resíduos e medir a autocorrelação espacial dos resíduos da regressão. No caso de existir dependência espacial os resíduos não são independentes e correlacionam-se.

Através do diagrama de dispersão é possível concluir sobre a existência de relação entre as variáveis, se os pontos se situam sobre uma linha reta existe uma relação estatística do tipo linear que é tanto mais forte quanto mais próximos estiverem os pontos da reta.

Os modelos de regressão espacial permitem estabelecer as relações entre as variáveis, levando em conta os efeitos espaciais. Para avaliar uma regressão é utilizado o valor do coeficiente de determinação R^2 , sendo este valor uma medida da proporção da variabilidade numa variável que é explicada pela variabilidade de outra. Este coeficiente varia de 0 a 1, sendo 0 nenhuma correlação entre as variações das variáveis em questão e 1 a correlação perfeita. Seguidamente apresenta-se os resultados da regressão clássica e o modelo.

```

REGRESSION
-----
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : Join_final
Dependent Variable : Focos  Number of Observations: 14
Mean dependent var : 113.643 Number of Variables : 4
S.D. dependent var : 112.774 Degrees of Freedom : 10

R-squared      : 0.411105  F-statistic      : 2.32699
Adjusted R-squared : 0.234437  Prob(F-statistic) : 0.136398
Sum squared residual: 104855  Log likelihood   : -82.3141
Sigma-square    : 10485.5  Akaike info criterion : 172.628
S.E. of regression : 102.399  Schwarz criterion : 175.184
Sigma-square ML  : 7489.62
S.E of regression ML: 86.5426
-----
Variable      Coefficient      Std.Error      t-Statistic      Probability
-----
CONSTANT      546.552          219.711        2.48759          0.03212
Taxa_Analf    -36.5535         15.9902        -2.28599         0.04532
Dens_Pop      -1.04318         2.20722        -0.47262         0.64663
Taxa_Des      -0.485858        8.30083        -0.0585313      0.95448
-----

```

Figura 32 - Relatório de Regressão Clássica

MODELO: $Y = 546.552 + (-36.553) X1 + (-1.043) X2 + (-0.486) X3 + \epsilon$

- Em que:
- X1 – Taxa de analfabetismo
 - X2 – Densidade populacional
 - X3 – Taxa de desemprego

Assim, fazendo uma análise dos resultados obtidos verifica-se que não existe grande correlação entre variáveis ($R^2=0.41$). Também o valor do Log Likelihood deveria ser o mais alto possível, assim como o valor de Akaike info criterion e o Schwarz criterion preferencialmente deveriam ser baixos.

No caso de haver uma correlação espacial forte entre as variáveis, o passo seguinte seria averiguar se a escolha do modelo foi a mais adequada. Nessa análise do modelo, seria fundamental mapear e fazer um diagnóstico dos resíduos (ϵ) no sentido de verificar se estes são independentes e não correlacionados com a variável dependente e, por outro lado, se apresentam uma distribuição normal com média zero e variância constante. Se estivermos na presença destas duas características, o modelo da regressão clássica é suficiente e aceitável.

Se verificarmos a existência de autocorrelação espacial, através de um indicador local de autocorrelação, por exemplo o Índice Local de Moran visto na secção anterior, devemos testar os modelos de regressão espacial.

Por forma a aplicar modelos de regressão espacial, assume-se nas secções seguintes que existe correlação entre as variáveis e que os resíduos apresentam autocorrelação espacial.

5.2. Modelos de Regressão Espacial (Spatial Lag e Spatial Error)

Para analisar a hipótese de que o processo espacial presente é estacionário, ou seja, não existe tendência e de que pode ser capturado por apenas um parâmetro aplicaram-se modelos com efeitos espaciais globais, nomeadamente Spatial Lag Model e Spatial Error Model.


```

REGRESSION
-----
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : Join_final
Spatial Weight : join_final_w
Dependent Variable : Focos      Number of Observations: 14
Mean dependent var : 113.643   Number of Variables : 5
S.D. dependent var : 112.774   Degrees of Freedom : 9
Lag coeff. (Rho) : -0.336839

R-squared      : 0.462048   Log likelihood : -81.8936
Sq. Correlation : -          Akaike info criterion : 173.787
Sigma-square   : 6841.72   Schwarz criterion : 176.982
S.E of regression : 82.7147

-----
Variable      Coefficient   Std.Error   z-value   Probability
-----
W_Focos      -0.336839    0.312779   -1.07692  0.28151
CONSTANT     666.744     181.658    3.67032  0.00024
Dens_Pop     -1.80363    1.79513    -1.00474  0.31502
Taxa_Des     -3.29378    7.00091    -0.470479 0.63801
Taxa_Analf   -38.1216    13.2944    -2.8675   0.00414
-----

```

Figura 33 - Spatial Lag Model

Neste modelo Spatial Lag podemos verificar a presença de um coeficiente Rho ($\rho = -0.34$) que reflete a não dependência linear entre as variáveis.

```

REGRESSION
-----
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : Join_final
Spatial Weight : join_final_w
Dependent Variable : Focos      Number of Observations: 14
Mean dependent var : 113.642857   Number of Variables : 4
S.D. dependent var : 112.774495   Degrees of Freedom : 10
Lag coeff. (Lambda) : -0.979842

R-squared      : 0.708911   R-squared (BUSE) : -
Sq. Correlation : -          Log likelihood : -79.422126
Sigma-square   : 3702.09   Akaike info criterion : 166.844
S.E of regression : 60.8448   Schwarz criterion : 169.4

-----
Variable      Coefficient   Std.Error   z-value   Probability
-----
CONSTANT     828.84      117.198    7.07211  0.00000
Dens_Pop     -2.26016    1.09685    -2.06059 0.03934
Taxa_Des     1.97853     3.96381    0.499147 0.61768
Taxa_Analf   -62.2362    11.1021    -5.60581 0.00000
LAMBDA      -0.979842   0.20677    -4.73879 0.00000
-----

```

Figura 34 - Spatial Error Model

Neste modelo Spatial Error podemos verificar a presença de um coeficiente Lambda ($\lambda = -0.98$) que também reflete a não dependência espacial inerente aos dados da amostra.

5.3. Comparação de Modelos de Regressão Espacial

MODELO	R ²	Log Likelihood	Akaike Criterion	Schwarz Criterion
OLS	0.41	-82.31	172.63	175.18
LAG	0.46	-81.89	173.787	176.982
ERROR	0.71	-79.42	166.84	169.4

Tabela 4 - Comparação de modelos

Da análise da tabela, pode-se verificar que o modelo que poderá ser melhor é o Spatial Error Model, pois é o que tem um coeficiente de determinação maior (R²) assim como o Log Likelihood, enquanto os valores de critério Akaike e Schwarz são menores. Apesar deste modelo ser o melhor de todos, é claro que nenhum deles indica a existência de dependência da ocorrência de incêndios face às variáveis independentes testadas.

6. Métodos de Interpolação

Uma interpolação de dados consiste em estimar valores de uma amostra tendo por base valores conhecidos na mesma amostra. Os métodos utilizados podem ser globais ou locais, dependendo do objetivo pretendido.

Um método global capta a tendência geral de uma superfície enquanto que um método local capta apenas a tendência de uma certa zona da região total. Utilizando os métodos de interpolação disponibilizados pelo ArcGis, foram utilizados quatro métodos de interpolação com o objetivo de determinar qual o melhor modelo que se ajusta ao modelo digital do terreno.

Para a realização desta etapa foi utilizado o concelho de Reguengos de Monsaraz¹². Para proceder à interpolação no software ArcGIS o modelo de elevação SRTM (em formato raster, com um tamanho de célula de 80x80 metros) foi convertido para pontos. Através dos pontos obtidos efetuou-se então a interpolação. O objetivo foi a refinação do modelo de elevação para uma malha mais fina de 15x15 metros usando o método de interpolação com resultados mais aceitáveis.

Os métodos utilizados foram Kriging¹³, IDW¹⁴, Natural Neighbor¹⁵ e Spline¹⁶ e os resultados estão representados nas Figuras 35, 36, 37 e 38 respetivamente.

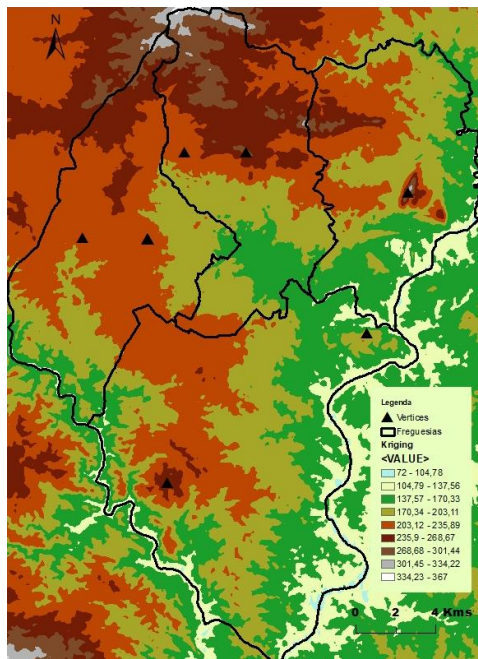


Figura 35 - Método de Interpolação kriging

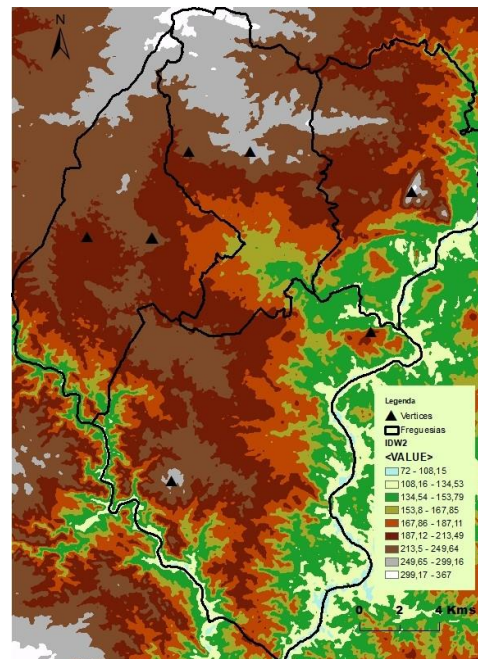


Figura 36 - Método de Interpolação IDW (Power:2)

¹² Dados extraídos do site <http://www.fc.up.pt/pessoas/jagoncal/srtm/>. Mosaico de quadrículas SRTM de 1 grau com malha de 3 segundos (versão 2). Preenchimento de pixels vazios por interpolação a partir dos pixels vizinhos. Conversão para coordenadas ETRS89-PTTM06 com pixel de 80 metros. Reamostragem bilinear. Formato Geotiff.

¹³ Método local probabilístico e exato.

¹⁴ Método local determinístico e exato.

¹⁵ Método local determinístico e exato.

¹⁶ Método local determinístico e exato.

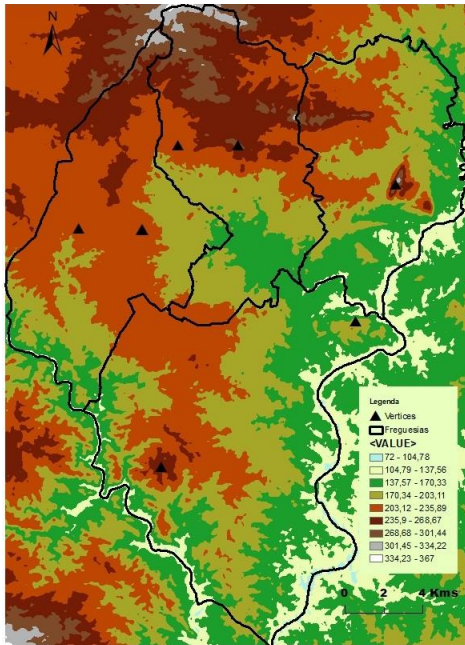


Figura 37 - Método de Interpolação Natural Neighbor

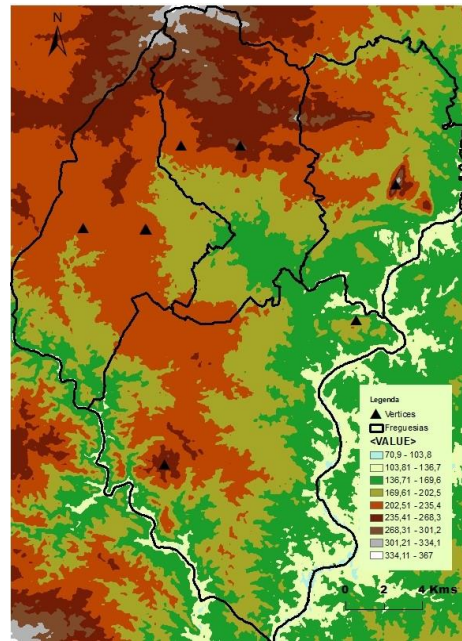


Figura 38 - Método de Interpolação Spline

A validação dos métodos de interpolação foi feita com base em vértices geodésicos. Comparando os valores dos vértices com alta precisão com os valores obtidos pelos métodos de interpolação podemos verificar, como apresentado na Tabela 5, que o interpolador mais aceitável é o que possui um erro menor do que 15m¹⁷. Denota-se que o ponto 76 foi considerado outlier e excluído desta análise por apresentar valores com elevadas discrepâncias relativamente aos restantes pontos.

PONTO	VERTICE	Z	SPLINE	(M2-Z)^2	NNEIGHBOR	(M2-Z)^2	KRIGING	(M2-Z)^2	IDW2	(M2-Z)^2
1242	VELEZ	297,56	283,10	208,98	283,22	205,58	283,02	211,44	283,36	201,70
1153	ZAMBUJOSA	251,71	234,28	303,87	234,00	313,64	233,29	339,19	233,57	328,95
1235	REGUENGOS	244,59	234,91	93,64	234,95	92,93	234,73	97,14	234,83	95,32
1236	REGUENGOS	245,37	235,00	107,54	235,00	107,54	235,00	107,54	235,00	107,54
1237	REGUENGOS	245,36	235,14	104,51	234,78	111,98	234,12	126,40	234,43	119,40
1244	ZAMBUJAL	227,86	215,60	150,43	215,22	159,72	214,33	183,09	214,75	171,90
729	PIPAS 1	225,94	211,00	223,20	211,00	223,20	211,00	223,20	211,00	223,20
835	SERRA	284,51	266,05	340,70	265,93	345,10	265,20	372,72	265,54	360,01
76	MONSARAZ	344,64	296,62	2306,30	296,52	2315,92	295,66	2399,04	296,58	2309,38
	Soma			1532,87		1559,70		1660,72		1608,01
	n-1		8,00							
	EQM			13,84229542		13,9628754		14,40799214		14,17751071

Tabela 5 - Validação do modelo usando 9 vértices geodésicos

Perante os resultados expostos na tabela, podemos ver que o método que possui menor erro é o Spline, ou seja, este será o método de interpolação aceite, comparando os quatro utilizados, para a freguesia de Reguengos de Monsaraz.

¹⁷ É a precisão dos dados utilizados inicialmente.

7. Análise de Redes

O objetivo nesta etapa seria, após a compreensão da distribuição espacial dos focos de incêndio no distrito de Évora, encontrar percursos ótimos entre os locais das corporações de bombeiros desse mesmo distrito e as áreas, anteriormente identificadas, com maior risco de incêndio, assim como gerar mapas de acessibilidades, em termos de distância, a partir da localização da cada corporação com o objetivo de fornecer elementos para um melhor planeamento ao combate de incêndios florestais. Devido ao processo de correção dos erros topológicos na área do distrito em estudo ser bastante moroso, optou-se por fazer uma análise de redes na freguesia de Reguengos de Monsaraz e considerou-se apenas as estradas principais.

7.1. Tratamento de dados

Em termos de dados foi necessário acrescentar a rede viária do município de Reguengos de Monsaraz através do OpenStreetMap e a partir desta criar uma nova Network Dataset em que se trata de uma coleção de elementos de rede topologicamente conectados (arestas, junções e curvas) que normalmente são utilizados para representar redes de estradas. Cada elemento de rede está associado com um conjunto de atributos de rede e os conjuntos de dados de rede são normalmente utilizados para modelar sistemas de fluxo sem direção.

7.2. Closest facility

Nesta fase para se determinar os percursos ótimos entre o local da corporação de bombeiros e as áreas com maior risco de incêndio, foi necessário recorrer ao algoritmo de análise de redes: Closest Facility. Com este algoritmo, determinamos as instalações (facilities) e os eventos (incidents) e os melhores percursos a percorrer face ao parâmetro distância. Assim, definiu-se o local de corporação de bombeiros como sendo a única facility e os locais com maior probabilidade de incêndios como incidents, resultando na Figura 37.

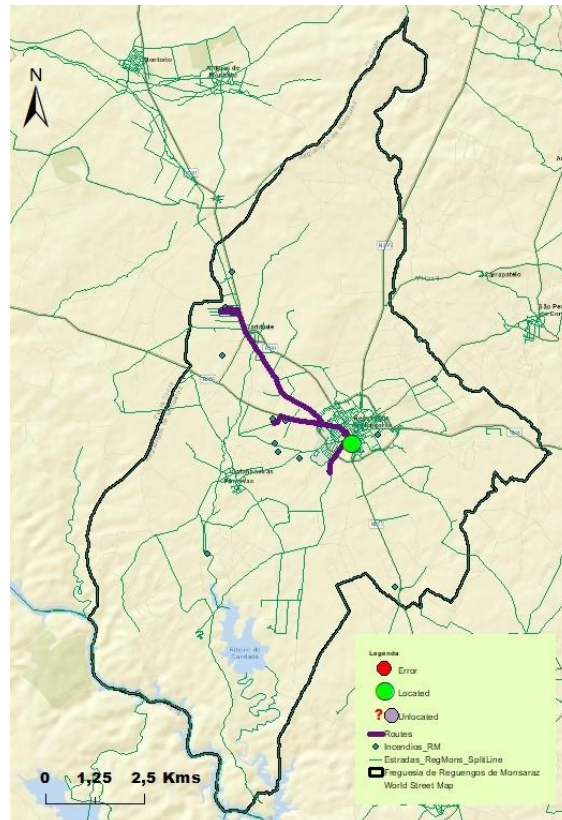


Figura 37 – Percursos ótimos entre a corporação de bombeiros e algumas zonas de elevada probabilidade de incêndio

7.3. Best Route

Nesta fase para se determinar os mapas de acessibilidade a partir do local da corporação de bombeiros para as áreas com maior risco de incêndio, foi necessário recorrer ao algoritmo de análise de redes: Route. Para um melhor resultado seria necessário aplicar este algoritmo a todas as corporações de bombeiros do distrito de Évora, considerando as localizações mais prováveis de incêndio situadas nas suas proximidades. Na figura 38, temos o exemplo do mapa de acessibilidade da corporação de bombeiros de Reguengos de Monsaraz face a alguns locais de incêndios mais próximos, em que está estabelecido o melhor percurso e a distância entre estes locais de modo a permitir um melhor planeamento por parte desta corporação face ao combate a incêndios florestais.

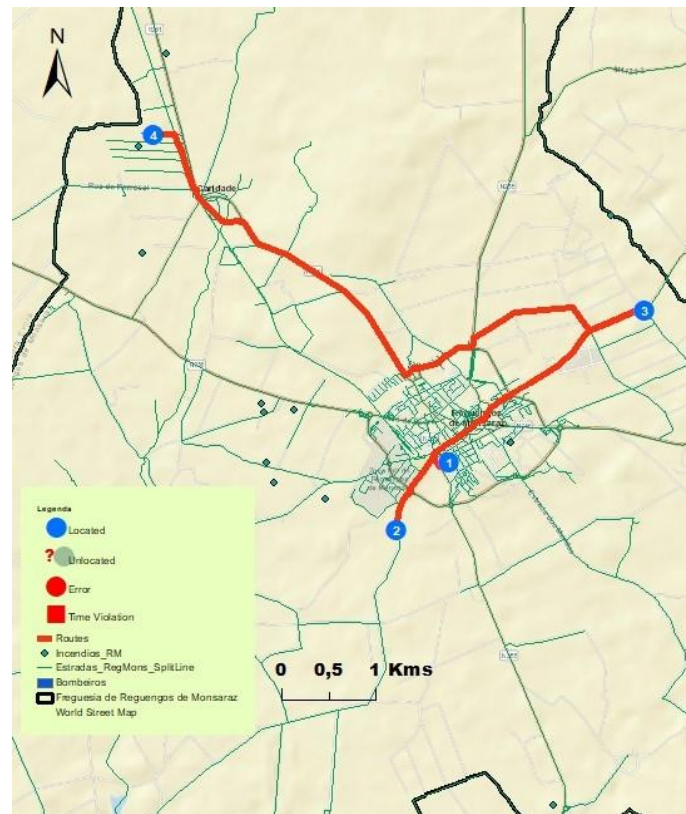


Figura 38 – Mapa de Acessibilidade dos Bombeiros de Reguengos de Monsaraz

7.4. Service Area

Nesta fase para se determinar as áreas de influência da corporação de bombeiros, de modo a avaliar e otimizar o combate a incêndios florestais, foi necessário recorrer ao algoritmo de análise de redes: Service Area. Com este algoritmo, determinou-se a área de influência da corporação de bombeiros, baseando-se em cinco distâncias: 1000, metros, 2000 metros, 3000 metros, 4000 metros e 5000 metros. Assim, definiram-se buffers em torno da corporação de bombeiros, tendo em consideração a rede viária existente, como se pode verificar pela Figura 39.

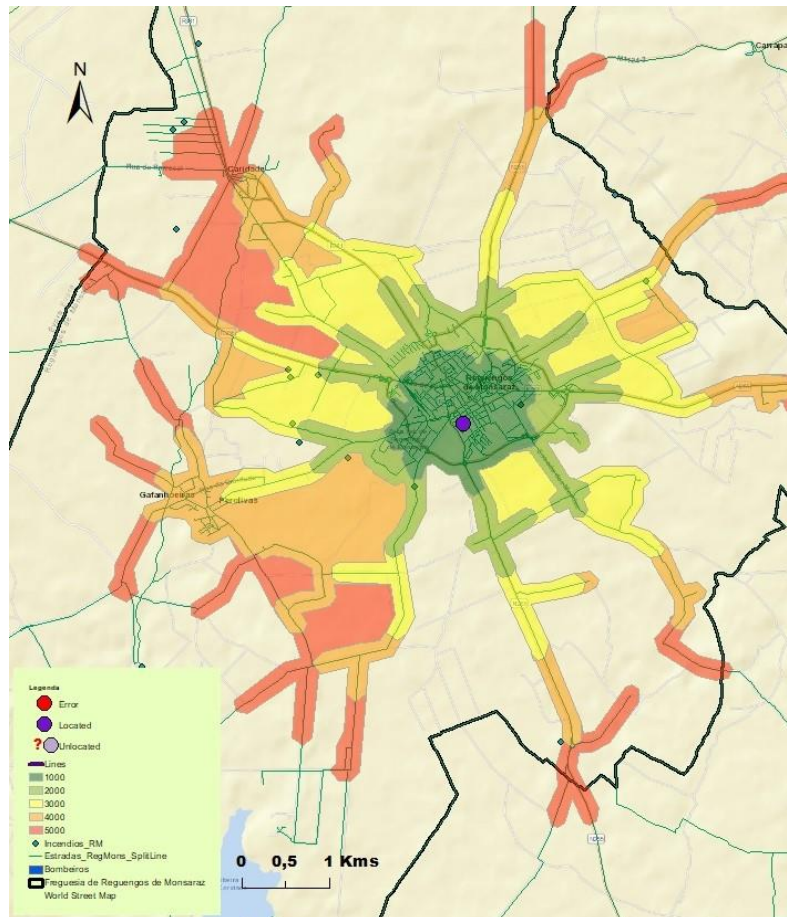


Figura 39 – Buffer 1000m, 2000m, 3000m, 4000m e 5000m em torno da corporação de bombeiros

8. Conclusões

A forma de execução deste projeto foi fundamental para ajudar a uma melhor compreensão e destreza dos conceitos ministrados durante as aulas teóricas e práticas. O facto deste trabalho ter sido faseado ao longo do semestre, de forma a acompanhar as aulas ministradas, levou a que tivessem sido tomadas opções numa perspetiva individual e não num conceito de projeto global. Refiro-me ao facto de se ter optado pela aplicação de métodos de interpolação a nível de um município e a análise de redes a nível de freguesia. Outra situação que poderia ter sido alterada foram as variáveis escolhidas e o período temporal. Para a realização de um trabalho mais completo, seria necessária uma abordagem diferente, mais rigorosa e exigente, especialmente nas escolhas já referidas.

Uma das grandes dificuldades sentidas durante a execução deste trabalho foi por vezes a interpretação dos resultados obtidos, revelando-se essencial o domínio das ferramentas estatísticas.

Referências Bibliográficas

Anselin, L. (2005), Exploring spatial data with GeoDaTM: a workbook, Texto não publicado, disponível em <http://geodacenter.asu.edu/system/files/geodaworkbook.pdf>;

Catita, C. (2018), Apontamentos da disciplina de Análise Espacial da Informação Geográfica. Texto não publicado, Universidade de Lisboa, Lisboa;