

Relatório da primeira parte do trabalho prático de Análise Espacial de Dados

João Gil

(fc42059@alunos.fc.ul.pt)

Resumo

O presente trabalho foi desenvolvido com base nas ferramentas analisadas nas aulas de Análise Espacial de dados e incidiu sobre os focos de incêndio e queimadas registadas no distrito da Guarda pela ICNF num período de 2001 e 2008. Foi feita uma análise de padrão pontual, uma análise de áreas e por fim uma regressão espacial. A análise de padrão pontual foi efetuada para a totalidade dos pontos e também para subconjuntos dos vários anos. Nesta análise foi concluído que tanto globalmente, como para cada ano os padrões pontuais são sempre agregados. Foram definidas as variáveis da população masculina, população desempregada e população analfabeta para o estudo de áreas. Concluiu-se que apenas a variável de número de indivíduos desempregados apresenta autocorrelação espacial. Concluiu-se também a existência de um *cluster* de taxas de desemprego altas na zona sul do concelho, com autocorrelação positiva local. Por fim foram testados vários modelos de regressão com e sem efeitos espaciais. Foi concluído que o modelo *Spatial Lag Model* seria o que mais se ajustava, apesar do nível de ajustamento não ser o mais adequado. Foi determinado que o interpolador por *kriging* seria o mais adequado para interpolar dados meteorológicos importantes para a determinação de risco de incêndio. Foram ainda determinados os poços mais adequados para o abastecimento dos carros de combate a incêndios das cooperações do concelho da Guarda.

1. Introdução

Este projeto foi desenvolvido no âmbito da unidade curricular de Análise Espacial de Dados. Os objetivos deste projeto prendiam-se com a análise da distribuição espacial dos pontos de ignição de incêndios florestais do distrito da Guarda compreendidos entre os anos de 2001 e 2008 e a investigação de possíveis variáveis que influenciem esta distribuição.

Este projeto é efetuado como o seguimento de outro projeto realizado numa unidade curricular anterior, em que foi elaborado um mapa de vulnerabilidade a incêndios da Guarda. Utilizando apenas os pontos de ignições de 2001 a 2008 e variáveis sociais recolhidas dos censos de população foi possível fazer vários testes para analisar a sua distribuição e possíveis correlações.

A área em estudo neste projeto foi o distrito da Guarda, na zona centro do país.

Segundo a Polícia Judiciária o incendiário português é essencialmente um indivíduo do sexo masculino, com nenhum ou fraco nível de escolaridade e desempregado ou com uma profissão de subsistência. Com estes dados foram recolhidos do portal PORDATA dados relativos a este perfil:

- Número de indivíduos do sexo Masculino por concelho
- Número de indivíduos analfabetos por concelho
- Número de indivíduos desempregados por concelho

Os dados recolhidos são referentes aos censos de 2001 por serem os mais próximos das datas de incêndios trabalhados.

Foram recolhidos dados meteorológicos para se determinar o melhor interpolador para este tipo de variáveis. Os dados meteorológicos foram considerados uma vez que estes são da maior importância na determinação do risco de incêndio.

Por fim foram utilizados algoritmos de análise de redes para determinar, na zona em estudo, os melhores percursos de cooperações de bombeiros, para poços onde os carros se possam abastecer de água de modo a melhorar a resposta a incêndios.

2.Dados utilizados

❖ Registos de Incêndios

Correspondem aos registos de incêndios do Instituto de Conservação da Natureza e Florestas. Destes registos foram retirados os correspondentes ao distrito da Guarda no período de 2001 a 2008. Os registos estão organizados em ocorrências de falsos alarmes, incêndios urbanos, queimadas e incêndios florestais. Para este trabalho apenas foram retirados os dois últimos. O sistema de coordenadas era o *Datum 73 Hayford Gauss IGeE*. Para este trabalho os dados foram reprojatados no sistema ETRS89-tm06.

❖ *Shapefile* do distrito

A *shapefile* do distrito da Guarda foi obtida da Carta Administrativa Oficial de Portugal elaborada pela DGT. Esta encontra-se no sistema de coordenadas ETRS89-tm06. Corresponde á versão de 2017 (em vigor).

❖ Dados de Censos

Foram recolhidos os dados de População Total, População do Sexo Masculino, População Desempregada e População Analfabeta correspondentes aos censos de 2001 de cada concelho integrante do distrito da Guarda.

De referir que os registos de incêndios e a *shapefile* do distrito da Guarda foram aproveitados de outro projeto pelo que o seu tratamento já havia sido assegurado.

3. Padrão Pontual

Um padrão pontual é um conjunto de dados correspondentes a localizações pontuais de eventos dentro de uma área em estudo. Neste tipo de dados existe interesse em estudar a localização geográfica de cada evento e o seu atributo associado. Pode ser também interessante analisar a frequência dos eventos na área em estudo. O estudo de um padrão pontual pretende então identificar os fatores que o controlam espacialmente, ou seja, verificar se os eventos apresentam algum padrão ou se pelo contrário apresentam uma distribuição aleatória.

A análise de um conjunto de dados espaciais passa por três passos:

1. Visualização dos dados espaciais
2. Análise Exploratória
3. Modelação

Os dados correspondentes aos registos de incêndios são eventos pontuais inseridos na área em estudo, o distrito da Guarda. Na figura 1 são apresentados os registos de incêndios já tratados correspondentes aos anos de 2001 a 2008.

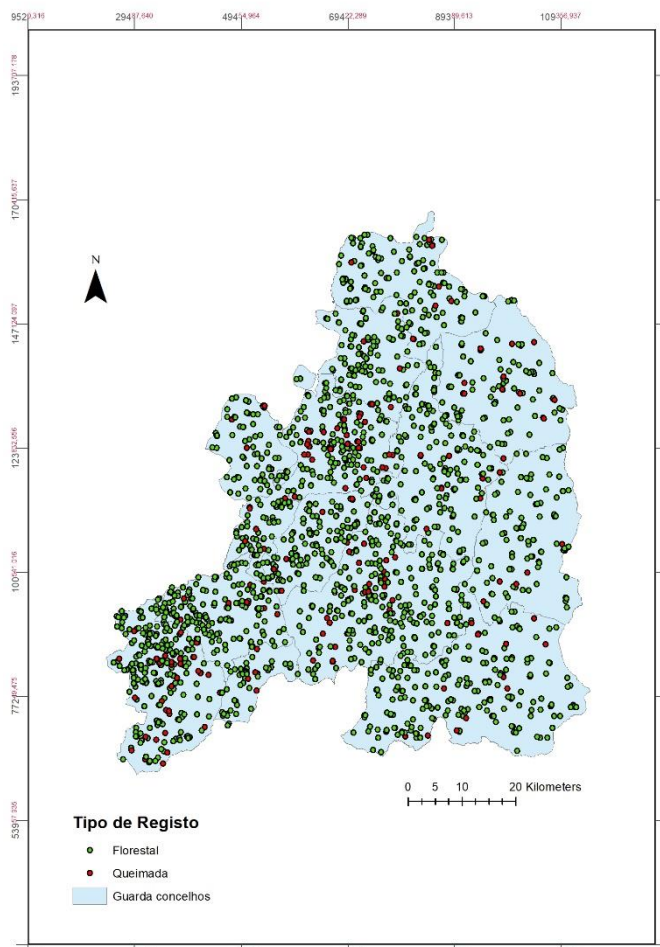


Figura 1 - Ocorrência de focos de Incêndios do Distrito da Guarda de 2001 a 2008

3.1 Visualização dos Dados Espaciais

Este passo permite ter uma percepção rápida da distribuição espacial dos dados bem como a utilização de alguns parâmetros estatísticos para a avaliar os dados. Uma vez que foram recolhidos dados correspondentes a vários anos foi verificada a distribuição temporal das ocorrências (figura 2). Verificou-se então que no ano de 2005 houve o maior número de ocorrências, 1405 e no de 2006 o menor 776 num universo de 8806 incêndios.

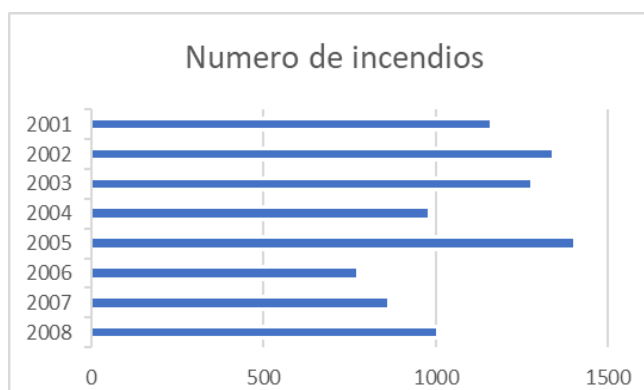


Figura 2 - Número de focos de Incêndio por ano no distrito da Guarda

Foi também possível verificar a densidade de incêndios por km² no distrito para cada ano. Foi utilizado a área do distrito da Guarda de 5518km². Os resultados são apresentados na tabela seguinte (tabela 1). Para o conjunto dos anos a densidade de incêndios por km² é de 1,1598 incêndios por km². É possível verificar que o ano de 2005 apresenta mais uma vez o maior valor.

Tabela 1 - Numero de fogos e numero de fogos por Km² para o distrito da Guarda - Valores de 2001 a 2008 e Total

ANO	NÚMERO DE INCÊNDIOS	DENSIDADE
2008	1007	0,182494
2007	864	0,156578
2006	776	0,140631
2005	1405	0,254621
2004	982	0,177963
2003	1281	0,232149
2002	1343	0,243385
2001	1164	0,210946
TOTAL	8822	1,598768

3.1.1 Centrografia

Foram utilizadas medidas de centrografia para caracterizar os dados. Para a construção destas medidas foram utilizadas as ferramentas do software ArcMap Central Feature, para a medida Central Feature, Mean Center para a medida Centro Geométrico. Para a medida Distância Padrão foi utilizada a ferramenta Standard Distance e para a elipse padrão foi utilizada a ferramenta Direcional Distribution.

Estas medidas permitem determinar geograficamente a zona de concentração dos eventos e medir a sua dispersão.

As medidas calculadas foram então as seguintes:

- ❖ *Central Feature* – localização do evento mais central.
- ❖ Centro Geométrico – localização de maior concentração de pontos.
- ❖ Distância padrão – Dispersão dos eventos em torno do centro geométrico
- ❖ Elipse Padrão – Orientação da distribuição espacial dos eventos.

É possível verificar que o centro Geométrico (a verde, figura 3) se localiza na zona Centro/Oeste do distrito, mais especificamente no concelho de Celorico da Beira. Dentro do círculo de menor diâmetro (68% de confiança) estão localizadas as maiores concentrações de pontos. No diâmetro maior (correspondente a 95% de confiança) os eventos encontram-se mais dispersos. A análise da figura 3b permite verificar através da elipse padrão que a orientação da distribuição dos eventos é aproximadamente SW-NE

Análise Espacial de Dados

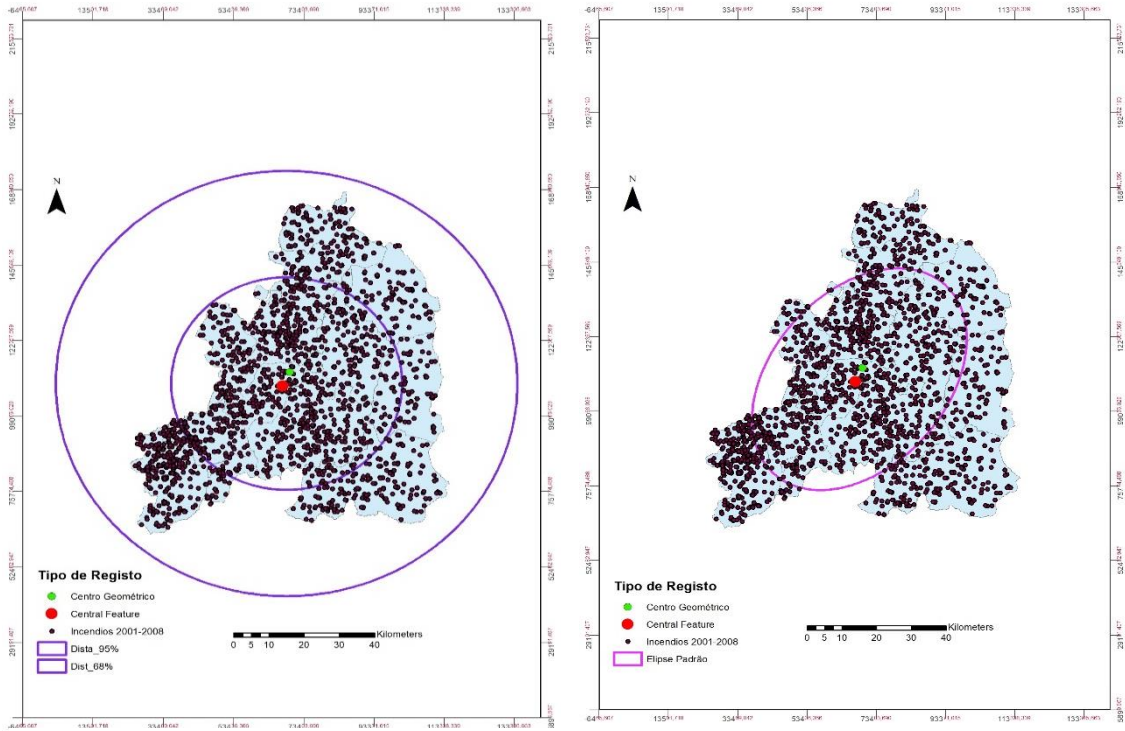


Figura 3a e 3b - Central Feature e Centro Geométrico para os focos de Incêndios totais do distrito da Guarda.
 Figura 3a - Distância padrão para 68 e 95% de confiança. Figura 3b - Elipse Padrão para 68% de confiança

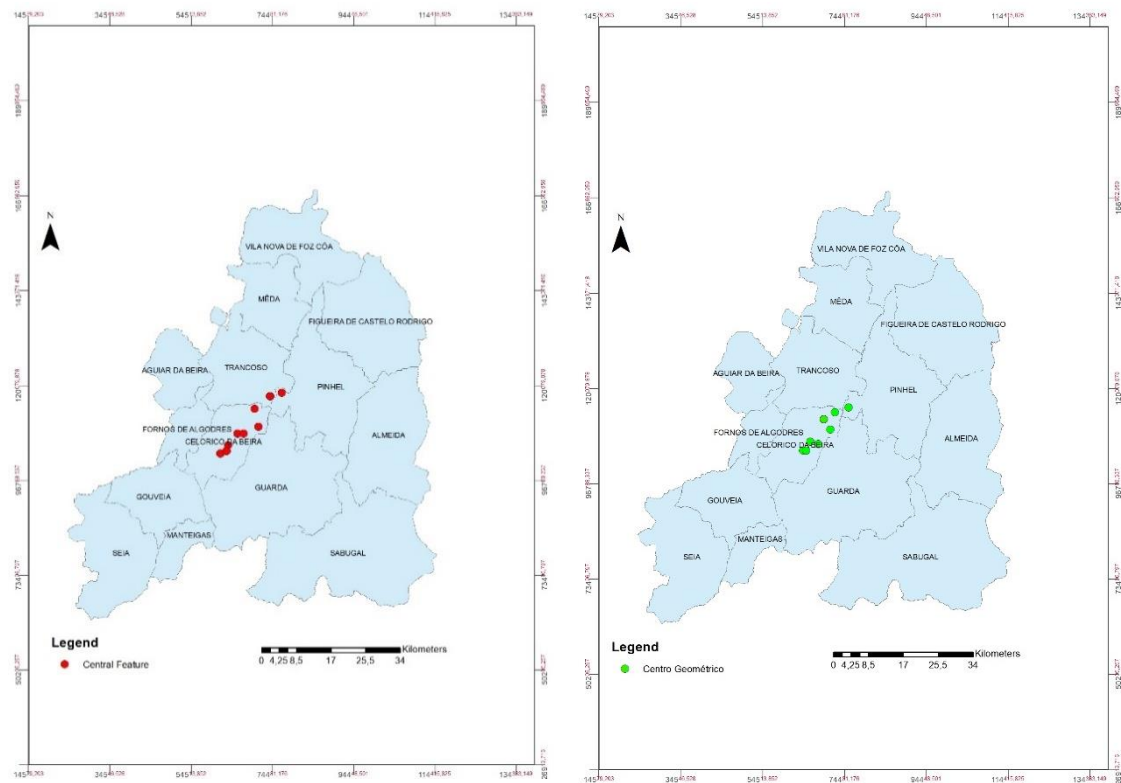


Figura 4a - Central Feature para todos os anos de focos de incêndio em análise. Figura 4b - Centro geométrico para todos os anos de focos de incêndios em análise

A análise dos mapas do centro geométrico (figura 4b) e *central feature* (figura 4a) permitem verificar que apesar da grande distribuição temporal dos eventos, estas medidas permanecem quase constantes na localização geográfica. Em relação ao centro geométrico, todos os anos calharam no concelho de Celorico da Beira, exceto no ano de 2004, que calhou no concelho de Pinhel. A *central feature* surgiu todos os anos também no concelho de Celorico da Beira, exceto no ano de 2004, em Pinhel e 2007 em Trancoso.

3.2 Análise Exploratória dos Dados Espaciais

Antes de passar á análise exploratória dos dados é importante definir o que são propriedades de primeira ordem e propriedades de segunda ordem. As propriedades de primeira ordem são propriedades consideradas globais. Ou seja, correspondem a variações nos valores médios dos eventos em toda a área. As propriedades de segunda ordem correspondem a efeitos locais e representam a covariância entre duas variáveis.

3.2.1 Técnicas para exploração de propriedades de 1ª Ordem (Globais)

Para a análise das propriedades de 1ª ordem do padrão pontual foram utilizadas as técnicas do Método de contagem por quadrantes e do *Kernel Density*.

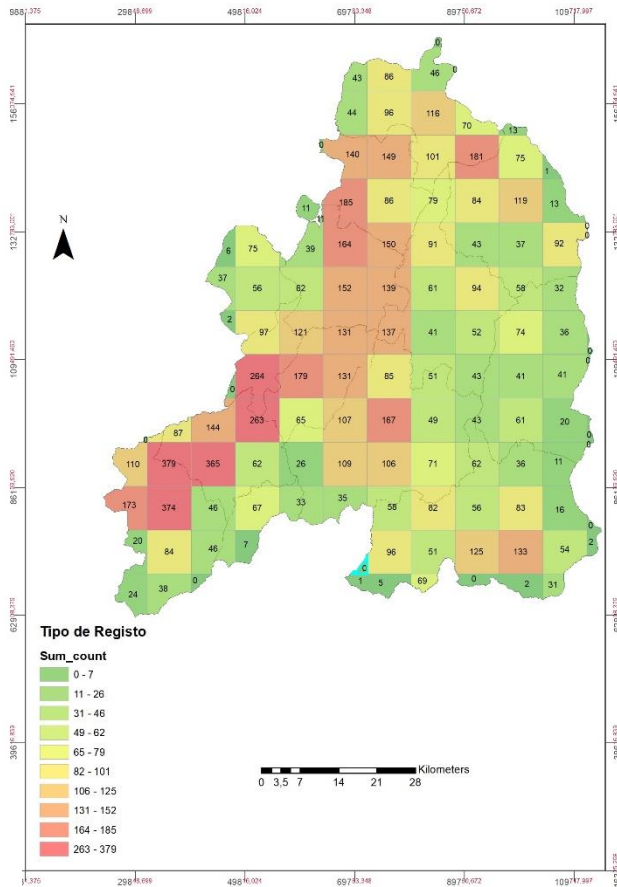
3.2.1.1 Método da Contagem por Quadrantes

O método da contagem por quadrantes é uma técnica de execução simples que envolve a criação de uma grelha regular simples e a contagem do número de eventos em cada quadrante. No entanto esta técnica apresenta limitações no que toca principalmente á definição dos quadrantes. Quando os quadrantes definidos têm uma área demasiado grande, a descrição do padrão pontual torna-se difícil e pouco precisa. No entanto, quadrantes demasiado pequenos podem não conter eventos, facto que deve ser evitado.

A figura seguinte (figura 5) mostra o resultado do método da contagem por quadrantes para a área do distrito da Guarda.

Para iniciar a aplicação desta ferramenta foi primeiro construída uma rede de quadrantes utilizando a ferramenta *Create Fishnet* do software *ArcMap*.

Figura 3 - Mapa do Método da contagem por quadrantes para os focos de incêndios do distrito da Guarda



Esta ferramenta permitiu criar uma malha quadrangular com quadrantes com um lado de 8500 metros.

Esta malha foi então recortada com o limite do distrito da Guarda. No passo seguinte adicionaram-se os pontos aos quadrantes e efetuou-se a contagem.

Uma análise do mapa gerado revela que a grande maioria dos focos de incêndios se situam na zona Sudoeste do distrito, com alguns quadrantes com um grande número de focos também na zona norte e centro/oeste.

Por fim esta técnica permite calcular a média da contagem de pontos por quadrante e também a variância.

Com base nos valores apresentados na tabela 2 foi possível calcular o rácio da variância com a média. Este rácio permite determinar o tipo de padrão que a área apresenta. A tabela seguinte apresenta o VTMR calculado para esta técnica. Valores de VTMR inferiores a 1 indicam um padrão regularmente disperso, enquanto valores próximos de zero indicam um padrão aleatório. Valores superiores a 1 indicam um padrão agregado (presença de *clusters*). Como o valor calculado é superior a 1 considera-se o padrão agregado.

Tabela 2 - Variáveis Calculadas para o método da contagem por quadrantes

Variáveis Calculadas	Valores
NÚMERO de Quadrantes (<i>n</i>)	118
NÚMERO de Focos de Incêndio	8806
Média	74,63
$\sum X_i$	8806
$\sum X_i^2$	1290982
$\frac{(\sum X_i)^2}{n}$	657166,41
Variância	5417,23
VTMR	72,59
X_i^2	8493,1002
Teste X^2	81,353

Esta técnica permite ainda efetuar a modelação do padrão pontual, através do teste estatístico Qui-Quadrado. O valor de X^2 determinado para este padrão pontual permite rejeitar a hipótese nula (que o padrão é aleatório) e pode-se afirmar que o padrão é aglomerado.

3.2.1.2 Kernel Density

A técnica da kernel density permite visualizar a intensidade dos conjuntos de dados pontuais numa superfície contínua. Esta técnica cria então uma superfície de influência e conta o número de ponto no seu interior, determinando o número de eventos por unidade de área. Esta técnica permite identificar mais facilmente os hot-spots na área em estudo, mas, no entanto, é uma técnica que também requer o ajustamento apropriado da superfície de influência

Para este projeto foi utilizada a ferramenta do software *ArcMap Kernel Density* para produzir a superfície. Foram utilizadas duas abordagens para a análise do padrão espacial utilizado esta técnica. A primeira consistiu na realização de um mapa de *Kernel Densities* para todos os focos de incêndios de todos os anos em causa.

Para a segunda abordagem foram construídos mapas de *Kernel Densities* para cada ano registado. Esta abordagem permitiu visualizar as diferenças de intensidade nos vários anos.

Analisando o mapa produzido pela primeira abordagem (figura 6) é possível visualizar que existem vários pontos na zona Sudoeste do distrito com uma grande intensidade de focos de incêndio.

Este mapa permite, portanto, verificar a existência de *hot-spots* com grande intensidade naquela zona, o que leva a sugerir que possam existir fatores naquela localização que são propícios ao começo de ignições de incêndios.

Em relação á segunda abordagem foram então criados mapas de *Kernel Density* para todos os anos em estudo. Estes mapas foram normalizados para os mesmos intervalos de valores para se poder efetuar uma comparação das densidades dos *hot-spots*. Os anos com maior número de incêndios possuem densidades consideravelmente superiores aos

restantes anos. Os anos em estudo podem ser divididos em anos com menos de 1000 focos e anos com mais de 1000 focos. Com mais de mil focos existem *Hot Spots* bastante intensos como pode ser visualizado na tabela 3.

Com o intuito de determinar se esses *hot-spots* se concentram numa mesma zona foram retiradas as zonas com uma densidade superior a 0,75 em todos os anos.

Tabela 3 - Densidades máximas dos hot-spots obtidos do Kernel Density para cada ano em estudo

Ano	Número de focos de incêndio	Densidade Máxima
2001	1164	1,023
2002	1343	1,770
2003	1281	1,411
2004	982	0,636
2005	1405	1,655
2006	776	0,613
2007	864	0,628
2008	1007	0,741

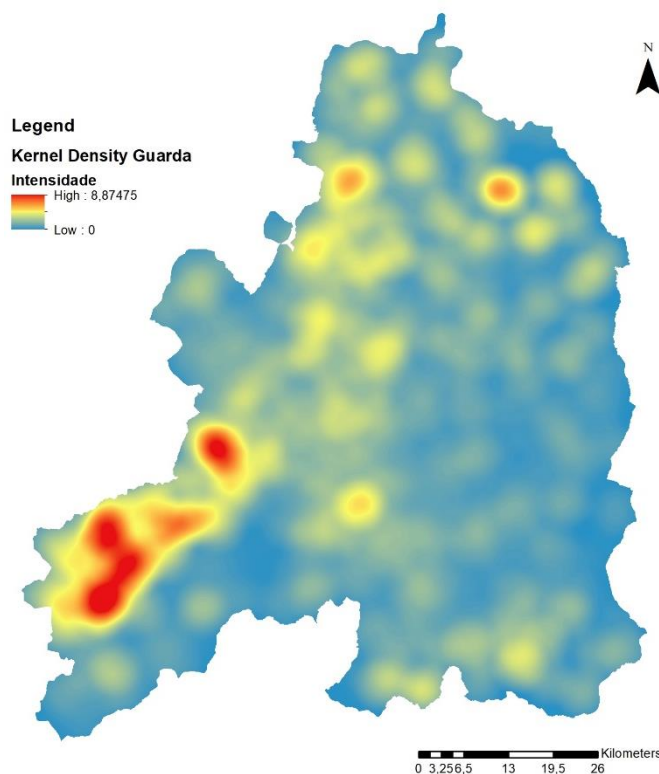


Figura 4 - Mapa de Kernel Density para a totalidade dos focos de incêndio do distrito da Guarda

Através de uma função de soma de rasters (Ferramenta *Raster Calculator*) é possível visualizar as zonas do distrito com vários anos de densidades de focos de incêndio grandes (figura 7). Estas zonas serão especialmente vulneráveis ou terão fatores que permitem esta concentração.

Análise Espacial de Dados

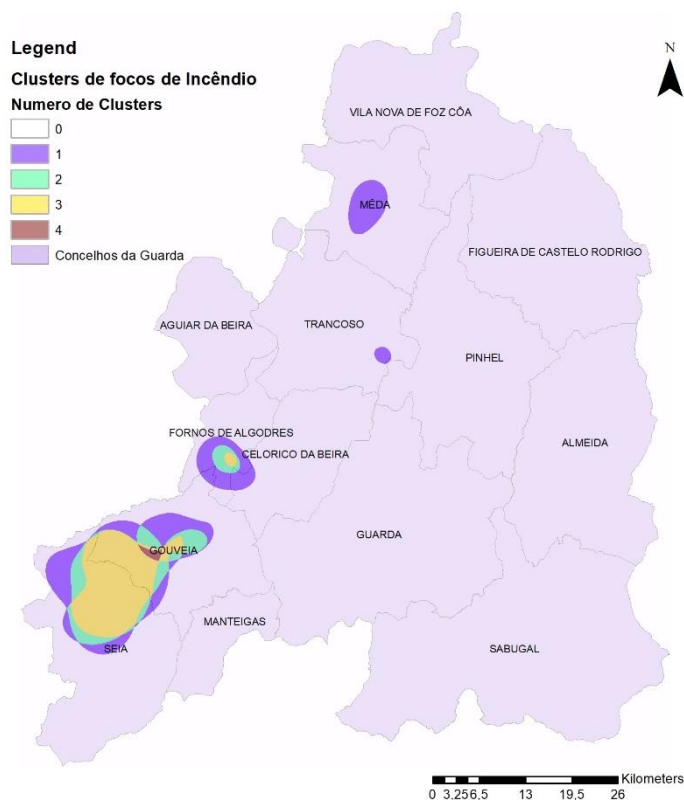


Figura 5 - Mapa de número de hot-spots para o distrito da Guarda

3.2.2 Técnicas de 2ª Ordem (Locais)

Para analisar as propriedades de 2ª ordem foram utilizadas as técnicas do vizinho mais próximo e da função k.

3.2.2.1 Vizinho mais próximo

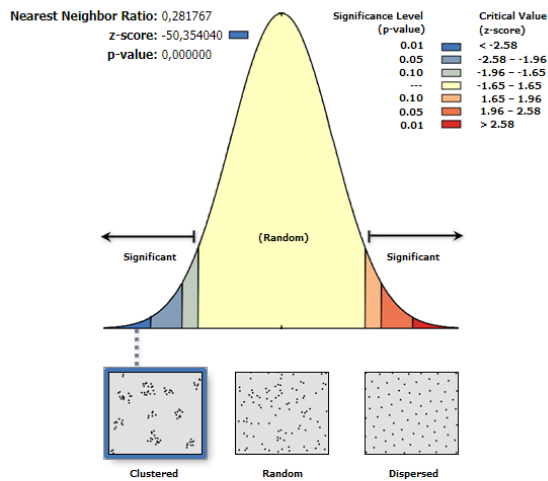
A técnica do vizinho mais próximo é uma técnica que determina as distâncias entre os eventos do padrão em estudo. Esta técnica pretende determinar as distâncias mínimas entre os eventos. Estas distâncias são acumuladas ao longo das distâncias entre todos os pontos. Posteriormente numa fase de modelação do processo pontual é aplicada uma função para determinar a possibilidade de rejeitar a hipótese nula (Aleatoriedade espacial)

Nesta técnica foi utilizada a ferramenta do software *ArcGis Average Nearest Neighbor*. Foram utilizadas nesta ferramenta a *shapefile* correspondente ao a todos os focos de incêndios de todos os anos e ainda as correspondentes a cada ano em estudo.

Foi possível determinar que todos os anos apresentam um padrão aglomerado, rejeitando-se a hipótese nula, existindo menos de 1% de hipótese de o padrão ser aleatório. O conjunto de todos os anos em estudo apresenta também um padrão agregado. Os valores para cada ano e para a totalidade dos focos encontram-se nas figuras 8 a 15.

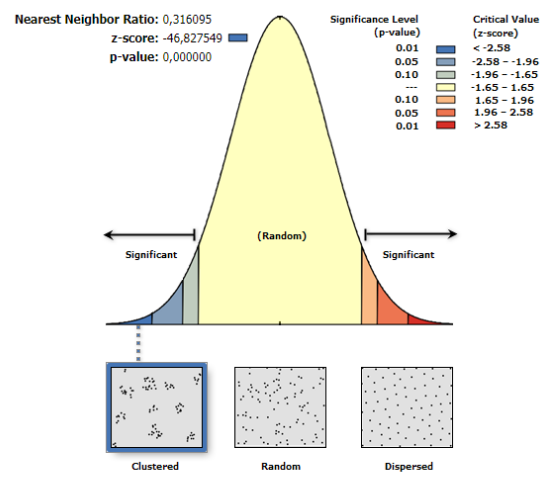
Análise Espacial de Dados

Figura 7 - Nearest Neighbor 2002

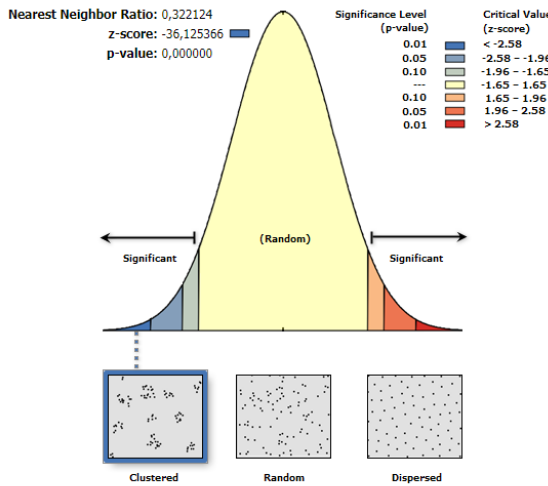


Given the z-score of -50.354039579, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figura 6 - Nearest Neighbor 2003

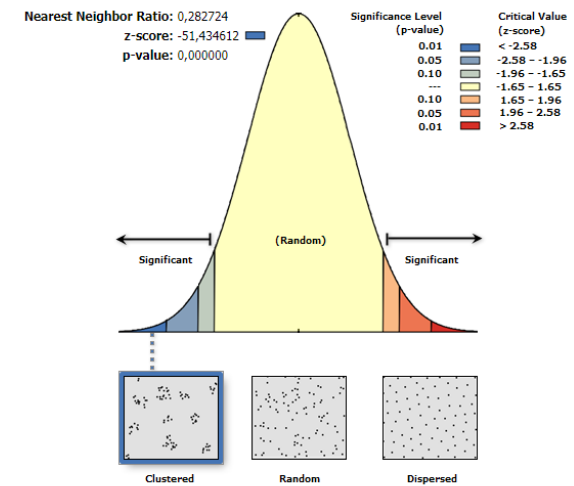


Given the z-score of -46.8275493732, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.



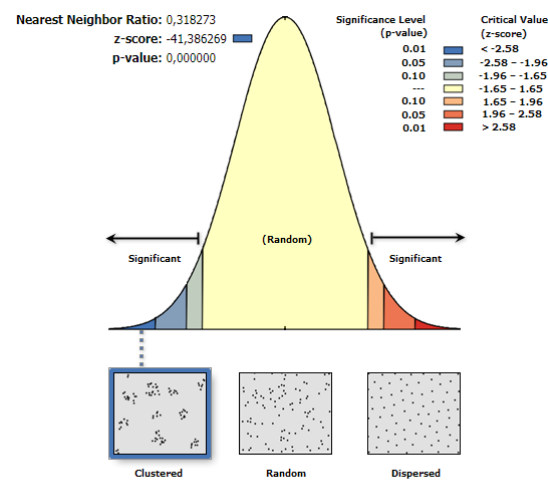
Given the z-score of -36.1253663499, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figura 10 - Nearest Neighbor 2004



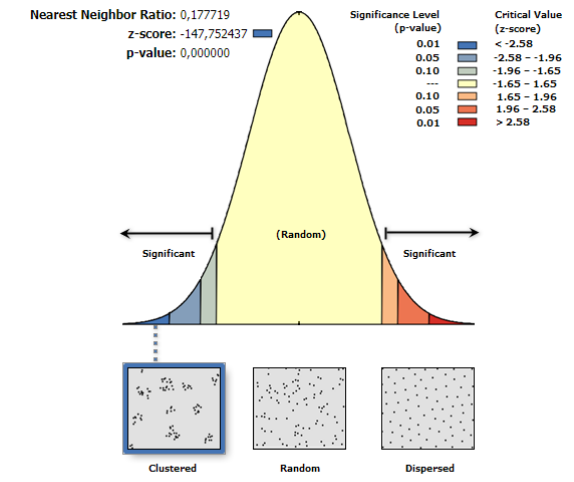
Given the z-score of -51.4346122466, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figura 11 - Nearest Neighbor 2005



Given the z-score of -41.3862694302, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

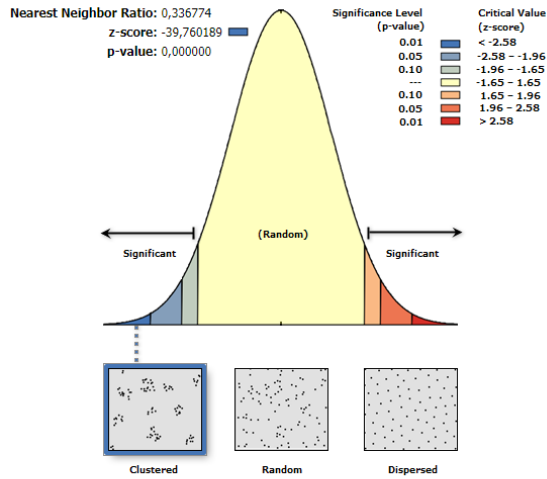
Figura 12 - Nearest Neighbor 2006



Given the z-score of -147.752436962, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

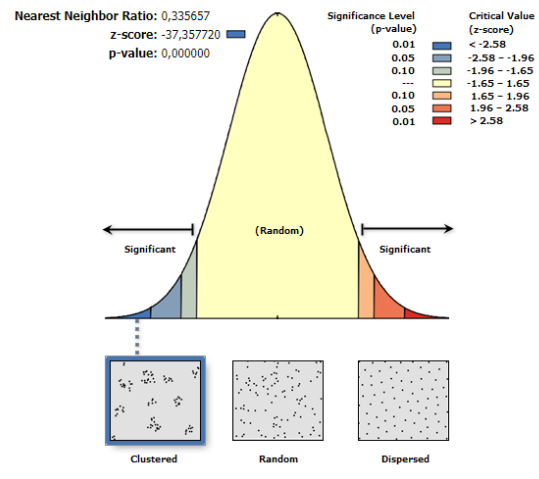
Figura 13 - Nearest Neighbor 2007

Análise Espacial de Dados



Given the z-score of -39.7601892904, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figura 14 - Nearest Neighbor 2008



Given the z-score of -37.3577204217, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Figura 15 - Nearest Neighbor Total dos focos de incêndios

3.2.2.2 Função K

A função K é uma função baseada na distância entre todos os pontos da área em estudo que permite colmatar as desvantagens que a técnica do vizinho mais próximo apresenta, nomeadamente o facto de em padrões muito aglomerados as distâncias serem muito pequenas. A função K é portanto mais sensível a intervalos maiores. Recorrendo a métodos estatísticos é possível realizar a modelação desta função para comparar a um modelo teórico que represente a aleatoriedade espacial. Para isto são criadas distribuições aleatórias e calculados os valores da função K para estes valores.

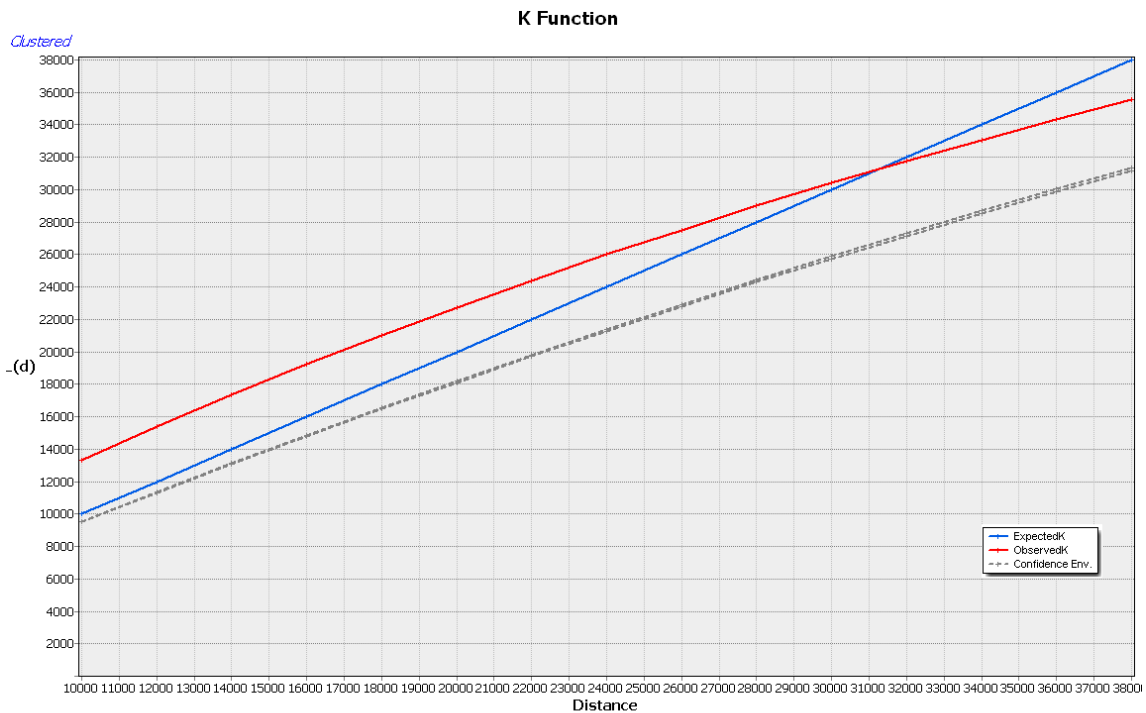


Figura 16 - Relatório da Função K para o teste de aleatoriedade espacial dos focos de incêndio

Quanto mais vezes este processo é repetido, com maior certeza é possível aceitar ou rejeitar a hipótese de existir aleatoriedade.

Esta função permite ainda determinar a distância a partir do qual a distribuição dos eventos passa de agregada para dispersa ou vice-versa.

Para a execução desta técnica foi utilizada a ferramenta *Multi-Distance Spatial Cluster Analysis (Ripley's K function)*. A ferramenta foi executada com 9 permutações. Havia ainda a hipótese de executar a ferramenta com 99 ou 999 permutações, mas o tempo de execução excessivo para este conjunto de dados impediu estas possibilidades. Na figura 16 apresentam-se os resultados da execução da ferramenta.

É possível verificar que até aproximadamente 31km o padrão se mantém aglomerado, passando a disperso a partir desta distância. De uma maneira geral os resultados desta função apontam para que exista aleatoriedade espacial.

4. Análise de Áreas

A análise espacial de áreas prevê um conjunto de técnicas que permite avaliar a autocorrelação espacial de atributos de áreas, investigar a existência de valores *outliers* no conjunto total dos dados ou apenas localmente. Como na análise de padrão pontual, podem ser estudados fenômenos globais ou locais. No caso de fenômenos globais avalia-se a variação da média na área em estudo. Nos fenômenos locais será avaliada a variância e a covariância na área em estudo.

Como na análise do processo pontual, a análise de áreas é dividida em 3 fases:

- Visualização de Dados Espaciais
- Análise Exploratória de Dados
- Significância Estatística

Devido ao *software* que será utilizado para a realização desta análise a fase da análise exploratória e a fase da Significância Estatística foram efetuadas em conjunto para cada técnica utilizada nesta fase.

4.1 Visualização dos dados espaciais

Para a fase da visualização de dados espaciais foram construídos mapas temáticos utilizando as variáveis estatísticas recolhidas: Percentagem de indivíduos analfabetos, percentagem de indivíduos desempregados e percentagem de indivíduos do sexo masculino. Os mapas foram construídos utilizando o *software ArcGis*. Para cada variável estatística foram construídos 3 mapas temáticos que representam as mudanças na visualização dos dados de acordo com o tipo de representação que é feita dos mesmos.

Na primeira representação (figuras 17) os dados foram divididos por intervalos iguais. Com esta representação podem surgir problemas se a variável a tiver muitos dados perto do valor máximo ou do mínimo, surgindo apenas uma cor no mapa temático, representando desses valores máximos ou mínimos.

Análise Espacial de Dados

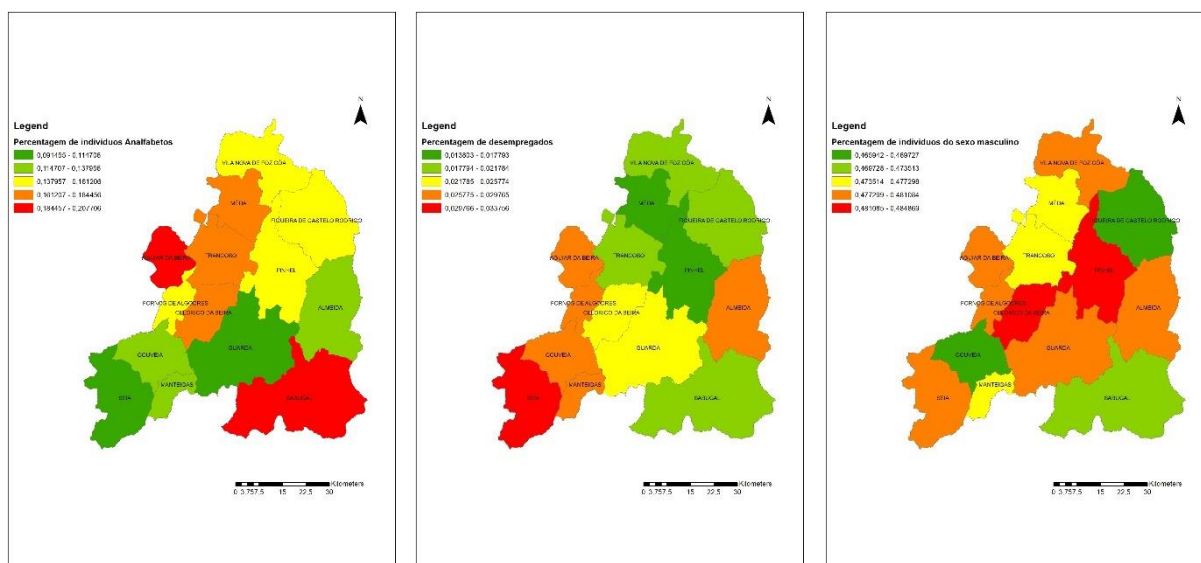


Figura 17 - Método de corte da variável com visualização por intervalos iguais. 17a – Indivíduos analfabetos. 17b – Indivíduos desempregados. 17c – Indivíduos do sexo masculino

É possível verificar nesta visualização que nos dados utilizados, os concelhos do Sabugal e Aguiar da Beira (figura 17a) apresentam o maior número de indivíduos analfabetos, enquanto que o concelho de Seia apresenta a maior taxa de desemprego (figura 17b). Os concelhos de Pinhel e Celorico da Beira apresentam o maior número de indivíduos do sexo masculino.

A representação por quantil (figuras 18) agrega os dados em quantidades iguais de eventos. Esta representação tem a desvantagem de poder camuflar diferenças significativas de valores de *outliers*.

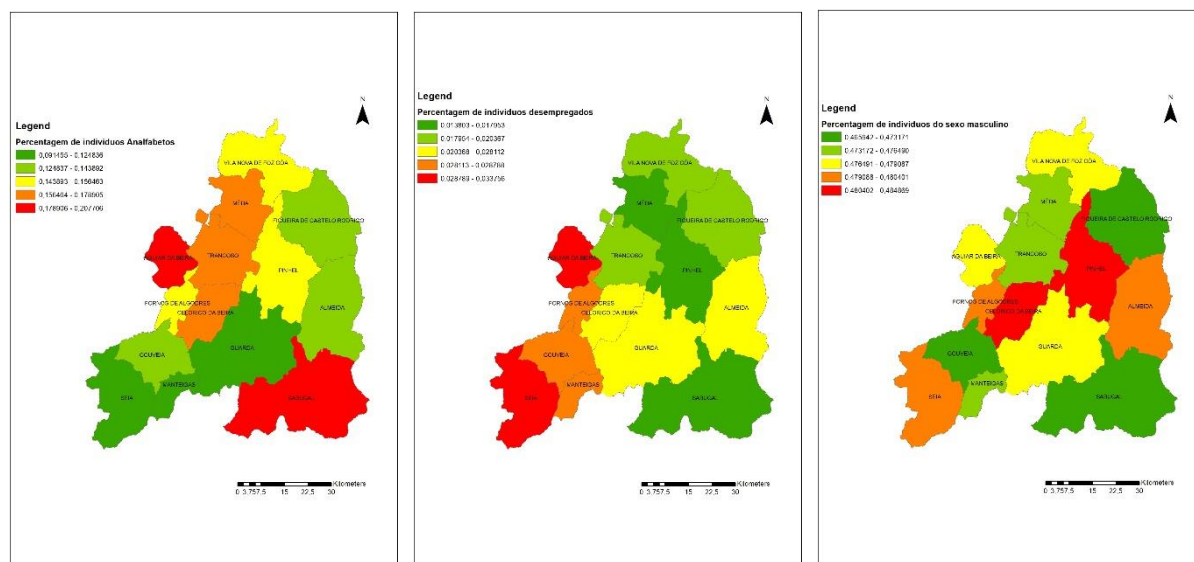


Figura 18 - Método de corte da variável com visualização por quantis. 17a – Indivíduos analfabetos. 17b – Indivíduos desempregados. 17c – Indivíduos do sexo masculino

Nesta representação foi possível verificar que os valores mais altos continuam representados nos mesmos concelhos que pelo método anterior. No entanto existem diferenças nas visualizações dos valores intermédios, nomeadamente nos dados referentes a indivíduos desempregados.

Por fim foram construídos mapas referentes á visualização por desvio padrão (figuras 19). Esta visualização mostra a dispersão em torno da média global da variável.

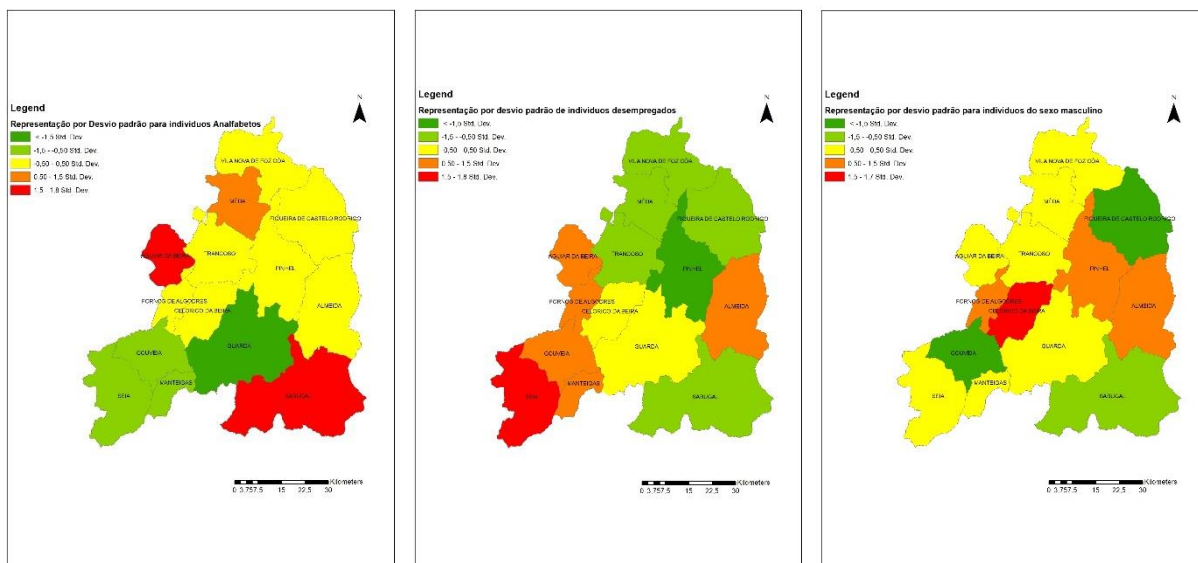


Figura 19 - Método de corte da variável com visualização por desvio padrão. 17a – Indivíduos analfabetos. 17b – Indivíduos desempregados. 17c – Indivíduos do sexo masculino

Nesta representação é possível verificar os concelhos que apresentavam os valores mais altos também apresentam os maiores desvios positivos ao desvio padrão.

4.2 Análise Exploratória de Áreas

A análise exploratória de áreas tem como objetivo principal medir a autocorrelação das variáveis das áreas da região em estudo. Como na análise de padrão pontual existem dois tipos de estatísticas de autocorrelação espacial, as globais e as locais.

As estatísticas globais vão considerar todas as observações da área em estudo e o pretendem mostrar a dependência espacial e mostrar como os valores estão correlacionados no espaço. As estatísticas locais pretendem identificar associações de valores entre as várias áreas vizinhas e a identificação de eventos extremos.

Para poder utilizar as ferramentas associadas a estas estatísticas é primeiro necessário criar uma matriz de vizinhança. Esta matriz descreve as relações de vizinhança ou inexistência das mesmas entre as várias áreas da zona em estudo. Para esta fase foram utilizados índices globais de autocorrelação espacial e índices locais.

4.2.1 Índices Globais de Autocorrelação Espacial

A autocorrelação espacial define-se como a medida do grau de influência de uma variável sobre si própria. A autocorrelação espacial avalia a influência que esse mesmo atributo tem no surgimento de valores semelhantes ou diferentes na sua vizinhança. Se a autocorrelação for positiva existe a presença de *clusters* espaciais, uma vez que ao valor de um atributo vão surgir valores semelhantes na vizinhança. Se pelo contrário os valores forem radicalmente diferentes na vizinhança existe autocorrelação negativa.

Para medir a existência ou não de autocorrelação espacial a nível global pode ser utilizado o índice de Moran I Global. Para utilizar este índice, bem como o teste de significância estatística utilizar-se-á o software *GeoDA*.

4.2.1.1 Índice de Moran I Global

Para a percentagem de indivíduos do sexo masculino no distrito foi verificado através deste índice que não existe autocorrelação espacial uma vez que o índice I para 999 permutações é de -0,235 (figuras 20), logo não se pode rejeitar a hipótese nula, ou seja a distribuição das percentagens de homens não dependem da localização espacial.

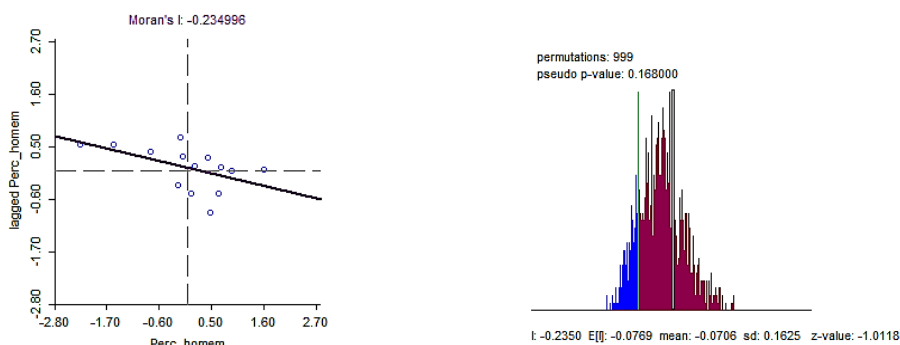


Figura 20a - Índice de Moran para a variável indivíduos do sexo masculino. Figura 20b- Teste de significância com 999 permutações

Também para a percentagem de indivíduos analfabetos (figura 21) no distrito não se pode rejeitar a hipótese nula uma vez que o valor do índice I (0,11) se aproxima do valor 0. Este teste foi também corrido com 999 permutações.

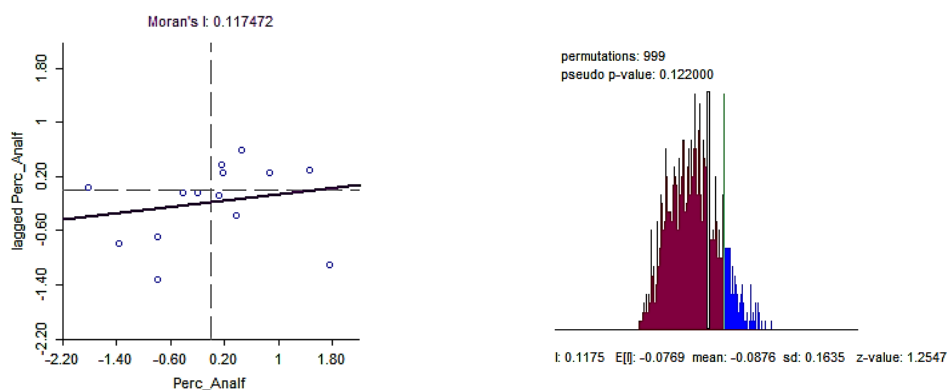


Figura 21 - Índice de Moran para a variável indivíduos analfabetos. Figura 21b- Teste de significância com 999 permutações

No caso da percentagem de indivíduos desempregados (figura 22) o valor de I de 0,375 e o facto de grande parte dos atributos se situarem no 1º e 3º quadrantes no *Moran's Scatter Plot* leva a que rejeitemos a hipótese nula e se possa afirmar que existe uma autocorrelação positiva para a percentagem de desempregados no distrito.



Figura 22 -Índice de Moran para a variável indivíduos desempregados. Figura 22b- Teste de significância com 999 permutações

Por fim foi testado número de focos incêndios por concelho no distrito da Guarda (figura 23). O valor do índice I é de -0,28 o que indica a ausência de autocorrelação espacial.



Figura 23 - Índice de Moran para a variável número de focos de incêndio. Figura 23b- Teste de significância com 999 permutações

4.2.2 Índices Locais de Autocorrelação Espacial

Para se determinar a autocorrelação espacial em cada área específica da região em estudo serão utilizados índices locais. Estes índices irão comparar o valor do atributo em estudo com o das áreas vizinhas, ou seja, comparam a covariância de um polígono com a sua vizinhança. Para a determinação destes índices irão ser utilizados dois indicadores: o índice local de Moran, e o índice G.

4.2.2.1 Índice Local de Moran

O índice Local de Moran vai comparar a covariância entre áreas para definir os seus resultados. Para a utilização deste índice, recorreu-se ao software *GeoDa*.

As figuras abaixo (figura 24) representam os mapas produzidos pelo índice Local de Moran para o atributo percentagem de indivíduos analfabetos. É possível verificar que existe um valor alto de autocorrelação no concelho do Trancoso, mas tal como no indicado no índice global não existem zonas de *clusters* no distrito. Para este mapa é possível verificar que o nível de confiança se situa entre 95% e 99,9%.

Análise Espacial de Dados

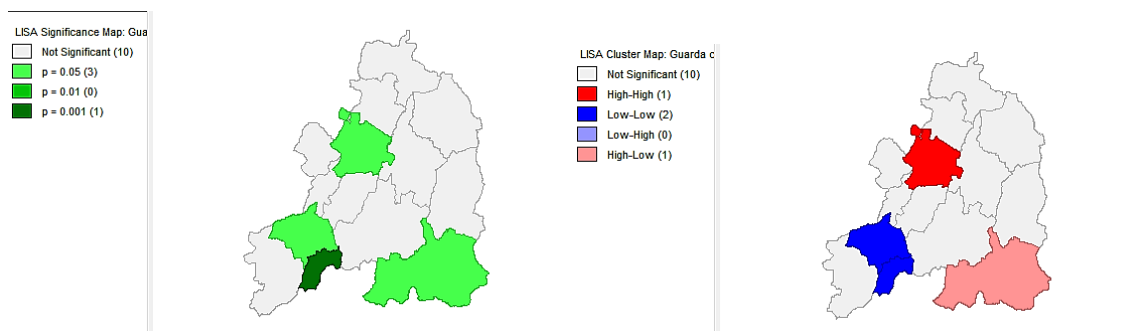


Figura 24 - Índice de Moran Local para a variável indivíduos analfabetos com mapa de nível de confiança

O mapa abaixo (figura 25) mostra os valores de índice Local de Moran para o atributo percentagem de indivíduos desempregados. Neste mapa é possível verificar a existência de *clusters* nos concelhos de Seia, Manteigas e Gouveia, na zona sul do distrito. Para estes distritos o nível de confiança é de 95%. É também perceptível neste mapa zonas de baixa autocorrelação espacial nos concelhos na zona norte do distrito, nomeadamente Pinhel, Mêda e Vila Nova de Foz

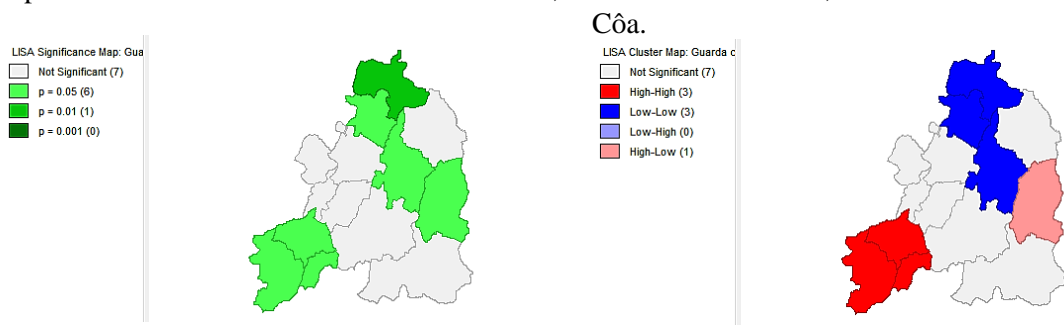


Figura 25 - Índice de Moran Local para a variável indivíduo desempregados com mapa de nível de confiança

A última variável em estudo foi a percentagem de indivíduos do sexo masculino no distrito (figura 26). Este atributo, tal como já havia sido constatado no índice global, não apresenta autocorrelação espacial.

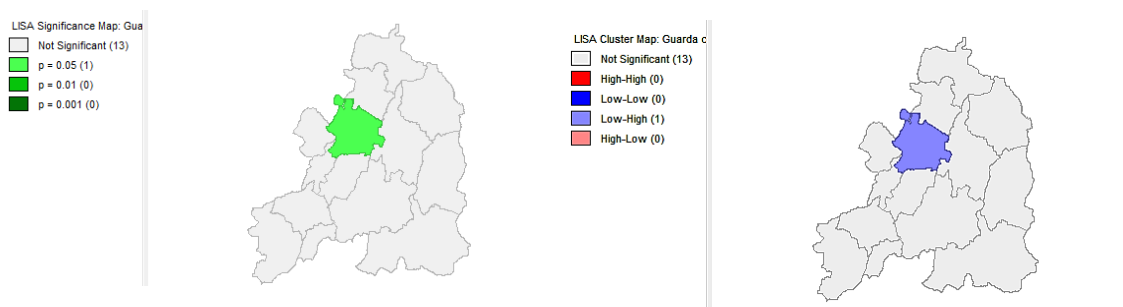


Figura 26 - Índice de Moran Local para a variável indivíduos do sexo masculino com mapa de nível de confiança

4.2.2.2 Índice G

A estatística G_i é um indicador local de autocorrelação espacial que utiliza a soma dos valores vizinhos definidos a partir de uma matriz de vizinhança. Foi utilizado o software *GeoDa* para a execução dos testes com este índice.

O índice G para a variável de percentagem de indivíduos analfabetos (figura 27) permite verificar a existência de um agrupamento de valores altos no concelho do Trancoso. Mostra também a existência de agrupamentos de valores baixos nos concelhos do Sabugal, Gouveia e Manteigas. A significância estatística para estas análises é de 99% para o concelho de Manteigas e 95% para os restantes enumerados.

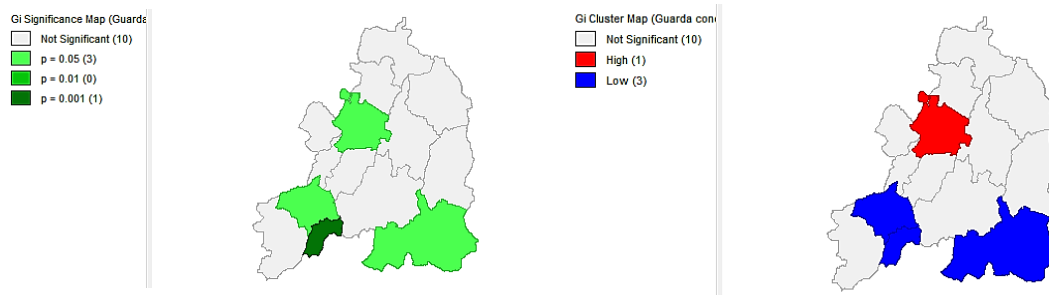


Figura 27 - Estatística G_i para a variável indivíduos analfabetos com mapa de nível de confiança

A análise com o índice G para a variável de número de indivíduos desempregados (figura 28) confirma os resultados do índice Local de Moran com a presença de agrupamento de valores altos do atributo nos concelhos de Seia, Gouveia e Manteigas. Na zona oriental do distrito existe um agrupamento de valores baixos deste atributo.

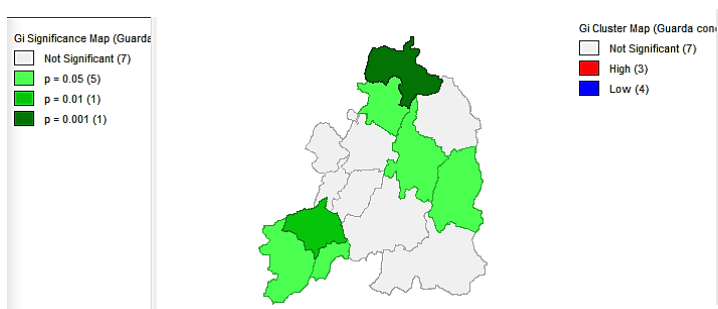


Figura 28 - Estatística G_i para a variável indivíduos desempregados com mapa de nível de confiança

Em relação á percentagem de indivíduos do sexo masculino (figura 29) verifica-se a existência de um concelho (Trancoso) com valores altos do atributo. O nível de significância para este atributo é de 95%.

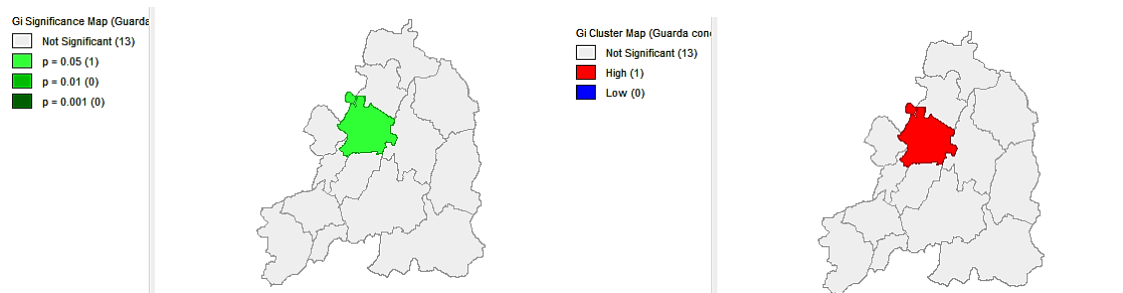


Figura 29 - Estatística G_i para a variáveis indivíduos do sexo masculino com mapa de nível de confiança

4.3 Conclusões sobre a autocorrelação espacial

Com as ferramentas de análises de área foi possível concluir que apenas a variável de percentagem de indivíduos desempregados apresenta algum nível de autocorrelação espacial. As restantes variáveis apresentam aleatoriedade. Em termos locais é possível verificar a existência de alguns clusters para as variáveis em estudo.

5. Regressão Espacial

Um modelo de regressão é uma ferramenta estatística que permite que dadas duas variáveis, uma delas (variável dependente) possa ser descrita ou estimada em relação à outra variável (variável independente). Esta ferramenta tem como objetivos encontrar um bom ajuste entre os valores estimados e os observados e descobrir que variáveis melhor contribuem para a relação linear.

Para se avaliar a existência ou não de uma relação entre variáveis utiliza-se um diagrama de dispersão. Para isto num diagrama são projetadas as duas variáveis, uma em cada eixo. Sobre a projeção de uma variável em relação à outra tenta-se ajustar uma reta de regressão. Quanto mais perto da reta os pontos projetados se situarem melhor é considerado o modelo.

Nesta fase do projeto serão utilizadas as variáveis sociais recolhidas para a fase da análise de áreas: Número de indivíduos desempregados, número de indivíduos analfabetos e número de indivíduos do sexo masculino.

Para avaliar o ajustamento dos dados à reta de regressão criada utilizar-se-á o coeficiente de determinação (R^2). Este valor indica em percentagem (de 0 a 1) o quanto o modelo consegue explicar os valores observados.

Na tentativa da criação de um modelo de regressão espacial a hipótese nula será sempre a de as variáveis não serem correlacionadas, ou seja, os resíduos dos modelos são independentes e não correlacionáveis com a variável dependente.

Para avaliar a rejeição ou não da hipótese nula irão ser mapeados os resíduos dos modelos escolhidos. Estes irão ser projetados num diagrama de dispersão e será ainda avaliada a sua autocorrelação espacial.

O passo final desta etapa será então adotar um modelo de regressão (espacial ou simples) que melhor se ajuste aos nossos dados.

Todos os passos descritos nesta fase foram executados recorrendo ao *software GeoDa*.

5.1 Diagrama de dispersão

Tabela 4 - Ajustamento das várias variáveis ao modelo linear simples a partir de um diagrama de dispersão

Variável	R^2	Percentagem de ajustamento
<i>Percentagem de indivíduos do sexo masculino</i>	0.139	13.9%
<i>Percentagem de indivíduos analfabetos</i>	0.058	5,8%
<i>Percentagem de indivíduos desempregados</i>	0.002	0,2%

Ao analisar a tabela 4 foi possível perceber que nenhuma das variáveis escolhidas tem um ajustamento aceitável ao modelo linear simples. Isto indica que nenhuma das variáveis escolhidas contribui de maneira decisiva para o número de fogos existentes em cada concelho.

Este resultado era de certo modo esperado, pois os fogos de origem criminosa são normalmente iniciados com um incendiário que é o conjunto das 3 variáveis aqui analisadas. Por esta razão foi decidido que o melhor modelo de regressão seria o modelo de regressão linear múltipla, que abrangeria todas as variáveis em estudo.

5.2 Escolha do modelo de Regressão

Os modelos de regressão que se procuram criar terão sempre como variável dependente o número de focos de incêndio por concelho, uma vez que é esse facto que procuramos explicar a partir do modelo.

Inicialmente foi criado um modelo de regressão múltipla que irá ser analisado e serão mapeados os seus resíduos para verificar se existe ou não dependência espacial das variáveis.

Para a escolha do modelo de regressão que melhor se ajusta aos dados, além da percentagem de ajustamento é importante analisar o critério AIC. Para o *Alkaline info criterion* (AIC) quanto menor for o seu valor, melhor será o ajustamento do modelo de regressão.

Tabela 5 - Variáveis e ajustamento para um modelo regressivo multivariável

Variáveis	R²	Percentagem de ajustamento	AIC	Índice de Moran
- Percentagem de indivíduos do sexo masculino - Percentagem de indivíduos analfabetos - Percentagem de indivíduos Desempregados	0.267	26.7%	201.575	-0.4003

Pode ser constato a partir dos valores obtidos, que este modelo já explica de maneira mais satisfatória os valores apresentados. No entanto a percentagem de ajustamento é ainda muito baixa. Seguidamente foram mapeados os resíduos. O índice de Moran (figura 30) obtido para o mapeamento dos resíduos indica autocorrelação espacial negativa, o que indica algum nível de dependência espacial dos mesmos.

Análise Espacial de Dados

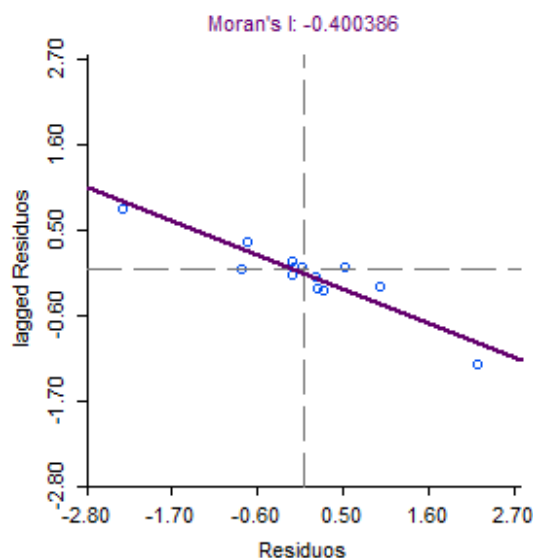


Figura 8 - Índice de Moran Global para os resíduos do modelo multivariável

Por fim foram experimentados os dois modelos de regressão com efeitos espaciais globais. Os modelos são o *Spatial Lag Model* e o *Spatial Error Model*. O primeiro atribui a autocorrelação espacial á variável independente e o segundo atribui a autocorrelação ao erro.

Tabela 6 - Modelos de regressão utilizados e o seu ajustamento aos dados

Modelo de Regressão	R²	Percentagem de Ajustamento	AIC
Regressão múltipla sem efeitos espaciais	0.267	26.7%	201.575
<i>Spatial Lag Model</i>	0.600	60.0%	197.959
<i>Spatial Error Model</i>	0.277	27.7%	191.64

A análise da tabela 6 permite concluir que o modelo *Spatial Lag Model* é o que apresenta melhor ajustamento aos dados utilizados. Apesar do valor de AIC não ser o menor, este é o que apresenta de longe o melhor ajustamento aos dados. A tabela 7 abaixo apresenta as equações determinadas para cada modelo de Regressão.

Tabela 7 - Equações para cada modelo de Regressão; NF=Número de focos por concelho; PA- Percentagem de indivíduos analfabetos; PD- Percentagem de indivíduos desempregados; PH- Percentagem de indivíduos do sexo masculino

Modelo de Regressão	Equação de Regressão
Regressão múltipla sem efeitos espaciais	NF=18355.1-3607.73PA-9851.77PD-35502PH
<i>Spatial Lag Model</i>	NF=16753.5-0.903784WY _i -3959.93PA-9160.16PD-30866.6PH
<i>Spatial Error Model</i>	NF=18390.8-3608.85PA-9905.46PD-35574.3PH-0.0181464

5.3 Conclusões sobre a Regressão Espacial

Apesar do modelo *Spatial lag Model* (com correlação espacial) ser o que melhor se ajusta aos dados, este ajustamento é bastante baixo (aproximadamente 60%) e explicaria deficientemente os valores dos números de focos de incêndios por todo o distrito. Este baixo ajustamento deve-se ao facto de os fatores escolhidos não serem os que melhor explicam os incêndios neste distrito. No entanto existem possivelmente outros fatores que levam aos focos de incêndio. No entanto esta análise não indica que os focos não tenham origem criminosa, apenas os fatores considerados não apresentam uma relação espacial com os focos suficientemente satisfatória que os permita explicar.

6. Aplicação de métodos de interpolação espacial

A interpolação espacial é a metodologia aplicada para a estimação de valores desconhecidos através de pontos de amostra conhecidos numa mesma área ou região. Esta converte pontos de observação em superfícies contínuas. Para se conseguir uma superfície de interpolação são necessários um conjunto de pontos a interpolar e um método de interpolação ótimo para esses mesmos pontos.

A precipitação é um fator importante a ter em conta para a determinação do risco de incêndio. Devido a este facto tentou-se criar uma superfície que mostre a variação do nível de precipitação

a nível nacional para o mês de junho (um dos meses com maior risco de incêndio). O ideal seria que a superfície fosse apenas o distrito da Guarda dado que esta é a área em que se localiza o foco deste projeto. No entanto, devido ao número de pontos ser baixo para uma análise mais completa decidiu-se que uma análise apenas deste distrito seria desadequada sendo mais interessante verificar a variação a nível nacional (continente).

Para tal foram então utilizados dados do Instituto Português do Mar e Atmosfera (IPMA) referentes á precipitação em Portugal Continental num período compreendido entre 1950 e 2003.

A base de dados foi construída com uma resolução horizontal de 0,2° e

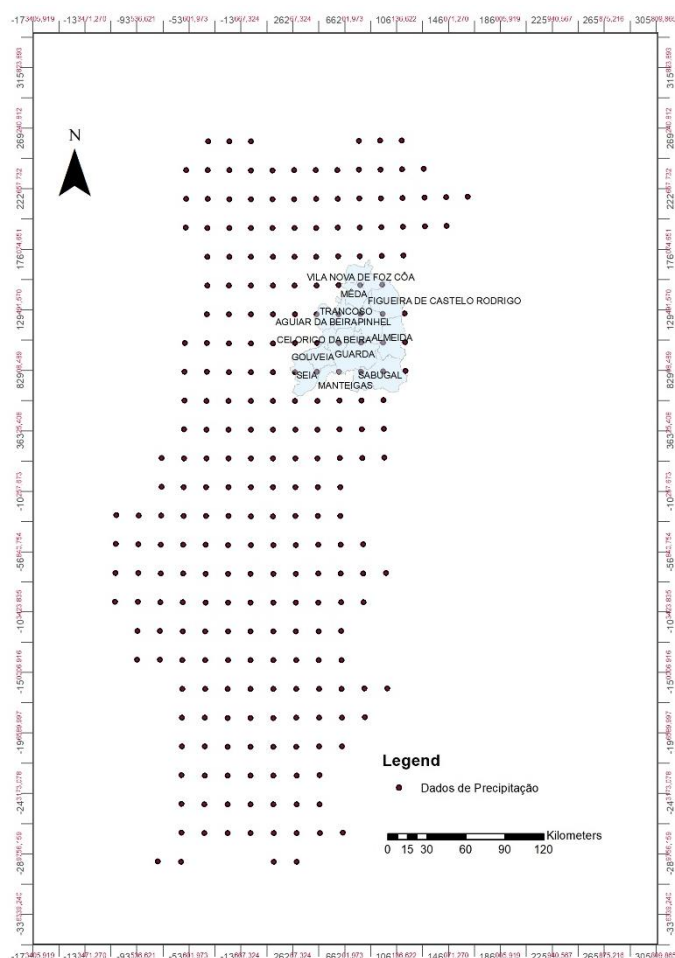


Figura 9 - Distribuição dos pontos a interpolar numa malha regular com o distrito da Guarda

foram utilizados dados de 188 estações meteorológicas e 618 pontos udométricos.

Os dados foram distribuídos em ficheiro de texto e apresentavam para cada mês todos os valores de precipitação de cada ano e de cada ponto de observação de uma malha regular.

Para a utilização em *ArcMap* foi utilizado um *script Python* que recolheu os valores de todos os anos e devolveu o valor médio de cada ponto e de cada mês na malha de pontos (figura 31). Os pontos encontravam-se no sistema de coordenadas WGS-84 pelo que foi necessário reprojeta-los no sistema ETRS98-tm6.

Para a execução dos métodos de interpolação foram utilizados os valores de precipitação correspondentes às médias de todos os anos do mês de Junho. Foram usados três métodos de interpolação: Kriging simples, Inverse Distance Weighting e Global Polinomial Interpolation. Para a avaliação da precisão dos métodos de interpolação foram retirados 10 pontos do conjunto. Embora estes pontos não sejam os ideais para realizar a avaliação, serão a melhor possibilidade para a verificação dos métodos. Os 10 pontos foram retirados aleatoriamente do conjunto de pontos de precipitação através de um *script Python*.

6.1 Global Polynomial Interpolation

O primeiro método utilizado foi a *Global Polynomial Interpolation* (GPI). Este interpolador é considerado determinístico e global.

Este gera uma superfície suavizada que vai procurar ajustar-se ao conjunto de pontos da amostra através de uma função matemática, neste caso um polinómio. Uma vez que se trata de uma função global esta função vai captar o padrão dos grandes comprimentos de onda dos dados. A superfície gerada raramente irá passar pelos pontos originais da amostra.

Para a interpolação utilizando este método será utilizado um polinómio de segundo grau. Foram testados polinómios de grau superior, mas estes começaram a apresentar problemas nas fronteiras.

Na figura seguinte (figura 32) apresenta-se o resultado do interpolador utilizando os valores de precipitação do mês de junho.

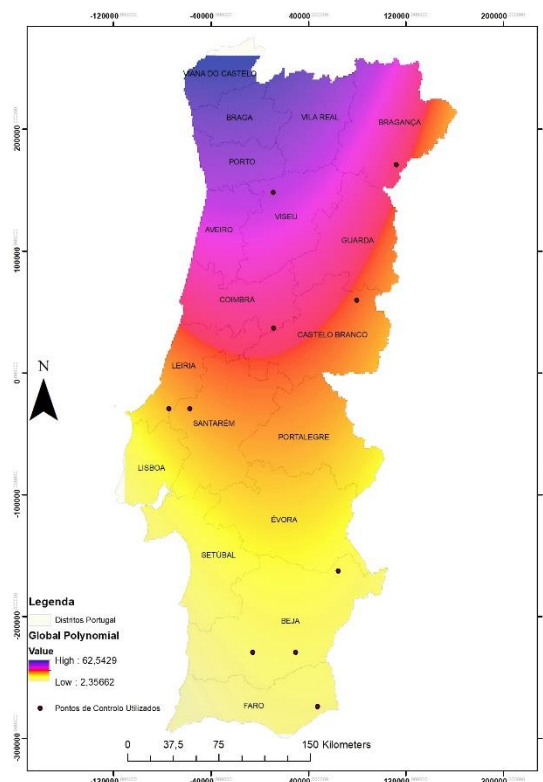


Figura 10 - Superfície gerada pela GPI com os pontos de controlo identificados

É possível verificar pela análise da figura que o interpolador captou a tendência global do mês em que os maiores valores de precipitação estão concentrados na zona NE do país.

6.2 Inverse Distance Weighted

O interpolador Inverse Distance Weighted (IDW) é um interpolador local determinístico que permite estimar o valor de um ponto com base nos pontos conhecidos. Este assume que quanto mais perto os pontos a estimar estiverem dos conhecidos, maior será o peso destes últimos, ou seja a influência dos pontos mais perto será maior que a dos pontos distantes.

A superfície criada por este interpolador será, portanto, baseada no peso dos pontos conhecidos mais próximos, nos pontos a estimar. Os parâmetros utilizados para a construção da superfície são os parâmetros ótimos (figura 33) estimados pelo software *ArcMap* através da ferramenta *Geostatistical Analyst*.

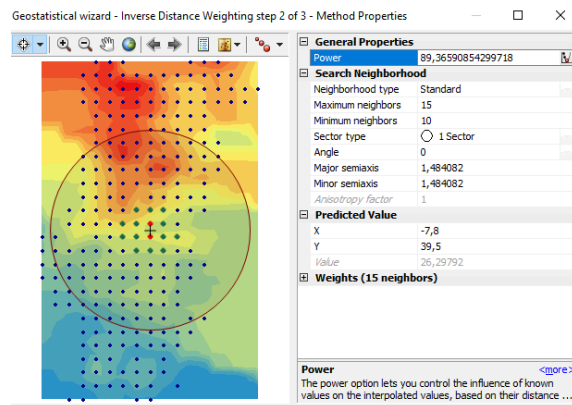


Figura 11 - Parâmetros estimados pela ferramenta do *ArcMap* ótimos para o interpolador IDW

Por fim foi gerado a superfície interpolado pelo IDW. Esta é apresentada na figura 34. Pode-se verificar ao analisar os resultados deste interpolador que devido á malha regular utilizada para os pontos as zonas de influência de cada ponto são também regulares, o que não corresponde ao visto na natureza. Este é, portanto, um mau interpolador para este tipo de dados. Esta conclusão é possível sem recorrer á *cross-validation* verificando apenas a superfície gerada.

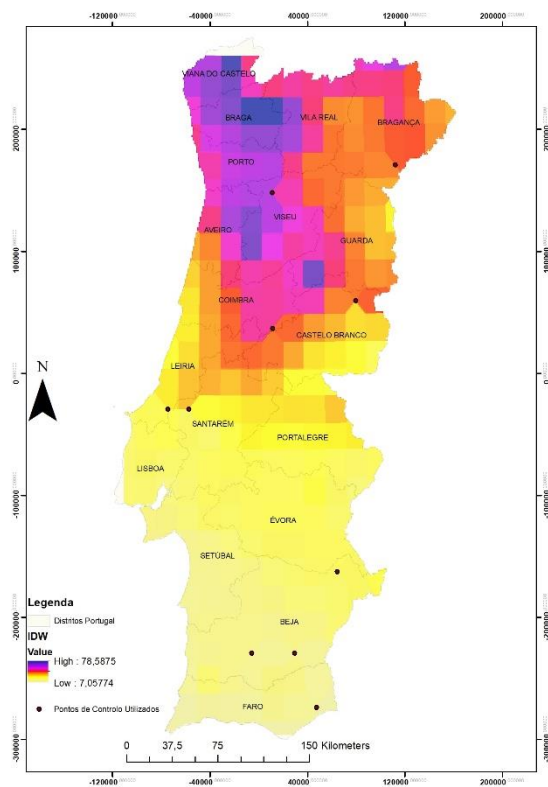


Figura 12 - Superfície gerada pelo IDW com os pontos de controlo identificados

6.3 Kriging

O kriging é um método de interpolação espacial local e probabilístico que permite construir uma superfície a partir dos dados pontuais disponíveis, recorrendo á modelação da estrutura da correlação espacial de dados.

Na construção da superfície interpolada a partir do kriging existem três fases, que incluem a análise exploratória dos dados, a modelação do variograma e por fim a interpolação estatística da superfície, baseada no variograma. Uma vez que a ferramenta *Geostatistical Analyst* do *ArcMap*, executa as primeiras tarefas automaticamente, irá ser construída logo a superfície interpolada. Os parâmetros para a construção desta superfície foram os considerados ótimos, pelo *software*.

Os parâmetros utilizados encontram-se na seguinte figura (figura 35a e 35b)

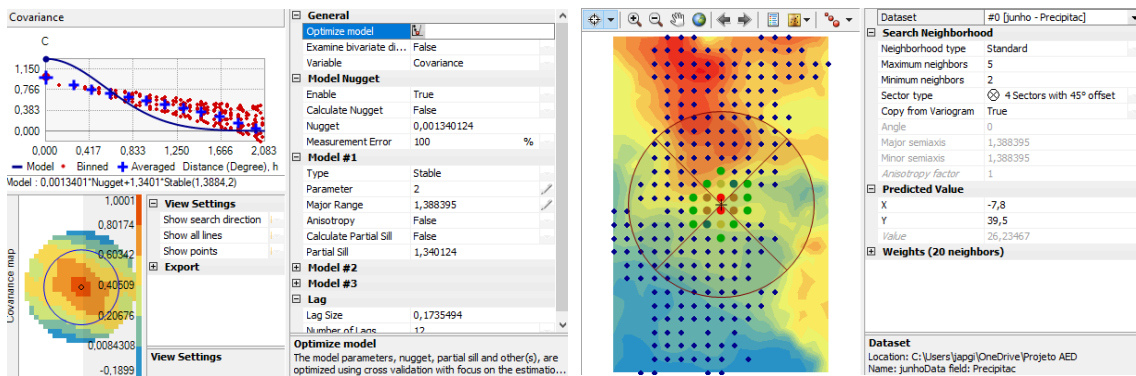


Figura 13a e 35b - Parâmetros ótimos determinados pelo ArcMap para o interpolador kriging simples

É possível verificar que o modelo ótimo considerou que o modelo de busca seria isotrópico, ou seja, sem variações em nenhuma das direções. Considerou ainda que o número máximo de pontos vizinhos a utilizar no cálculo dos valores para a superfície seria 5 e o mínimo seriam 2. Por fim foi também apresentado a *Cross-Validation* possível do *kriging*, ou seja, o erro da superfície gerada em relação aos pontos utilizados (figura 36).

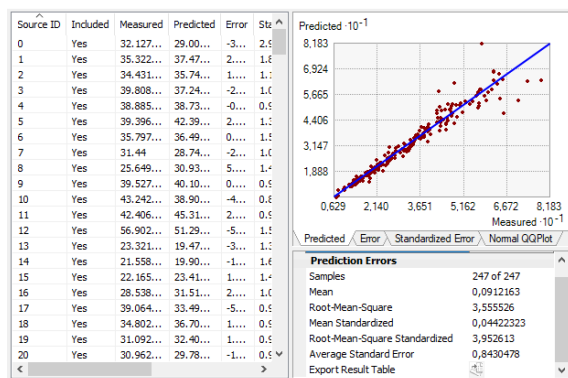


Figura 14 - Cross Validation do interpolador kriging

O erro médio em relação aos pontos utilizados foi de aproximadamente 0,09 enquanto que o RMS foi de 3,55. Estes valores podem-se considerar bons para este método, mas não se consegue ainda verificar a diferença para os restantes métodos.

Por fim apresenta-se a superfície interpolada (figura 37).

Análise Espacial de Dados

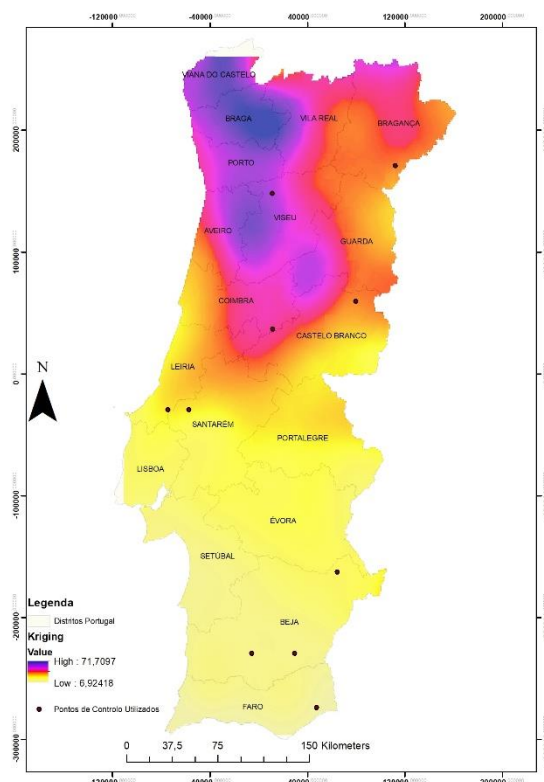


Figura 15 - Superfície gerado por kriging com os pontos de controlo identificados

6.4 Avaliação dos erro e Conclusões sobre os métodos de interpolação

Para a comparação dos métodos de interpolação foi utilizada a ferramenta do *software ArcMap Extract Multi Values to Points* para permitir extrair o valor, em cada *Raster* de superfície de interpolação, dos pontos de controlo utilizados. Os valores foram retirados para uma tabela *excel* onde foi possível calcular o erro médio de cada método de interpolação, ou seja, a diferença média, entre os valores estimados e os valores reais no ponto. Com esta metodologia é possível verificar na tabela 7 que o método que apresenta o erro menor é o *kriging* simples. Utilizando o *Root Mean Square* (RMS) que utiliza a média dos erros ao quadrado, podemos verificar que mais uma vez o *kriging* simples que apresenta o erro mais baixo.

Tabela 8 - Erro médio e RMS para cada interpolador

Métodos de Interpolação	Erro Médio (mm)	RMS
<i>Global Polinomial Interpolation</i>	2,833359355	35,7346465
<i>Inversed Distance Weighted</i>	2,147649305	13,10297724
<i>Kriging Simples</i>	1,714828885	10,30196668

É possível por isto concluir que o interpolador de *kriging* simples é o mais adequado á interpolação dos dados recolhidos. O interpolador IDW apresenta também erros baixos, mas isto deve-se á utilização de pontos retirados de uma malha regular, como pontos de controlo.

Verifica-se que a superfície gerada pelo IDW não é representativa de fenómenos de natureza meteorológica, pelo que este método para este tipo de dados, deve ser rejeitado. Quero por fim agradecer ao IPMA pelos dados utilizados neste estudo (Dataset de precipitação PT02).

7. Análise de Redes

Por fim foi feita uma análise de redes que permita determinar a possibilidade de utilização de poços de abastecimento de água, por parte das cooperações de bombeiros da Guarda.

Para realizar esta análise foram utilizados 3 algoritmos de análises de redes: *Route (Optimal Path)*, *Service Area e Closest Facility*. Para reduzir a quantidade de processamento necessário no tratamento de redes optou-se por realizar apenas esta análise utilizando vias, poços e cooperações de bombeiros apenas do concelho da Guarda, o mais populoso do distrito.

7.1 Recolhas dos dados

Para esta etapa do projeto foi necessário recolher novos dados, nomeadamente:

- ❖ Cooperações de Bombeiros a nível nacional

Dados pontuais correspondentes á localização geográfica das cooperações de bombeiros a nível nacional. Foram retirados deste *shapefile* apenas os dados referentes ao concelho da Guarda, utilizando a ferramenta *intersect* do *ArcMap*. Dados com o sistema de coordenadas ETRS89-tm06 pelo que não necessitavam de tratamento adicional.

- ❖ Rede viária do concelho da Guarda

Vias principais e secundárias obtidas a partir do *software Open Source Qgis*, do website <http://www.openstreetmap.org/>. Dados reprojatados no sistema de coordenadas ETRS89-tm06. Foram selecionadas as vias consideradas principais, secundárias e terciárias por se considerar que são as mais indicadas para a circulação de viaturas de bombeiros.

- ❖ Poços de água do concelho da Guarda

Os poços de água do concelho foram também obtidos a partir do *software Open Source Qgis* e retirados do website <http://www.openstreetmap.org/>. Os dados foram mais uma vez reprojatados no sistema de coordenadas ETRS89-tm06 e guardados numa nova *shapefile*.

7.2 Limpeza da rede viária e criação do *Network Dataset*

Para poder ser utilizada em algoritmos de redes foi necessário efetuar uma limpeza da rede viária. Esta incluiu a ligação de vias que não estavam completas, por erros dos dados e ainda a limpeza de rotundas e outras interseções de vias que deviam ser feitas apenas com a interseção de vários arcos num nó. Após este processo ser concluído em *software ArcMap*, a *shapefile* foi importada para o *ArcCatalog* onde se procedeu á criação de um *Network Dataset* a partir do qual se podem utilizar as ferramentas de análises de redes.

7.3 Aplicação das ferramentas de Análise de Redes

A primeira ferramenta a ser aplicada foi a do *optimal path*. Para a aplicação desta ferramenta foram utilizados como pontos de paragem os poços de água presentes no concelho da Guarda, numa tentativa de verificar qual seria o percurso ideal para um carro da cooperação que necessite de se abastecer de água em vários pontos do concelho.

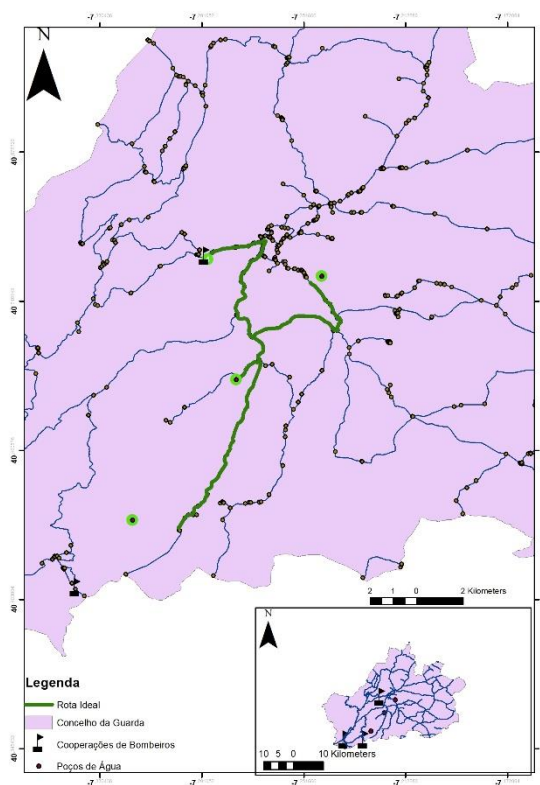


Figura 16 - Ferramenta de análise de redes de *optimal path* com as cooperações de bombeiros e poços de água identificados

É possível verificar na figura acima (figura 38) que o percurso construído permite verificar a maneira mais indicada para passar pelos vários poços. No entanto também se verifica que dois dos poços de água não têm vias principais ou secundárias perto, pelo que o acesso dos carros dos bombeiros fica dificultado.

Verificada a deficiência da abordagem anterior, a falta de vias adequadas á passagem dos carros de bombeiros, foi executado a ferramenta de *Closest Facility* que permite determinar os pontos mais próximos a uma infraestrutura pela rede criada. Para esta ferramenta foram utilizadas todas as cooperações de bombeiros do concelho, bem como todos os poços do concelho. O resultado é apresentado na figura abaixo (figura 39).

Com a análise providenciada é possível agora verificar que a presença de vias de comunicação adequadas até aos poços determinam que estes sejam os mais próximos às cooperações. Neste mapa verifica-se que todas as cooperações de bombeiros do concelho têm apenas um poço considerado o mais próximo. No entanto esta análise apenas considerou poços de água no concelho da Guarda, pelo que a utilização de uma rede de todo o distrito com poços para todos os distritos poderá fornecer resultados diferentes. A natureza académica deste relatório torna, no entanto, estes resultados adequados.

Análise Espacial de Dados

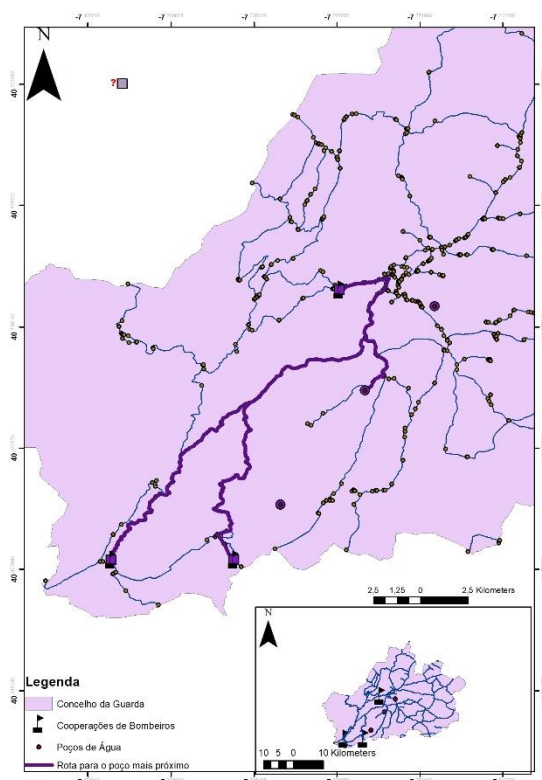


Figura 17 - Ferramenta de análise de redes de Closest Facility com as cooperações de bombeiros e poços de água identificados

Por fim, para determinar a área de influência de cada poço na rede viária foi utilizado a ferramenta *Service Area*. Esta ferramenta permite então verificar em vários patamares até onde chega a influência dos vários poços. Os resultados surgem na figura abaixo (figura 40).

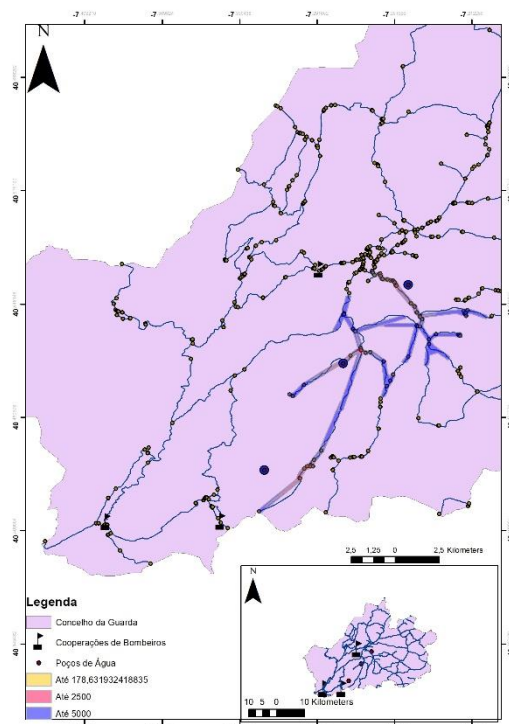


Figura 18 - Ferramenta de análise de redes de Service Area com as cooperações de bombeiros e poços de água identificados

Este mapa mostra que os vários poços têm áreas de influência grandes, apesar de nenhuma cooperação de bombeiros aqui cair, facto que leva a que apenas um poço seja o ideal para a utilização.

7.4 Conclusões da análise de redes

A análise de redes permitiu verificar que apesar de existirem vários poços de água no concelho da Guarda, apenas um é o ideal para a utilização pelas cooperações de bombeiros, se necessário. Isto deve-se á presença de uma rede viária de estradas mais adequadas á passagem dos carros dos bombeiros. O poço mais adequado tem o nome de “Fonte de Mergulho”.

8. Conclusões

A execução deste projeto permitiu concluir que os focos de incêndios no distrito da Guarda apresentam um perfil agregado, dispondo-se em clusters ao longo do distrito. A zona onde os clusters mais se concentram é a zona oeste do distrito. Analisando cada ano individualmente é possível verificar que anos com menor número de focos de incendio vão apresentar *clusters* com uma densidade mais baixa que os anos com pelo menos mil focos. É possível identificar na zona sudoeste do distrito uma área onde se focam os *clusters* mais densos de focos de incêndio anualmente, levando a concluir que aquela zona apresenta algum fator de risco.

A análise de áreas permite verificar que das variáveis sociais escolhidas a partir do perfil do incendiário, apenas a percentagem de indivíduos desempregados apresenta algum grau de autocorrelação espacial. A nível local existem concelhos com *clusters* destas variáveis, nomeadamente para a percentagem de indivíduos desempregados e percentagem de indivíduos analfabetos.

Na tentativa de desenvolver um modelo de regressão adequado aos focos de incêndios foi determinado que um modelo que contabilize todos as variáveis sociais seria o ideal. Foi também verificado através dos indicadores R^2 e AIC que um modelo que contabilize as variáveis espaciais seria o mais adequado. Isto foi determinado pois os resíduos do modelo regressivo simples apresenta autocorrelação espacial. O modelo escolhido foi o *Spatial Lag Model*. No entanto a percentagem de ajustamento deste modelo aos dados reais continua baixo (60%) pelo que se verifica que possivelmente as variáveis escolhidas não são as mais corretas para a explicação dos dados.

Uma vez que as variáveis meteorológicas são de grande importância para o risco de incêndios florestais foram utilizados dados de precipitação médios para o mês de junho de modo a criar uma superfície que mostre a distribuição da precipitação para o país, uma vez que os dados não apresentam detalhe suficiente para se fazer uma avaliação mais adequada a nível do distrito. A partir deste trabalho foi verificado que o melhor interpolador para este tipo de dados seria o *kriging* simples.

Por fim foi verificado pela análise de redes do concelho da Guarda, distrito da Guarda, que a presença de uma rede viária principal melhora o acesso a infraestruturas (poços de água) por parte das cooperações de bombeiros.

Bibliografia

- Catita, Cristina. Apontamentos da disciplina Análise Espacial de Dados (2018)
- Belo-Pereira, M., E. Dutra, and P. Viterbo, 2011: 'Evaluation of global precipitation data sets over the Iberian Peninsula', *J. Geophys. Res.*, 116, D20101, doi:10.1029/2010JD015481