

ANÁLISE DA DISTRIBUIÇÃO ESPACIAL DE INCÊNDIOS NO DISTRITO DE SETÚBAL

Pedro Dias

(dias296@hotmail.com)

RESUMO

O objectivo deste projecto é realizar uma análise espacial dos incêndios florestais e agrícolas em Portugal Continental durante o período de 2006 até 2010, mais especificamente para o distrito de Setúbal. Este projecto foi organizado em 3 trabalhos parciais consoante o tipo de dados e os objectivos analisados.

No 1º trabalho parcial pretendeu-se estudar a existência de aglomerados ou clusters de incêndios no distrito de Setúbal através da aplicação de técnicas globais de 1ª ordem como o Kernel Density e técnicas locais de 2ª ordem como o método do vizinho mais próximo e a função k. Identificaram-se cerca de 66 clusters principais para um nível de confiança superior a 95%, localizando-se a maior parte na zona norte do distrito.

No 2º trabalho parcial pretendeu-se avaliar a autocorrelação espacial do número de fogos em Portugal Continental, tentando definir um modelo de regressão explicativo para a ocorrência desses mesmos fogos. Para estudar a autocorrelação espacial, utilizaram-se técnicas globais de 1ª ordem como o índice de Moran Global e técnicas locais de 2ª ordem como a técnica LISA, identificando-se claramente duas zonas de aglomeradas muito semelhantes, uma na zona litoral Norte e outra na zona E-SE do país, para uma significância de 99.9%. Relativamente ao modelo de regressão, utilizou-se modelos de regressão espaciais, SLM e SEM, para justificar a ocorrência de incêndios florestais e agrícolas para o período de 2006 a 2010, com base em variáveis independentes como a temperatura, a precipitação e o número de quarteis de bombeiros por concelho. Os resultados obtidos não foram conclusivos, demonstrando que as variáveis independentes são pouco explicativas do número de incêndios agrícolas ou florestais.

Por último, no 3º trabalho parcial, pretendeu-se numa primeira fase analisar se a distribuição das áreas aridas no distrito de Setúbal, durante o período em estudo, poderá ser justificado com as condições meteorológicas de temperatura e precipitação. Posteriormente foi também realizado um estudo de optimização dos percursos percorridos pelas viaturas dos bombeiros, desde os seus aquartelamentos até aos focos de incêndios mais prováveis identificados no trabalho parcial 1.

Relativamente à primeira fase, utilizaram-se métodos de interpolação exactos determinísticos (IDW) e probabilísticos (Kriging) para produzir padrões espaciais a partir de dados meteorológicos pontuais. Concluiu-se que apenas se consegue identificar uma correlação espacial entre os focos de incêndio mais prováveis e as zonas com temperaturas mais elevadas. Relativamente à segunda parte, utilizaram-se algoritmos de análise de redes, como o Closest facility e o Service Area. Os resultados foram uma lista de trajectos mais curtos entre quarteis bombeiros e focos de incêndios mais prováveis e gráficos com zonas alcançadas a partir de cada quartel para distâncias de 5 e 10 km.

1. Introdução

O objectivo deste trabalho é realizar uma análise espacial de um fenómeno ocorrido no espaço, através da utilização de um sistema de informação geográfico (SIG). O tema escolhido foi a análise da distribuição espacial de incêndios florestais e agrícolas no distrito de Setúbal que ocorreram entre 2006 e 2010.

A análise espacial visa compreender a distribuição dos dados com origem em fenómenos ocorridos no espaço, medindo as suas relações face à sua localização. De forma geral pode-se definir a análise espacial como uma metodologia que visa detectar padrões ou anomalias nos dados e que se subdivide em 3 áreas fundamentais: A visualização dos dados, a análise exploratória dos dados e a modelação

dos mesmos. As técnicas utilizadas para cada fase, dependem do tipo de dados em estudo, assim, pode-se efectuar uma análise pontual, areal ou de rede.

Este projecto visa aplicar as técnicas dos 3 tipos distintos de análises possíveis assim, vai-se organizar em 3 trabalhos parciais. O trabalho parcial 1, com o tema análise da distribuição espacial pontual de incêndios no distrito Setúbal; o trabalho parcial 2, com o tema análise de autocorrelação e regressão espacial de incêndios por concelho e o trabalho parcial 3, com o tema métodos de interpolação e análise de redes.

Ao longo de todo o projecto, foram utilizados operadores disponibilizados por vários sistemas de informação geográfica (SIG) como o software *ARCGIS* Versão 9.3 [ESRI, 1999-2008] e o software *OpenGeoDa* Versão 0.9.9.8 [GeoDa Center for Geospatial Analysis and Computation, 2011]. Todos os dados utilizados no software SIG, foram uniformemente transformados para o sistema de coordenadas PT-TM06/ETRS89.

2. Trabalho Parcelar 1

Este primeiro trabalho parcial consiste na análise espacial de incêndios no distrito, de forma a estudar a sua distribuição e concluir sobre a existência de aglomerados ou clusters.

Seleccionaram-se todos os registos de incêndios florestais e agrícolas, no site AFN [2012], que ocorreram entre 2006 e 2010, no distrito de Setúbal. Eliminaram-se os registos de falsos alarmes e queimadas.

Foram também utilizados dados sobre a rede viária, do site CloudMade [2012], sobre a rede hidrográfica, do site Atlas do Ambiente [2012] e sobre a ocupação do solo, do site do Instituto Geográfico Português [2012], para realizar análises comparativas com os clusters identificados ao longo deste trabalho parcelar.

Após importar os dados verificou-se a existência de alguns outliers que se teve que corrigir, conforme ilustra a figura 1.

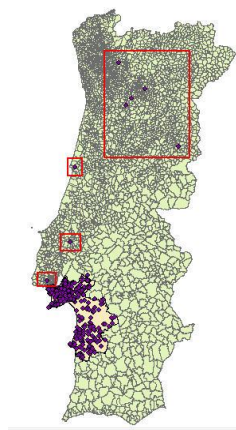


Figura 1– Outliers de incêndios no distrito de Setúbal

A metodologia adoptada para resolver este trabalho parcelar foi organizada nas três etapas já enumeradas anteriormente. A visualização, a análise exploratória de dados (propriedades de 1ª e 2ª ordem) e a modelação de dados (*Complete Spatial Randomness*) do conjunto em análise.

2.1 – Visualização de dados do tipo pontual

Esta fase da análise da distribuição de padrões de pontos consiste na utilização de um conjunto de operadores que permitem uma rápida percepção de como os pontos se distribuem no espaço, figura 2,

com base na orientação espacial ou da sua variação. Procedeu-se então à identificação e localização de ocorrências no espaço através da técnica do centro geométrico e das elipses padrão (68%, 95% e 99%).

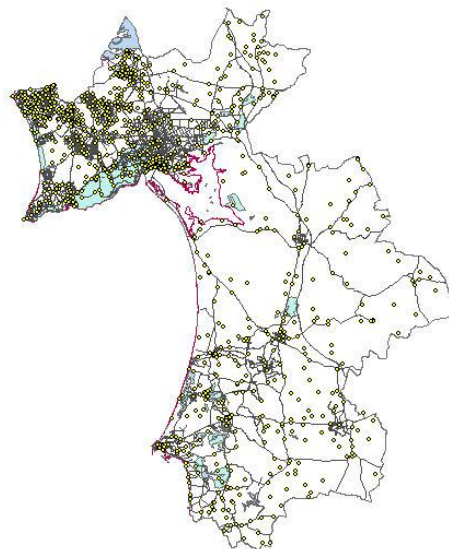


Figura 2– Fogos registados no distrito Setúbal entre 2006 e 2010

Pode-se inferir de acordo com a figura 3, que o centro geométrico, ou a média das coordenadas X e Y dos eventos, localiza-se em pleno parque natural da Arrábida e que a maior concentração de fogos (68%) se encontra a Norte da cidade de Setúbal. Pode-se verificar pela análise da imagem seguinte um efeito direccional forte na elipse padrão. Isto justifica-se pelo facto do centro geométrico se localizar a Norte do distrito e de se registar uma concentração considerável a Sul deste.

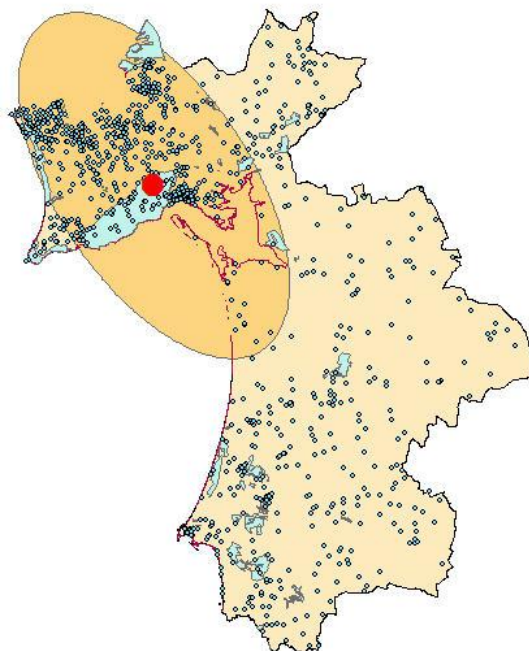


Figura 3– Centro Geométrico e Elipse Padrão (68%)

Na figura 4, pode-se analisar a dispersão dos dados para uma elipse padrão a 95% (amarelo) e para uma elipse padrão a 99% (vermelho)

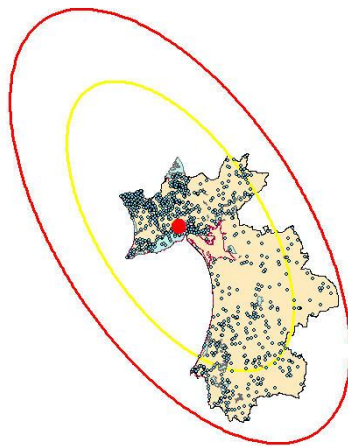


Figura 4– Centro Geométrico e Elipse Padrão (95% e 99%)

2.2 – Análise Exploratória Dados

O objectivo desta etapa é determinar se estamos perante um padrão pontual disperso, aleatório ou aglomerado. Para tal, pode-se utilizar técnicas que explorem as propriedades de 1ª ordem ou técnicas que explorem as propriedades de 2ª ordem.

Os efeitos de 1ª ordem, considerados globais ou de grande escala, correspondem a variações no valor médio do processo, ou seja, variações na sua densidade ao longo de uma superfície contínua. Quando recorremos a técnicas que exploram efeitos de 2ª ordem, também designados por locais ou de pequena escala, estamos a analisar a dependência espacial no processo, nomeadamente a covariâncias entre duas variáveis aleatórias.

2.2.1 – Propriedades de 1ª Ordem (*Kernel Density*)

Nesta fase do projecto, utilizou-se o operador kernel density de forma a identificar hot-spots na área em estudo. Para tal, e após várias tentativas, calculou-se as variações de densidade para uma largura de banda de 1000m, que pareceu ser a mais adequada, obtendo-se os seguintes resultados:

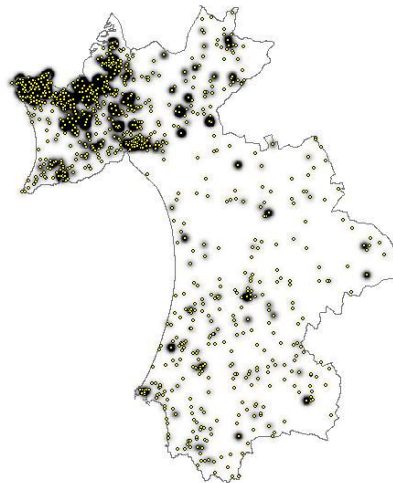


Figura 5– Kernel density para incêndios no distrito de Setúbal

Da análise da figura conclui-se que os maiores aglomerados de eventos se localizam na zona norte do distrito.

2.2.2 – Propriedades de 2ª Ordem (Método do vizinho mais próximo)

Este método procura distâncias entre os eventos do padrão de pontos. É considerado o método mais directo para descrever as propriedades de 2ª ordem, estimando a função de distribuição cumulativa a partir de distâncias entre os eventos da área em análise.

Utilizando a opção *Average Nearest Neighbor*, obteve-se os seguintes resultados, que indicam estar na presença de um padrão de aglomerado:

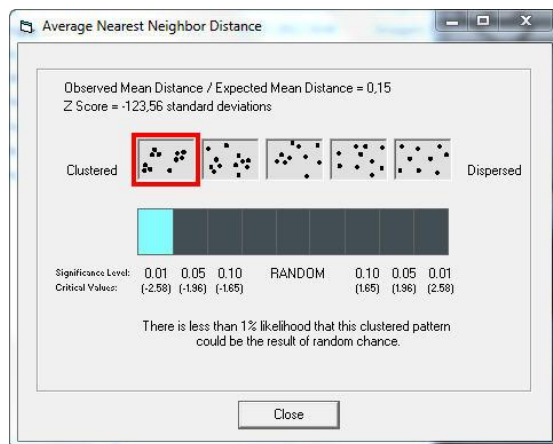


Figura 6– Operador Average Nearest Neighbor Distance

Esta técnica funciona como uma boa indicação da distribuição do padrão pontual, no entanto, é necessário complementar essa análise com funções específicas como a função k.

2.2.3 – Propriedades de 2ª Ordem (Função K)

Esta função complementa a análise anterior e baseia-se na distância entre todos os pontos. Utilizou-se esta função de forma a evitar a limitação das funções G e F, uma vez que estas apenas utilizam as distâncias entre dois eventos. O raio dos círculos utilizado pela função K foi de 1000 m com 10 incrementos, de forma a efectuar uma análise até aos 10 km. Obteve-se os seguintes resultados:

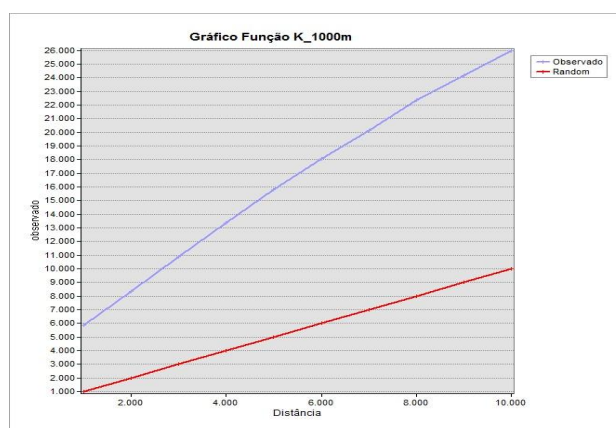


Figura 7– Gráfico da função K

Os valores observados (a azul) são significativamente superiores aos valores esperados (a vermelho) e aumentam consideravelmente com o aumento da distância. Logo, pode-se concluir que o padrão pontual dos fogos florestais e agrícolas é um padrão essencialmente de aglomerados.

2.3 – Modelação dos dados espaciais

A análise estatística dos padrões de dados pontuais necessita de um modelo de referência de modo a verificar a significância dos métodos exploratórios realizados anteriormente. Assim, para se testar a hipótese de aleatoriedade espacial ao conjunto de dados em análise, utilizam-se modelos teóricos.

2.3.1– Complete Spatial Randomness (CSR)

É o modelo teórico mais utilizado e pressupõe duas bases fundamentais. Existe igual probabilidade de ocorrência de um evento em qualquer parte do distrito Setúbal (hipótese 1) e que existe independência na posição dos incêndios (hipótese 2). Existem dois métodos para testar as hipóteses do CSR, mas, neste caso utilizou-se apenas as técnicas do método 2 uma vez que o processo espacial é complexo. A técnica utilizada foi a função K com simulações.

Esta técnica estima a significância dos desvios da distribuição em relação à aleatoriedade, determinando um envelope (limite) superior e inferior. Assim, utilizou-se a função K com 5 simulações aleatórias, obtendo-se os seguintes resultados:

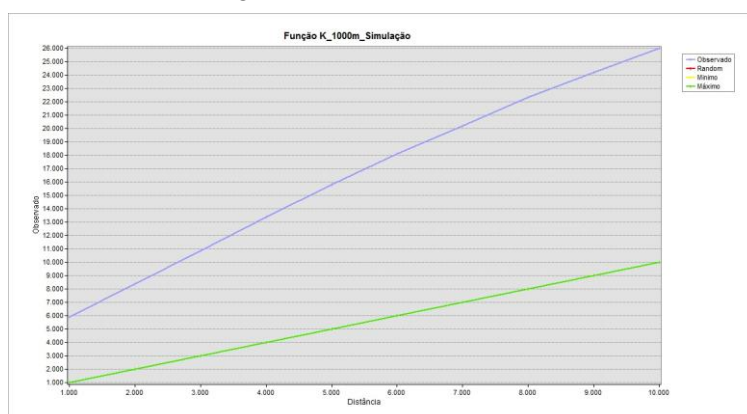


Figura 8– Função K com simulação

Como a linha observada (a azul) está fora dos intervalos de confiança, independentemente da distância considerada, e encontra-se acima dos valores máximos (a verde) pode-se concluir que o padrão pontual de dados em estudo é significativamente aglomerado. Desta forma rejeita-se a hipótese de aleatoriedade, para um nível de confiança de 95%.

2.4 – Análise de Resultados

As três etapas da análise da distribuição espacial do fenómeno pontual de incêndios no distrito de Setúbal durante o período de 2006 até 2010, demonstram que nos encontramos perante um padrão de aglomerado ou cluster, para um intervalo de confiança de 95%. Outra ilação que se pode retirar da análise da figura 5, é sobre as localizações geográficas dos aglomerados do padrão pontual, que se registam maioritariamente na zona norte do distrito de Setúbal.

Nesta fase do projecto, pretende-se integrar essa informação com dados da ocupação de solo, da rede hidrográfica e da rede viária, de forma a se poder extrair conclusões que poderão explicar o fenómeno.

Para analisar a influência da presença de bacias de água com a distribuição dos incêndios, criou-se um buffer de 5km (distância considerada óptima para meios aéreos) à volta de cada bacia e comparou-se com os aglomerados determinados. Na figura 9 (mapa da esquerda), as setas a vermelho indicam os aglomerados que se encontram fora dos círculos de 5km. Facilmente se concluí que a proximidade de bacias de água é inversamente proporcional à existência de aglomerados de incêndios, como era expectável.

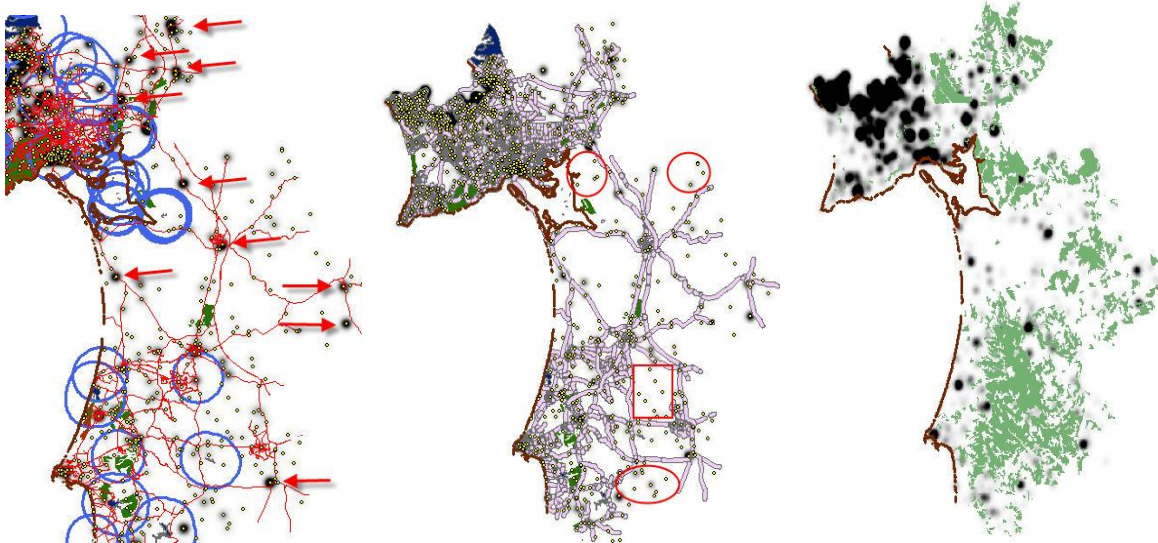


Figura 9 – Estudo da influência da rede hidrográfica, rede viária e ocupação de solo nos incêndios do distrito Setúbal.

Adoptando a mesma metodologia para a rede viária, criou-se um buffer de 500m de diâmetro que corresponde à distância fora de estrada acessível às viaturas autotanque todo-o-terreno dos bombeiros. Analisando a dispersão dos aglomerados de incêndios da figura 9 (mapa central), conclui-se que o factor rede viária não contribui da mesma forma que as bacias de água, uma vez que não se registaram muitos eventos fora do buffer determinado.

Por último comparou-se a localização dos aglomerados de incêndios com as zonas de floresta (folhosas, resinosas, mistas e aberta), zonas de vegetação herbácea natural e vegetação esclerófila. Verificou-se que os aglomerados não coincidem com estas localizações pouco exploradas pelo Homem (a verde no mapa da direita), mas sim com outras áreas dedicadas à agricultura e próximas de grandes aglomerados populacionais, como é exemplo os concelhos localizados a norte do distrito de Setúbal e que se localizam próximos do litoral.

3. Trabalho Parcelar 2

O objectivo deste trabalho parcial é avaliar a autocorrelação espacial do número de fogos em Portugal Continental, tentando definir um modelo de regressão explicativo para a ocorrência desses mesmos fogos. Para tal, definiu-se que se iriam organizar os dados em unidades de área de concelho, por preservar o fenómeno através da utilização de técnicas de agregação e realizar um estudo de autocorrelação a nível nacional, também para o período entre 2006 e 2010.

As fontes de dados utilizadas foram as mesmas do trabalho parcial 1.

3.1 – Visualização de dados do tipo areal

Para visualizar os dados agrupados por áreas, utilizaram-se métodos de corte da variável dependente, para se produzir os mapas temáticos. Assim, considerando o número de eventos por concelho e utilizando as ferramentas de visualização disponibilizadas pelo software *OpenGeoDa*, obteve-se, de forma bastante intuitiva, a distribuição espacial dos dados em dois mapas distintos, figura 10.

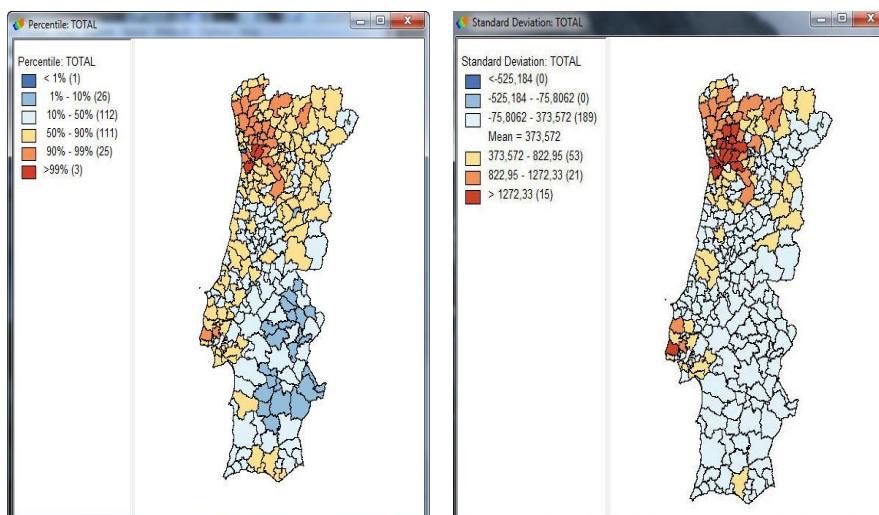


Figura 10 – Mapa por Percentil e Desvio Padrão

Através do mapa de percentis, podemos deduzir que mais de 50 % das ocorrências registadas na série temporal em estudo se localizaram na zona centro/norte do país. Através do mapa do desvio padrão, podemos analisar o comportamento estatístico da variável a 87%, 95% e 99,9%. À semelhança do gráfico dos percentis, também aqui se evidencia que os valores que se afastam consideravelmente da média se localizam na zona noroeste de Portugal.

3.2 – Análise Exploratória Dados

O objectivo desta etapa é medir a estrutura da correlação espacial e estimar a magnitude da autocorrelação entre áreas. Para tal, podemos utilizar técnicas que exploram as propriedades de 1ª ordem ou técnicas que exploram as propriedades de 2ª ordem.

Os efeitos de 1ª ordem, considerados globais ou de grande escala, determinam a variação do valor médio em todos os polígonos em estudo. Quando recorremos a técnicas que exploram efeitos de 2ª ordem, estamos apenas a avaliar a covariância entre um determinado polígono e a sua vizinhança em função de uma distância, de forma a identificar diferentes regimes espaciais que “escapam” aos métodos globais. Para realizar estas análises é necessário definir a matriz de proximidade espacial ou matriz de vizinhança, usando o método de conectividade de 1ª ordem, conforme se pode verificar na figura 11.

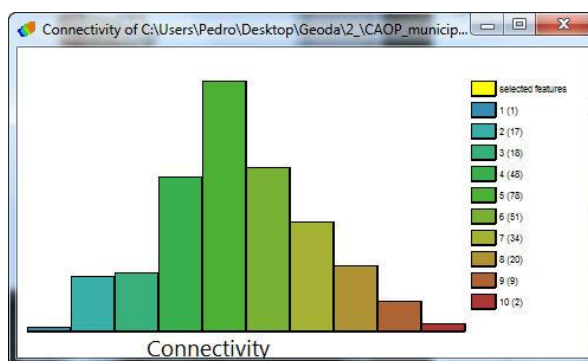


Figura 11 – Matriz de vizinhança

3.2.1 – Propriedades de 1ª Ordem (Índice Moran Global)

Este índice determina a dependência espacial do número de ocorrências de incêndios em todos os concelhos do país. Podemos visualizar esta dispersão através do gráfico de dispersão de Moran da figura 12.

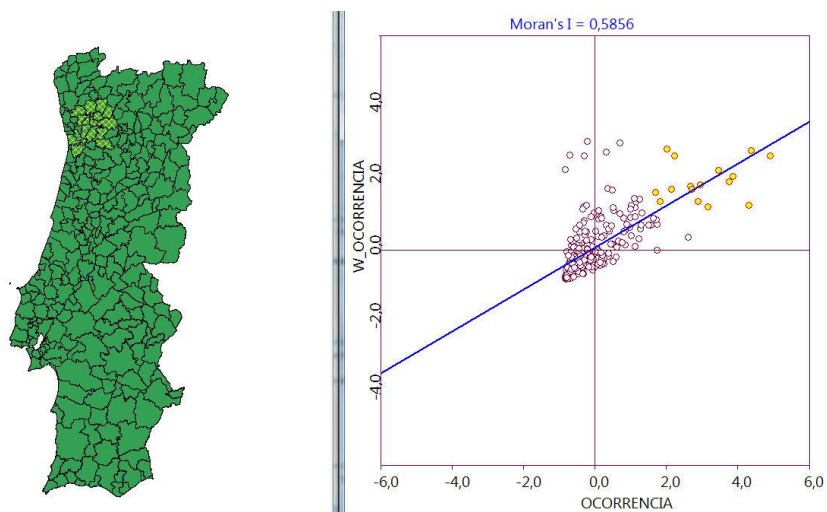


Figura 12 – Índice global de Moran

Valores do índice de Moran próximos de 1, indicam a existência de aglomerados. A autocorrelação espacial positiva indica que os valores altos tendem a estar na vizinhança de valores altos e que valores baixos tendem a estar próximos de valores baixos. Se o coeficiente de correlação for nulo, significa que estamos perante um caso de aleatoriedade e, no caso de ser negativa, que valores altos estão rodeados de valores baixos e vice-versa.

Neste caso, o índice de Moran com o valor positivo de 0,5856 indicia que existe uma distribuição aglomerada de eventos, ou seja de clusters. Estão seleccionadas no gráfico de dispersão as áreas com maiores valores de autocorrelação espacial, ou seja, são zonas muito semelhantes entre si.

Para analisar a validade estatística deste índice de Moran e avaliar se os valores medidos representam ou não uma correlação espacial significativa, utilizou-se o teste da pseudo-distância do software *OpenGeoDa*, que se baseia numa técnica de permutações já referenciada no primeiro trabalho parcial 1. Utilizou-se a opção de 999 permutações para simular uma distribuição normal e compará-la com o índice medido.

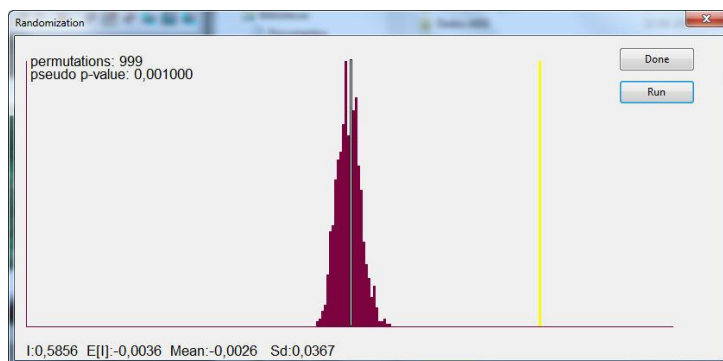


Figura 13 – Teste da pseudo-distância

Na figura 13 pode-se verificar que devemos rejeitar a hipótese nula do modelo CSR e assumir que a variável número de incêndios depende da localização espacial, uma vez que o índice Moran (a amarelo) se encontra num extremo da distribuição simulada, próximo do valor um.

3.2.2 – Propriedades de 2ª Ordem (*Local Indicators of Spatial Autocorrelation -LISA*)

Este tipo de técnicas utilizam-se para se identificar a existência de valores extremos e de aglomerados (*clusters*), que não tenham sido evidenciados pela autocorrelação global.

Com o *Lisa Map* da figura 14, pode-se identificar os polígonos que pertencem a cada quadrante de dispersão de Moran. Desta forma obteve-se uma imagem clara do território nacional subdividido em 2

zonas distintas. A noroeste (vermelho) os polígonos que contribuem significativamente para uma autocorrelação positiva e a sudeste (azul) os que correspondem a uma autocorrelação negativa.

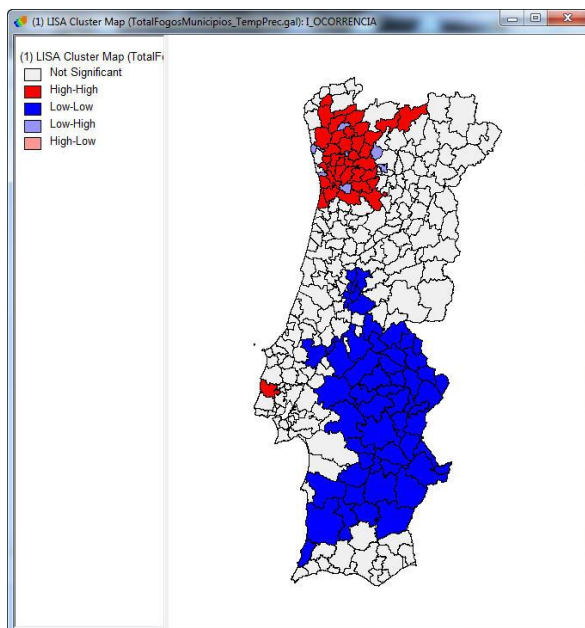


Figura 14 – Lisa Map

O modelo *Geographically Weighted Regression*, aqui utilizado, ajusta um modelo de regressão a cada ponto observado, ponderando todas as observações em função da distância desse ponto aos pontos vizinhos, através do modelo:

$$Y(s) = \beta (s) X + \varepsilon$$

A significância estatística com que cada concelho é classificado, no sentido de contribuir para o valor global de autocorrelação, obtêm-se com o *Lisa significance map*. O mapa da figura 16 foi determinado para 999 permutações ou simulações.

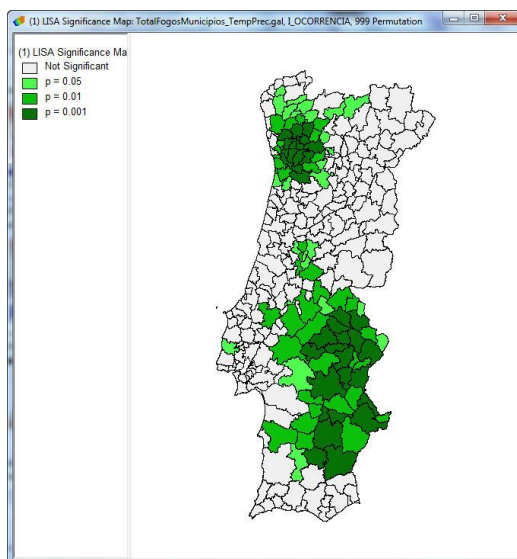


Figura 15 – Mapa de Significância

Nesta imagem pode-se observar a significância estatística com que os polígonos anteriores foram classificados. Assim, a verde-escuro, temos os polígonos classificados com uma certeza de 99.99% e a verde-claro com uma certeza de 95%.

3.3 – Regressão Espacial

Nesta 2ª fase deste trabalho parcial pretende-se utilizar técnicas de regressão espacial para estudar o relacionamento existente entre duas ou mais variáveis. O objectivo é descrever o número de incêndios (variável dependente) em relação a outras variáveis independentes.

3.3.1 – Identificação de variáveis

Para se identificar as variáveis a utilizar no modelo de regressão espacial, recorreu-se ao site do Instituto de Meteorologia, segundo o qual informa que: “índice meteorológico de risco de incêndio *FWI*, desenvolvido pelo Serviço Canadano de Florestas (...), entra em consideração com os valores observados como a temperatura do ar, a humidade relativa, a velocidade do vento e a quantidade de precipitação ocorrida nas últimas 24 horas.”

Assim, identificou-se como variáveis independentes para esta análise, dados médios anuais de precipitação e de temperatura, figura 16. Os dados utilizados para a regressão são relativos a uma série temporal compreendida entre 1981 e 2010, obtidos no site do Instituto Meteorologia [2012], na qual se enquadra as ocorrências de incêndios (2006-2010). Para além destes, incluiu-se também o número de corporações de bombeiros existentes em cada concelho de Portugal Continental do site AFN [2012].



Figura 16 – Fonte de dados de temperaturas e precipitação média.

3.3.2 – Modelos de regressão clássica (OLS)

A qualidade do ajustamento dos modelos de regressão é determinada pela análise dos resíduos. Para se determinar se um modelo de regressão deve ou não ser espacial, é necessário medir a autocorrelação espacial dos resíduos da regressão ou mapeá-los em gráficos.

Após se testar o modelo de regressão não espacial OLS, no software *OpenGeoDa* obteve-se os seguintes resultados da figura 17.

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	54,95679	237,2905	0,2316013	0,8170130
QUARTEISBO	94,47664	13,94038	6,777192	0,0000000
TEMPERATUR	-20,22641	13,61557	-1,485536	0,1385520
PRECIPITAC	5,74427	0,903687	6,356482	0,0000000

Figura 17 – Resultados dos coeficientes das variáveis independentes pela regressão clássica.

O modelo OLS é então:

$$Y = 54,95679 + 94,47664x - 20,22641y + 5,74427z + \epsilon$$

Onde: x = Quartel bombeiros; Y = Temperatura e Z = Precipitação.

Existem vários parâmetros que indicam se um modelo de regressão clássica se ajusta bem aos dados. O parâmetro *Log likelihood*, que deve ser o maior possível, os parâmetros *Akaike info criteriom* (AIC)

e *Schawrtz criteriom* que devem ser preferencialmente baixos e o parâmetro *R-squared* que varia entre 0 e 1, sendo que quanto mais próximo de 1 estiver, melhor é o ajuste do modelo.

Neste ajustamento isto não se verificou conforme se pode analisar na figura 18. Pode-se concluir que o modelo clássico se ajusta mal aos nossos dados ou que as variáveis independentes são pouco explicativas do processo.

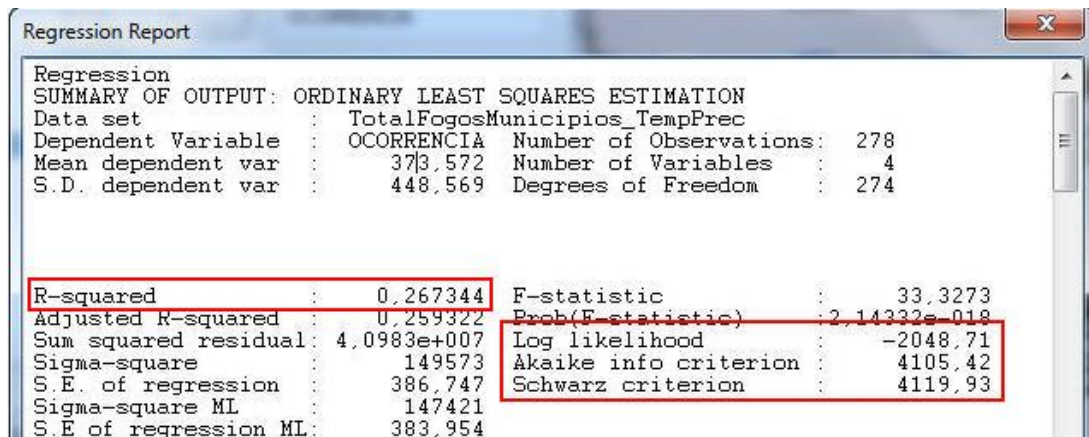


Figura 18 – Relatório da regressão clássica1

3.3.3 – Diagnósticos ao modelos de regressão clássica (OLS)

Existem vários diagnósticos que permitem analisar os resultados da OLS. São exemplo disso os diagnósticos da condição de multicolinearidade, o diagnóstico de heterocedasticidade (3 testes) e o diagnóstico para a dependência espacial (6 testes).

Após se realizarem esses diagnósticos, obtiveram-se os resultados patentes na figura 19.

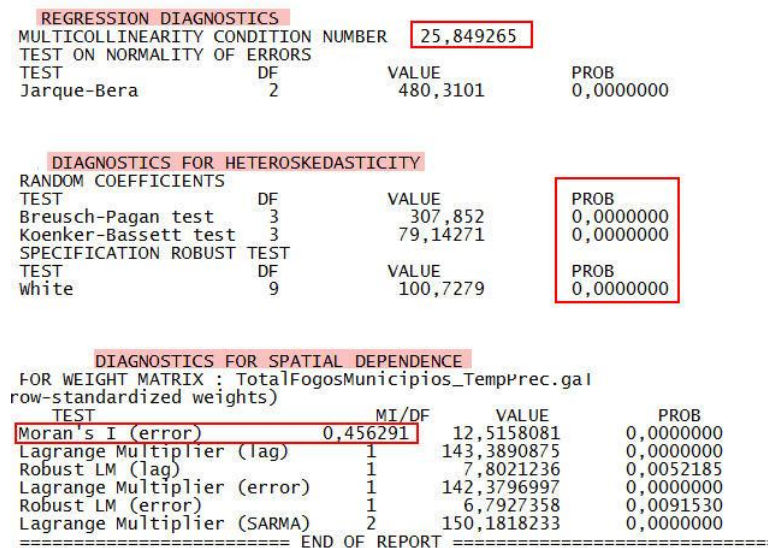


Figura 19 – Relatório da regressão clássica2

Após se analisarem os resultados, chegou-se às seguintes conclusões:

- O facto do número de condição de multicolinearidade ser superior a 20 indica problemas no modelo. O facto de a probabilidade do teste Jarque-Bera ser zero indicia também que não se trata de uma distribuição normal.

- Relativamente ao segundo diagnóstico, a baixa probabilidade dos três testes (zero), evidencia a existência de heterocedasticidade (variância não constante ou forte dispersão dos dados ao longo de uma recta) e que condiciona a validade do modelo OLS.

- Por último, o facto de o índice de Moran do diagnóstico da dependência espacial ser um valor positivo elevado (0,4563), indicia que existe uma autocorrelação espacial nos resíduos.

Outra forma de verificar a existência de uma correlação espacial, conforme já foi referido anteriormente, é recorrendo ao mapeamento dos resíduos. Assim, mapeando os resíduos de duas variáveis independentes pode-se verificar que a variância não é constante conforme se observa na figura 20. Pela análise do histograma, conclui-se que estes não possuem uma distribuição normal.

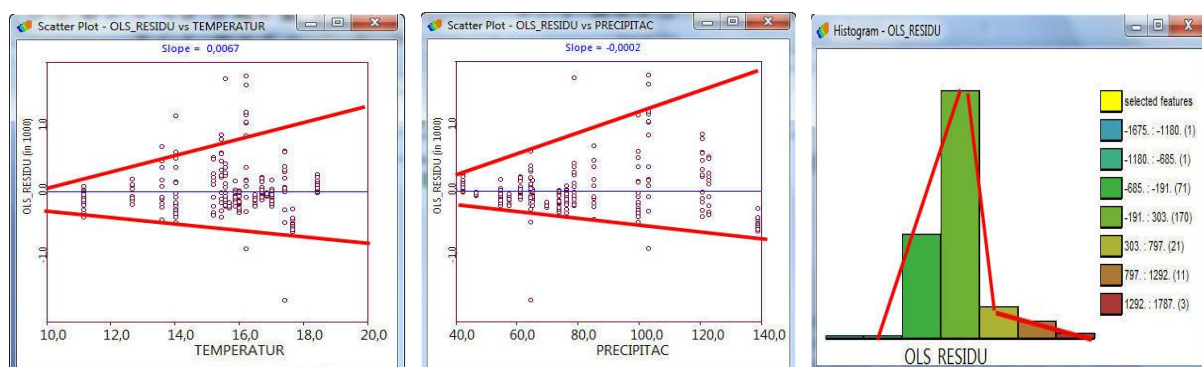


Figura 20 – Mapa de resíduos temperatura, precipitação e histograma de resíduos.

Resumindo, pela análise dos 3 gráficos da figura anterior, conclui-se que existe uma dependência espacial entre as variáveis.

3.3.4 – Modelos de regressão espacial

De seguida, para se estudar a hipótese de que o processo espacial é estacionário e que pode ser captado apenas por um parâmetro, testou-se os modelos globais de regressão espacial, nomeadamente o *Spatial Lag Model* (SLM) e o *Spatial Error Model* (SEM). Enquanto o SLM atribui a autocorrelação espacial à variável dependente Y, o SEM atribui a autocorrelação espacial ao erro. Os resultados são descritos na tabela 1.

Tabela 1 – Resultados dos modelos de autocorrelação clássicos e espaciais.

<i>Modelo</i>	<i>R²</i>	<i>Log Likelihood</i>	<i>AIC</i>	<i>Schwartz Criterion</i>	<i>Coefficiente espacial autocorrelação</i>
<i>OLS</i>	0,27	-2048,71	4105,42	4119,93	-
<i>SLM</i>	0,53	-1998,64	4007,27	4025,41	$\rho = 0,582431$
<i>SEM</i>	0,55	-1996,47	4007,95	4015,46	$\lambda = 0,655113$

Da análise da tabela conclui-se que o melhor modelo é o SEM. Para além de possuir o maior coeficiente de determinação e de *Log Likelihood*, apresenta também um parâmetro *Schwartz* inferior aos restantes e um parâmetro *AIC* inferior ao modelo OLS. Outro indicador é o facto do coeficiente espacial de autocorrelação ($\lambda = 0,655113$) indicar uma autocorrelação positiva bastante significativa nos nossos dados.

3.4 – Análise de Resultados

O facto de os parâmetros de estudo principais possuírem sinais opostos ao esperado, demonstra que as variáveis independentes não modelam bem a dependente. O facto de se utilizar para análise médias da temperatura e precipitação de uma série temporal superior (1981-2010) à série de dados da variável dependente (2006-2010) influencia o modelo de regressão, assim como o facto de se efectuar um modelo de regressão a nível nacional.

4. Trabalho Parcelar 3

O 3º trabalho parcial deste projecto consiste numa análise espacial da distribuição de áreas ardidas no distrito de Setúbal, de forma a identificar factores meteorológicos que poderão estar na sua origem. Para tal, identificou-se inicialmente que essa análise iria recair sobre dados de temperatura e de precipitação. As fontes de dados utilizadas foram obtidos nas estações da rede meteorológica do sistema nacional de informação de recursos hídricos, através do site <http://snirh.pt> [2012]. Foram seleccionadas todas as estações do distrito que forneciam os dados meteorológicos em estudo. Assim, para dados de precipitação foram seleccionadas 13 estações: Alcochete, Canha, Monte da Caparica, Vila Nogueira de Azeitão, Moinhola, São Martinho, Comporta, Grândola, Torrão, Azinheira Barros, São Domingos, Barragem de Campilhas e Cercal do Alentejo. As 11 estações fontes de dados de temperatura seleccionadas foram a de Alcochete, Canha, Monte da Caparica, Moinhola, Comporta, Grândola, Azinheira de Barros, Barragem de Campilhas, Cercal do Alentejo, Albufeira do Pego do Altar e a Barragem do Vale de Gaio.

Como o período em análise ao longo de todo o projeto são 4 anos, recolheram-se os dados disponíveis da rede meteorológica para essas estações para o período de 2006 até 2010. Os parâmetros escolhidos foram a precipitação anual, mensal e temperatura do ar média mensal. Para se obter um valor de referência para cada parâmetro em cada estação, foram determinados os valores médios anuais de precipitação e temperatura e, posteriormente, feita uma média para o período em análise.

Após se modelar superfícies de precipitação e temperatura através do uso de métodos geostatísticos, realizou-se uma análise à rede viária do distrito para identificar percursos óptimos entre as corporações de bombeiros e os locais mais prováveis de incêndio. Produziu-se também mapas de acessibilidade para todos os quartéis de bombeiros com base em distâncias, nomeadamente para os 5 e os 10 km. As fontes de dados utilizadas para esta fase foram a localização das 31 corporações de bombeiros que se obtiveram no site <http://srif.igeo.pt> [2012] e a rede viária, do distrito de setúbal.

4.1 – Análise Exploratória dos dados

Realizando a análise exploratória aos dados meteorológicos de precipitação e temperatura através de um histograma e de estatísticas básicas, obteve-se os seguintes resultados da tabela 2.

Tabela 2 – Resultados estatísticos da análise exploratória de dados.

<i>Parâmetro Meteorológico (2006-2010)</i>	<i>Nº Observações</i>	<i>Mínimo</i>	<i>Máximo</i>	<i>Média</i>	<i>Desvio Padrão</i>
<i>Temperatura média Anual (°C)</i>	11	13.513	17.084	15.848	0.93281
<i>Precipitação Anual(mm)</i>	13	378.27	872.3	518.84	123.26

Para analisar se a distribuição dos dados é normal, utilizou-se a ferramenta *Normal QQPlot* e verificou-se visualmente que o gráfico descreve uma linha diagonal, confirmando esse pressuposto. Para além disso, esse gráfico permite também identificar possíveis outliers, conforme ilustra os gráficos da figura 21, onde os outliers são identificados através de uma seta a vermelho.

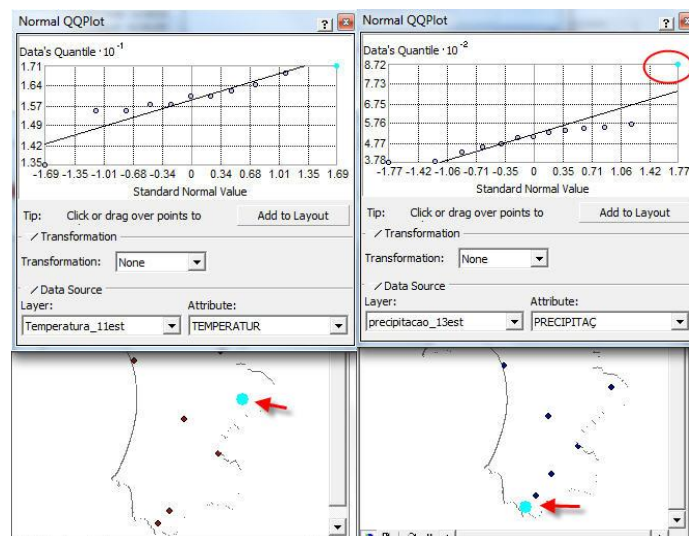


Figura 21 – Outliers na temperatura e precipitação, respectivamente.

De seguida, analisou-se a tendência espacial dos dados em estudo, uma vez que a ausência de autocorrelação é uma das hipóteses fundamentais adjacentes à modelação do variograma. Para tal, utilizou-se a ferramenta *Trend Analysis*, obtendo-se os resultados da figura 22.

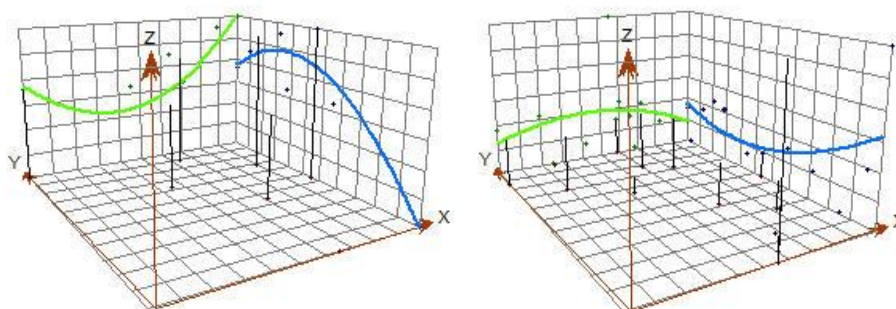


Figura 22 – Tendência espacial da variável temperatura e precipitação, respectivamente.

Pela análise dos gráficos pode-se concluir que a precipitação é ligeiramente superior no Centro do distrito do que nas extremidades Este e Oeste, e que é menor no Norte em comparação com o Sul. Relativamente à temperatura, esta é menor no Centro e aumenta consideravelmente para Este do distrito. Em contrapartida, a temperatura é superior no Centro do distrito que nas extremidades Norte/Sul, sendo consideravelmente inferior a Sul.

4.2 – Construção do semivariograma

O semivariograma que se obtém através do operador *Semivariogram/Covariance*, corresponde à nuvem de pontos em análise e mostra para cada distância “bin”, o quadrado da diferença de cada par de observações. As opções tomadas para o *lag size* e para o *número de lags*, 5000 e 10 respectivamente, visa constranger o variograma para metade da distância máxima que se verificou entre estações do distrito de Setúbal, que neste caso é aproximadamente 108 km.

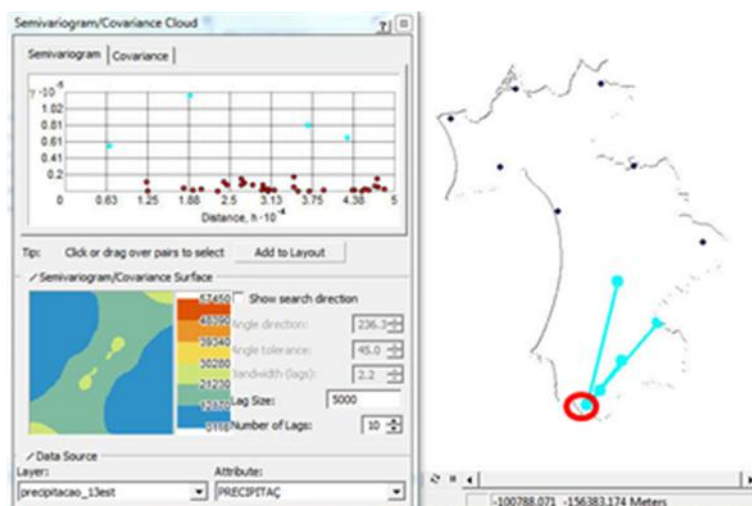


Figura 23 – Semivariograma da precipitação

Analisando o semivariograma da precipitação, figura 23, identifica-se um conjunto de pares de pontos com origem comum, onde a diferença quadrada das observações é muito superior às restantes, dentro do mesmo bin (ou seja, para as mesmas distâncias). Este facto confirma a análise anterior sobre a normalidade dos dados, sugerindo que se está perante um outlier ou um ponto de não estacionaridade.

Verificou-se ainda no semivariograma da temperatura, figura 24, um efeito direcional no sentido NW → SE, conforme se pode verificar pelas assimetrias do mapa que contém a média das diferenças quadradas para todos os pares de pontos da amostra.

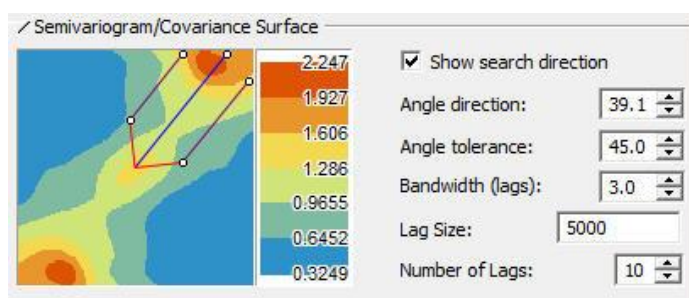


Figura 24 – Efeito direcional do semivariograma da temperatura

4.3 – Kriging

Os métodos de interpolação utilizados na geoestatística diferenciam-se dos tradicionais porque têm em consideração a localização geográfica e a dependência espacial. O Kriging é um dos métodos mais utilizados e caracteriza-se por ser um estimador exacto, local e probabilístico, onde se assume que os pontos da amostra apresentam uma correlação espacial em direcção e distância que irão ajudar a descrever a superfície. Independentemente do tipo de kriging que se utilize, este método consiste sempre em 3 etapas: a análise exploratória dos dados, a modelação do variograma e a interpolação estatística com base nesse variograma.

O kriging utilizado para estimar as superfícies de temperatura e precipitação para o distrito de Setúbal foi o kriging ordinário que se baseia no facto de as médias locais não serem necessariamente próximas das médias da população total. Ou seja, utiliza apenas os pontos vizinhos para realizar a estimação das superfícies. O valor é então estimado com uma média ponderada dos valores de todas as amostras, em que o somatório de todos os pesos utilizado é igual a um. O estimador é definido da seguinte forma:

$$Z^*x_0 = \lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i), \quad \lambda_i = 0 \text{ e } \sum_{i=1}^n \lambda_i = 1$$

Nesta fase modelou-se o variograma com base nos valores obtidos anteriormente e definiu-se qual o modelo teórico a usar para modelar a variação espacial para todo o espaço. Foi escolhido o modelo esférico por ser um dos mais comuns. O modelo teórico esférico mais comum é descrito desta forma:

$$y(h) = \begin{cases} 0, & |h| = 0 \\ C0 + C1 \left[\frac{3}{2} \left(\frac{|h|}{a} \right) - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right], & 0 < |h| \leq a \\ C0 + C1, & |h| > a \end{cases}$$

A figura 25 ilustra todas as opções tomadas para interpolar uma superfície através da opção do kriging ordinário.

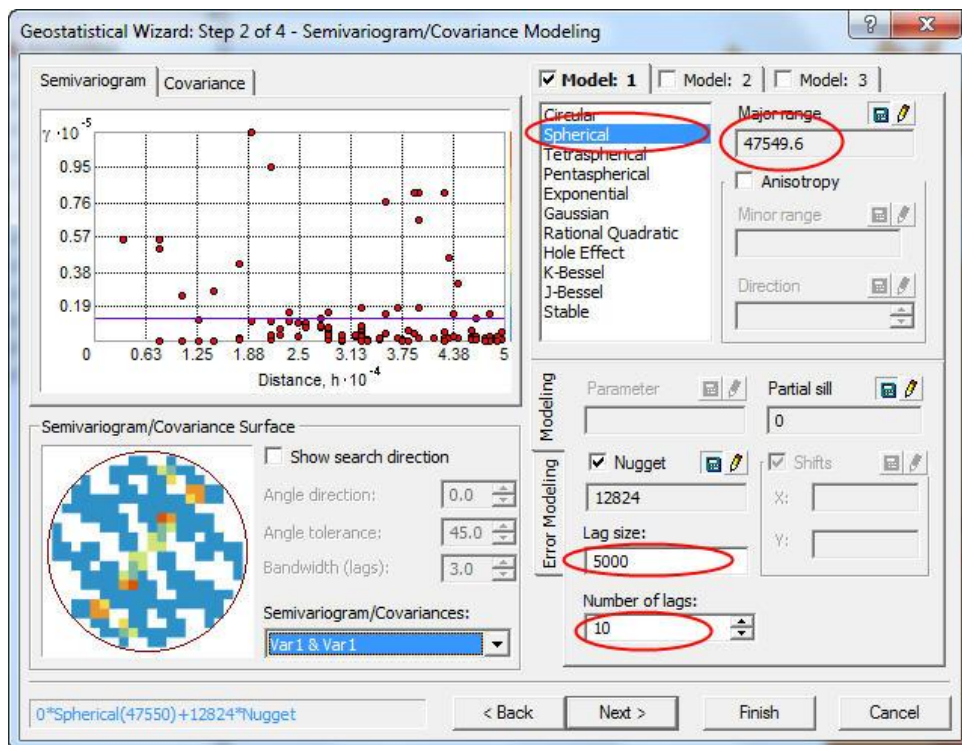


Figura 25 – Configuração do kriging ordinário

As superfícies de precipitação e temperatura interpoladas com esta técnica são as ilustradas na figura 26.

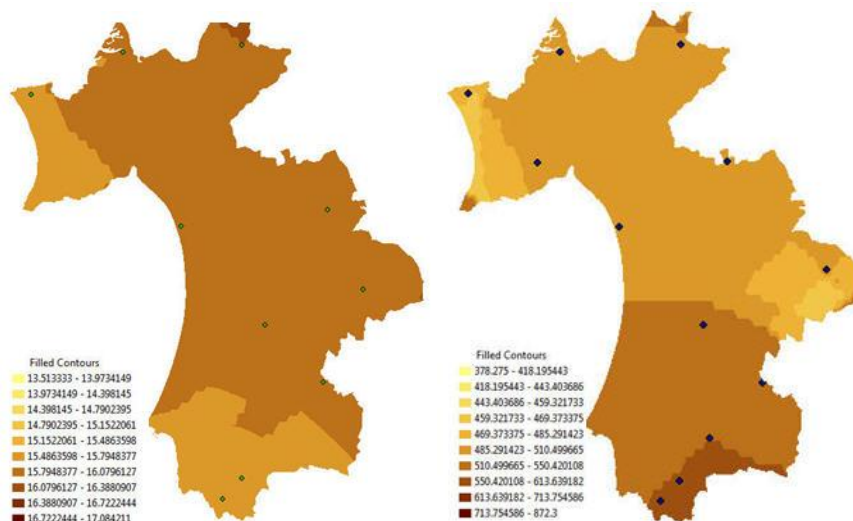


Figura 26 – Kriging ordinário da temperatura e precipitação, respectivamente.

De modo a avaliar a qualidade das superfícies interpoladas, a qualidade do ajustamento necessita de ser quantificada. Neste projecto foi utilizada a técnica de *cross-validation* que consiste na eliminação de uma observação da amostra e usar as restantes para criar uma superfície interpolada. O erro médio quadrático entre a observação e a respectiva estimativa, avalia a qualidade da interpolação. Para a superfície de precipitação eliminou-se a estação de Vila Nogueira de Azeitão e para a de temperatura eliminou-se a estação do Monte da Caparica para fazer a respectiva validação.

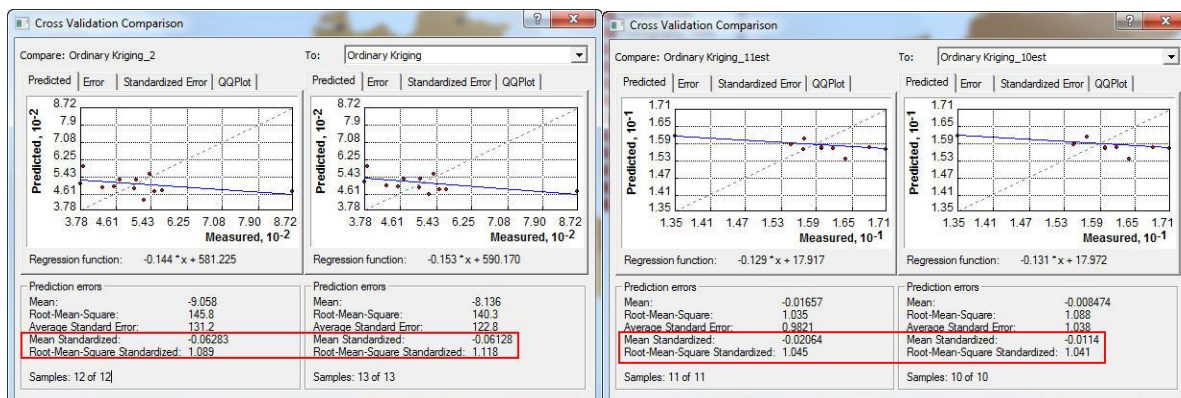


Figura 27 – Cross-validation precipitação e temperatura respectivamente

Os valores obtidos por *cross-validation* para ambas as superfícies interpoladas, figura 27, mostram que os resultados do interpolador são adequados uma vez que os valores do erro médio quadrático são próximos de zero e os valores do erro médio quadrático normalizado é próximo de 1 como se esperava.

Outra forma de analisar a qualidade da interpolação é utilizando o operador *validation*, que consiste em quantificar directamente a diferença entre o valor estimado e o valor observado, na superfície interpolada sem esse ponto de controlo. As figuras 28 e 29 ilustram os resultados da validação para a precipitação e temperatura, respectivamente.

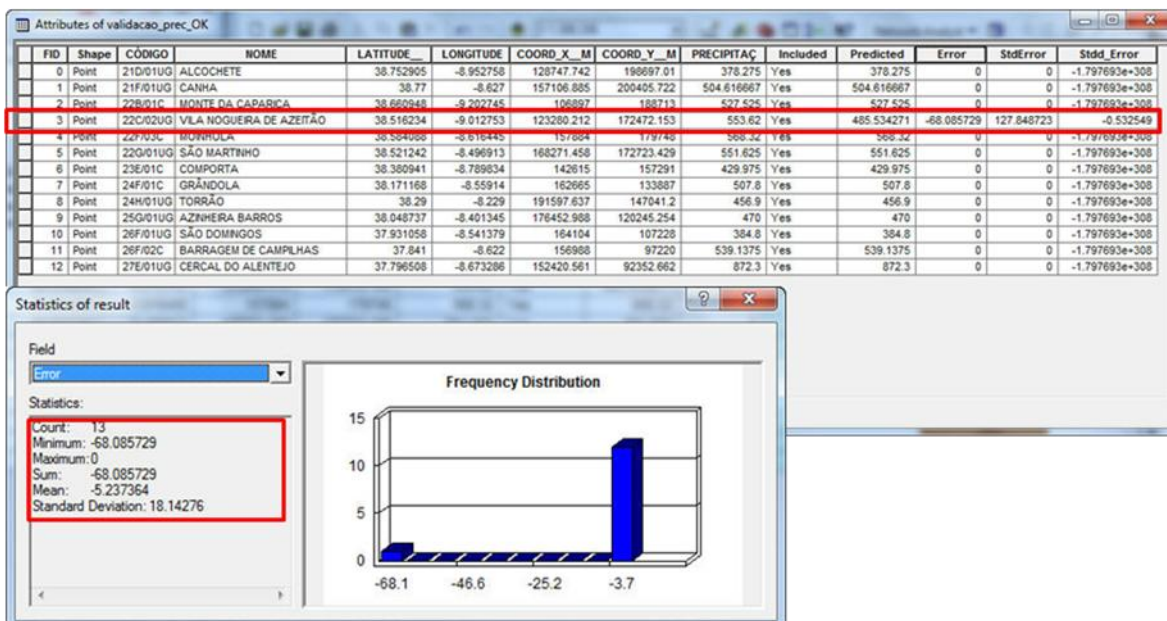


Figura 28 – Validação precipitação.

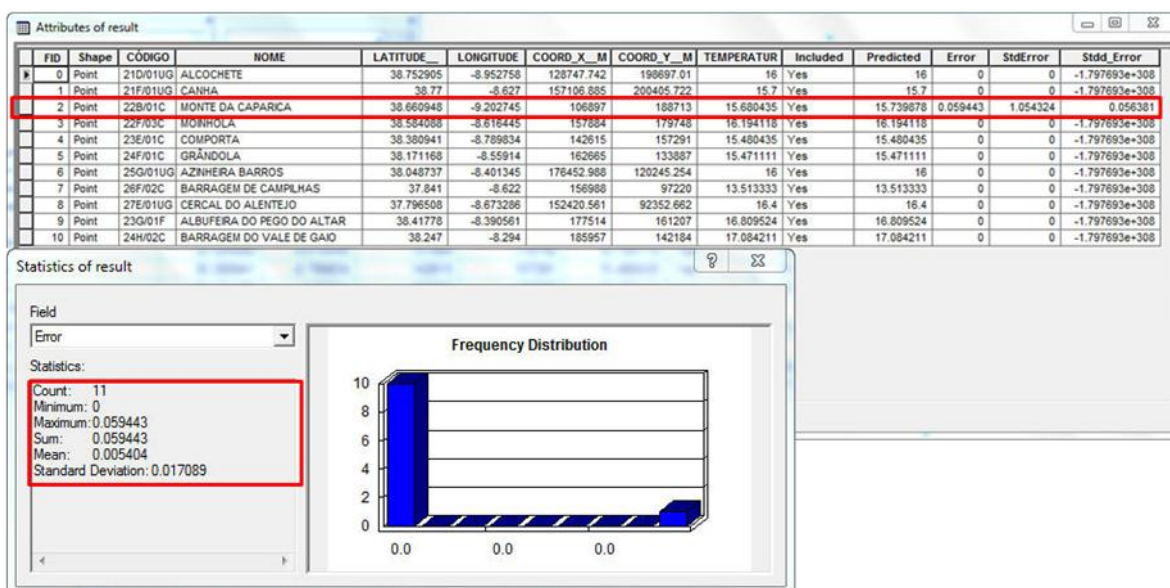


Figura 29 – Validação temperatura.

Pode-se verificar que o desvio padrão obtido para a interpolação da precipitação é consideravelmente superior aos valores obtidos para a interpolação da temperatura.

4.4 – Análise de Resultados

Após se obter superfícies interpoladas com qualidade, resta compará-las com as zonas de maior risco identificadas no trabalho parcial I do projecto, de forma a se retirar ilações sobre a importância destes factores meteorológicos na localização dos focos de incêndios no distrito de Setúbal. Para tal, apresenta-se de seguida duas figuras com as superfícies de precipitação e temperatura interpoladas e as respectivas localizações dos principais focos de incêndio no distrito.

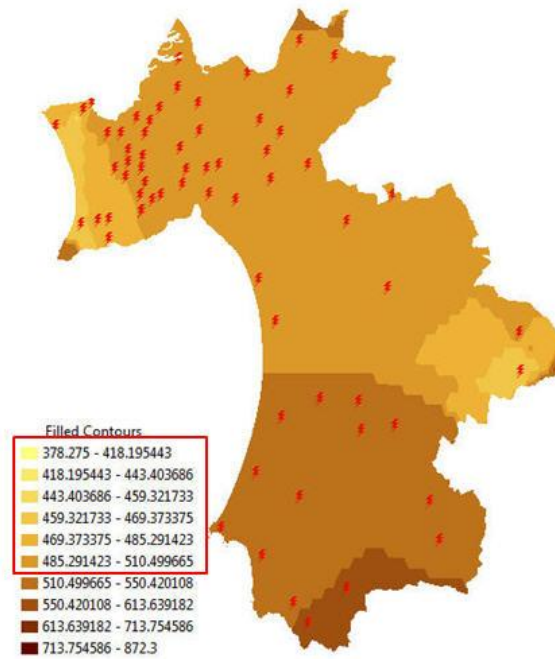


Figura 30 – Estudo da influência da precipitação nos incêndios do distrito Setúbal.

Na figura 30, pode-se deduzir que a grande maioria dos incêndios registados no distrito (na zona centro e norte), localizam-se em zonas com menor precipitação média anual, com valores que oscilam entre os 378.3 até 510.5mm.

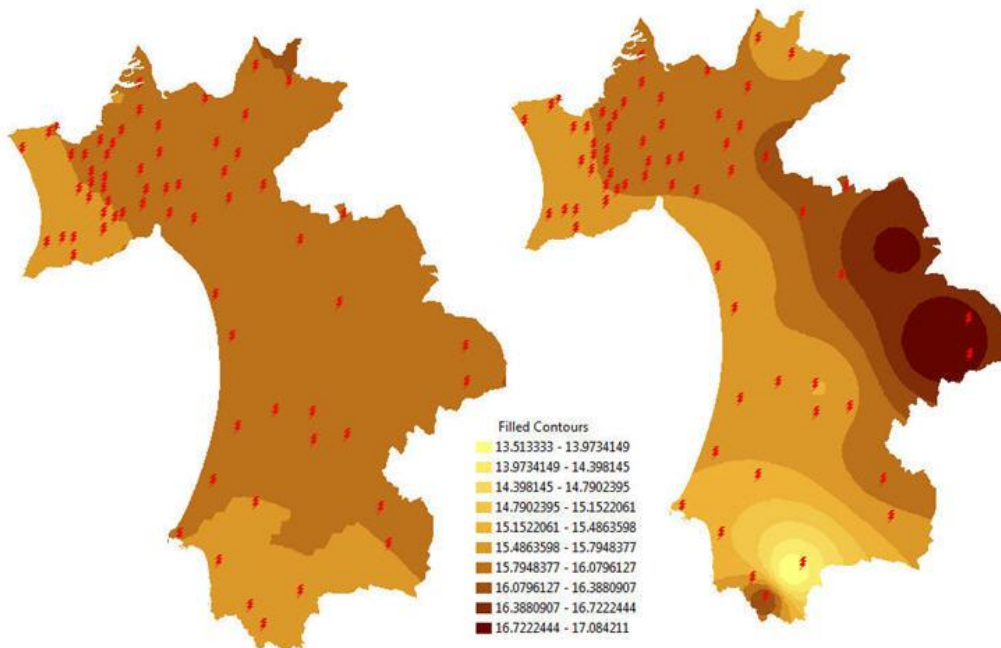


Figura 31 – Estudo da influência da temperatura nos incêndios do distrito Setúbal. Mapa interpolado com o operador kriging ordinário e IDW, respectivamente.

Como se pode observar no mapa da esquerda da figura 31, o distrito subdivide-se em 3 grandes zonas térmicas, registando-se a maior parte dos incêndios na faixa intermediária de temperaturas que oscila entre os 17.79 e os 16.08 °C. Não se pode então estabelecer uma relação directa entre as temperaturas médias anuais e a existência dos focos de incêndio.

Como a interpolação fornece apenas uma estimativa de um valor baseado noutros valores conhecidos, consoante o estimador usado assim é o resultado final obtido. Assim, e de forma a estudar diferenças entre estimadores probabilísticos e determinísticos locais, calculou-se também um mapa interpolado com o operador *Inverse Distance Weight (IDW)*. O mapa da direita foi interpolado a partir do conceito de proximidade e não considera a autocorrelação espacial dos valores da amostra. O modelo aplicado neste interpolador é o seguinte:

$$Z(x) = \frac{\sum_i w_i z_i}{\sum_i w_i}, \quad w_i = 1/d_i^2$$

A análise baseada neste mapa já evidencia que a grande maioria dos focos de incêndio (mais de 95%) se localiza numa faixa térmica inferior aos 16.08 °C.

O facto de o mapa interpolado com o kriging ordinário ser muito homogéneo deve-se ao facto de terem sido utilizadas poucas estações no processo de interpolação (apenas 11) e de se registar uma pequena variabilidade entre elas. Como o mapa interpolado com o *IDW* não considera a variabilidade da amostra, produz um resultado que depende exclusivamente dos valores da amostra mais próximos do ponto a estimar, produzindo esse resultado menos homogéneo.

4.5 – Análise de Redes

Nesta 2ª fase do trabalho parcelar 3 procurou-se responder a duas solicitações que envolvem directamente algoritmos de análise de redes. Primeiro, pretendia-se identificar percursos óptimos entre locais das corporações de bombeiros do distrito de Setúbal e as áreas de incêndio de maior risco identificadas no trabalho parcelar 1. Depois, pretendia-se criar mapas de acessibilidades de distância ou tempo para cada quartel de forma a facilitar posteriores acções de planeamento.

A metodologia aplicada para resolver este problema foi a seguinte. Numa primeira fase, criou-se uma *Network Dataset* com base na rede viária do distrito de Setúbal, definindo-se para a impedância da rede a distância. Esta impedância representa o custo do movimento na rede em análises futuras. Depois criou-se uma *shapefile* de trabalho que possui a rede criada anteriormente, os 31 quartéis de bombeiros e os 66 focos de incêndios de maior risco existentes no distrito de Setúbal, figura 32.

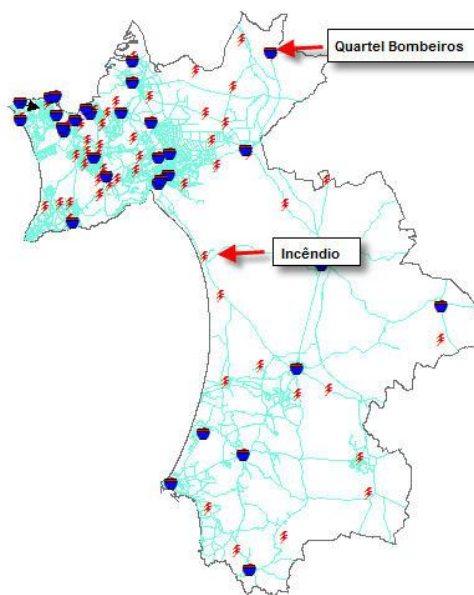


Figura 32 – Rede viária, quartéis bombeiros e focos de incêndio do distrito de Setúbal.

De seguida, para responder à primeira questão, utilizou-se o algoritmo *Closest facility*. Este algoritmo consiste em identificar as *facility*, que neste caso serão os quartéis de bombeiros, que se encontram mais próximos em termos de distância dos *eventos*, que no nosso caso serão os focos de incêndios mais prováveis. Assim, obteve-se o seguinte mapa com as rotas que ligam as áreas de maior risco aos aquartelamentos.

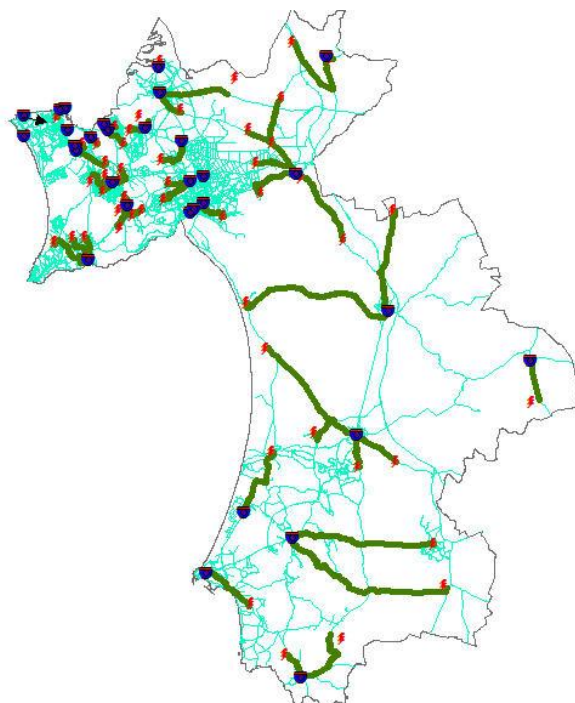


Figura 33 – Rotas com percursos óptimos entre incêndios e quartéis de bombeiros.

Da análise da figura 33, conclui-se que as distâncias médias percorridas pelos bombeiros para aceder aos locais de incêndios mais prováveis nas zonas centro e sul do distrito, são consideravelmente superiores às distâncias percorridas na zona norte. Este facto não é preocupante uma vez que, como

descrito no trabalho parcial 1, a grande concentração de focos de incêndio reside no litoral norte do distrito.

Por último, para se gerar os mapas de acessibilidades de distâncias a partir da localização de cada corporação, utilizou-se o algoritmo Service Area. Este algoritmo define polígonos ou buffers ao longo da rede viária que podem ser alcançados a partir de um ponto para um determinado custo de trajecto, que normalmente se define como a distância ou o tempo. Para este caso, de forma a facilitar acções de planeamento e coordenação de intervenção a incêndios não previstos, criaram-se polígonos para distâncias de 5 e 10 km para todas as corporações. A figura 34 representa mapas de acessibilidade para as duas zonas mais fustigadas pelos incêndios florestais do distrito de Setúbal.

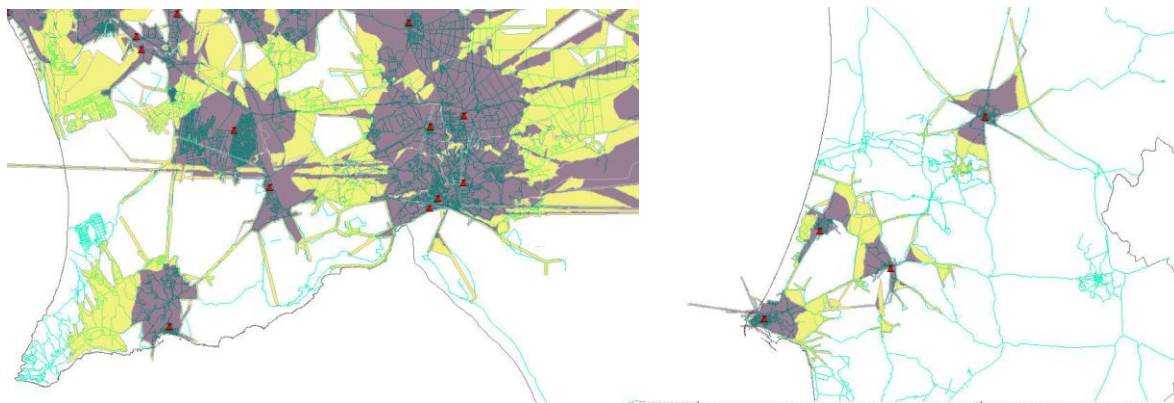


Figura 34 – Mapas de acessibilidade. (5 km a cinzento e 10 km a amarelo)

5. Conclusões

A forma de execução deste projecto foi fundamental para ajudar a uma melhor compreensão e destreza dos conceitos ministrados durante as aulas teóricas e práticas. O facto de este ter sido faseado ao longo do semestre, de forma a acompanhar as aulas ministradas, levou a que tivessem sido tomadas opções numa perspectiva individual e não num conceito de projecto global. Refiro-me ao facto de durante o trabalho parcial 2 ter sido optado por realizar uma regressão espacial a nível nacional em vez de distrito como seria mais adequado. Outra situação que poderia ter sido alterada foi o período temporal escolhido. O facto de se ter seleccionado um período de 4 anos para o trabalho parcial 1, condicionou todos os restantes e conduziu a resultados que poderão não ser tão explícitos comparativamente com análises de 1 ano. Para além disso, os dados meteorológicos disponibilizados não são completos para o período em análise neste projecto, adulterando-se os valores usados nas interpolações com a utilização de valores médios em vez de valores exactos medidos.

Uma das grandes dificuldades sentidas durante a execução deste trabalho foi principalmente a interpretação dos resultados obtidos, nomeadamente no trabalho parcelar 2. Embora as fichas práticas sejam um bom tutorial para acompanhar as aulas, muitas vezes senti ser insuficiente a explicação fornecida por estas para os resultados obtidos.

Referências Bibliográficas

Catita, C. (2011), Apontamentos das Aulas de Análise Espacial de Informação Geográfica, texto não publicado.

Fontes de dados

Dados incêndios florestais totais por concelho (2012): Acedido em Abril de 2012, em <http://www.afn.min-agricultura.pt/portal/dudf/estatisticas>

Dados temperatura e precipitação (2012): Acedido em Abril de 2012, em

<http://www.meteo.pt/pt/oclima/normais.clima/1981-2010>

Dados corporações bombeiros (2012): Acedido em Abril de 2012, em

http://scrif.igeo.pt/ASP/cb_f.asp

Dados da rede de estações meteorológicas (2012): Acedido em Junho de 2012, em

<http://snirh.pt/index.php?idMain=2&idItem=1>

Dados da rede viária mundial (2012): Acedido em Junho de 2012, em

http://downloads.cloudmade.com/europe/southern_europe/portugal/

Software utilizado

ARCGIS Versão 9.3 [ESRI Inc., 1999-2008]

OpenGeoDa Versão 0.9.9.8 (GeoDa Center for Geospatial Analysis and Computation, 2011).