

1. Introdução

1.1 – Enquadramento geral

A região de estudo neste trabalho foi definida em função da área total de abrangência do Comando Metropolitano da Polícia de Segurança Pública de Lisboa. Desta forma, o raio de intervenção da PSP compreende os concelhos da região da grande Lisboa indicados na figura 1.1.

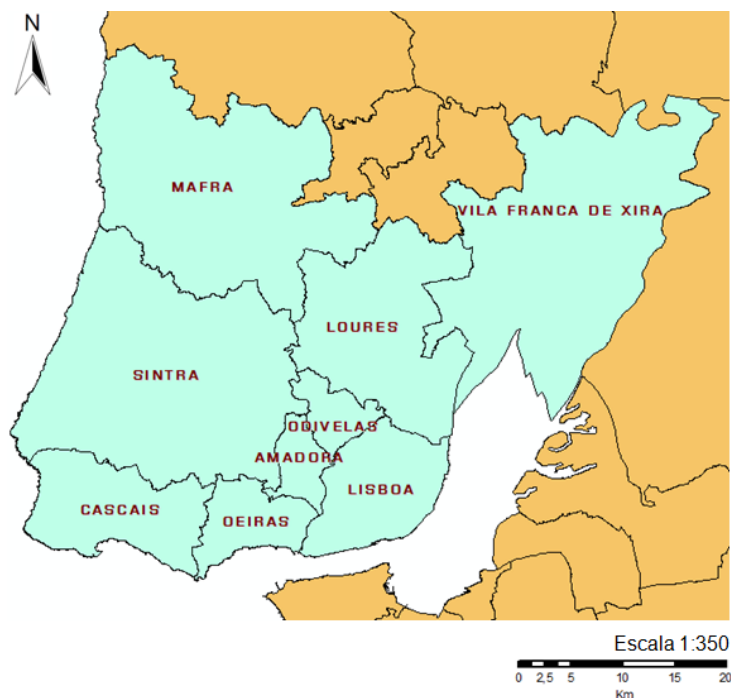


Figura 1.1 – Mapa da área de estudo (concelhos da grande Lisboa)

1.2 – Estrutura do trabalho

De seguida é apresentada uma breve descrição de cada capítulo presente neste trabalho.

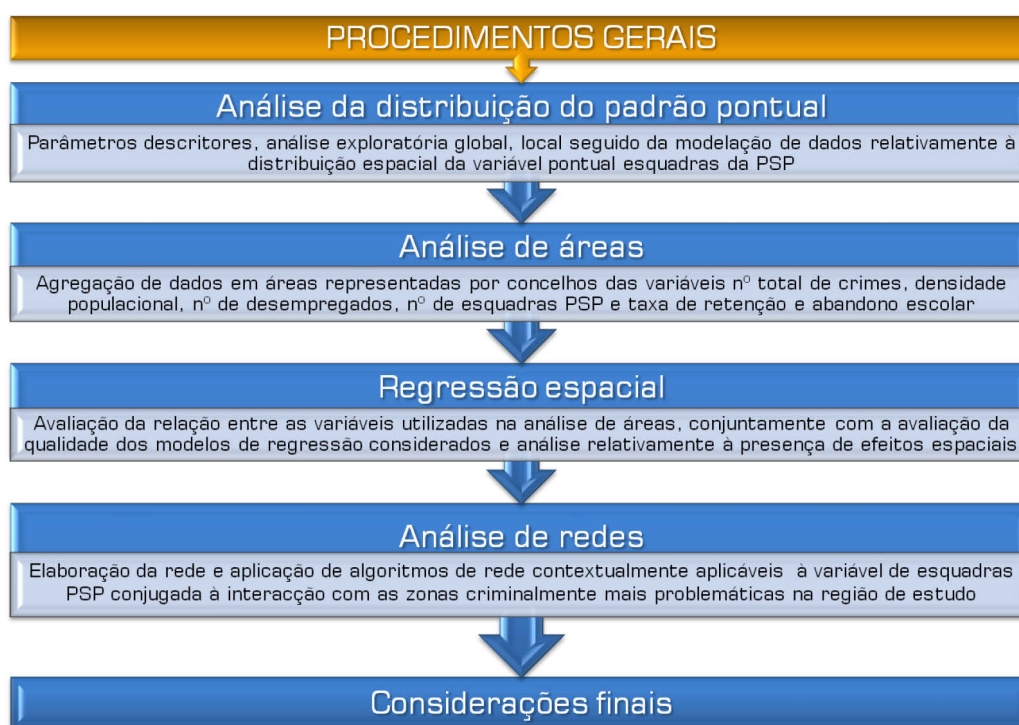


Figura 1.2 – Descrição do procedimento geral em cada capítulo

A acabar esta breve introdução, são expressas na figura seguinte, quais as entidades responsáveis pelo fornecimento dos dados utilizados neste projecto.



Figura 1.3 – Fontes utilizadas na aquisição de dados em cada capítulo deste trabalho (aquisição efectuada no período Abril - Junho 2012)

2. Análise da distribuição do padrão pontual

2.1 – Visualização dos dados espaciais

Parâmetros estatísticos básicos de caracterização (descrição) dos dados espaciais de forma a permitir uma percepção de como é que os pontos se distribuem no espaço, (i.e. variação e orientação espacial) possibilitando a identificação de padrões espaciais de interesse.

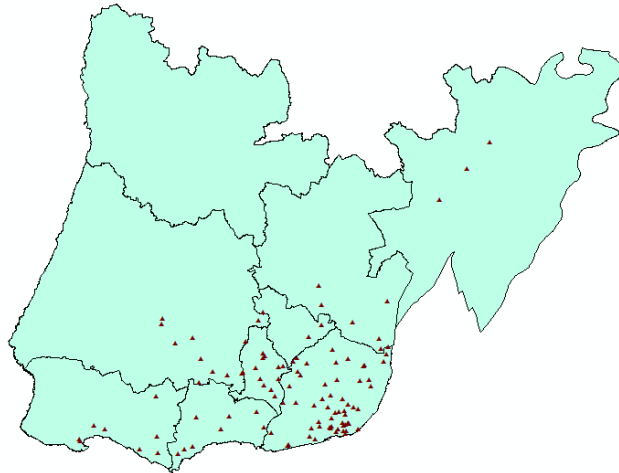


Figura 2.1 – Mapa com as localizações dos eventos (esquadras PSP) na área de estudo

2.1.1 – Descritores de padrões de pontos

2.1.1.1 - Frequência (o nº de ocorrências na área de estudo)

$n = 107$ eventos (esquadras PSP)

2.1.1.2 - Densidade (λ)

$$\lambda = \text{Frequência} / \text{Área}_{\text{grande Lisboa}} \quad (\Leftrightarrow) \quad \lambda = \frac{n}{a} = 0,078 \text{ n/Km}^2$$

2.1.1.3 - Centrografia

Dispersão espacial ou distância padrão

Observa-se que a elipse de distância padrão é relativamente pronunciada, manifestando uma configuração direccional ao longo do eixo próximo de NE-SO (ENE-OSO). Para melhor estudo da tendência direccional e da sua propagação ao longo da área de estudo, realizou-se uma interpolação de análise de tendências (figura 2.2) adquirindo-se a mesma tendência para quase toda a área, excepto no concelho de Vila Franca de Xira (direcção NE-SO), o que se traduz na presença de efeitos espaciais (dependência espacial) na área de estudo.

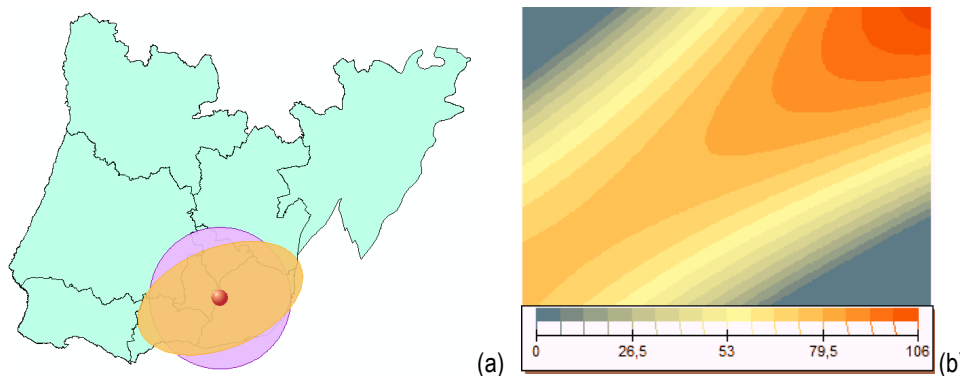


Figura 2.2 – (a) Dispersão espacial. (b) Análise de tendências espaciais (interpolação da superfície com equação polinomial de 2º grau) da distribuição dos eventos

Correspondendo o raio de abrangência do círculo e elipse padrão a uma concentração de 68% dos pontos, observa-se um aglomerado de pontos sobretudo nos concelhos de Lisboa, Amadora seguidos de Oeiras e Cascais.

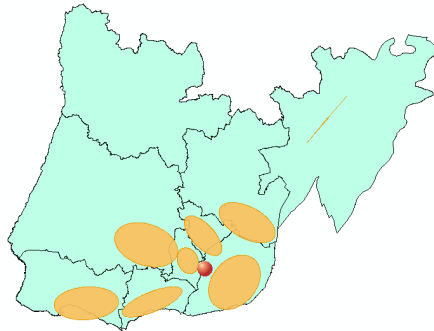


Figura 2.3 – Dispersão espacial (elipses de distância padrão) por concelho

A efectuar uma análise mais detalhada da dispersão (por concelho), na figura anterior observa-se adicionalmente uma dispersão pontual com orientação paralela ao longo da costa em Cascais e Oeiras.

Centro geométrico do padrão de pontos (ou media padrão)

Na figura anterior observa-se que o centro geométrico (indicador da tendência central da distribuição) dos pontos se localiza na fronteira entre o concelho de Lisboa e da Amadora, local correspondente à maior concentração de pontos.

2.2 – Análise exploratória de dados

Neste ponto procederam-se a técnicas de detecção de padrões de pontos, globais e locais, que permitem quantificar e medir a distribuição espacial do padrão pontual em análise. Para tal foram analisadas, respectivamente, as propriedades de 1ª e 2ª ordem da variável Esquadras PSP.

2.2.1 - Técnicas de análise Exploratória de 1ªOrdem (Globais)

2.2.1.1 - Método de contagem por Quadrantes (Censos)

Para este método foi definida uma rede de quadrantes 6x16, com uma média=1,1146 ($\mu=N^\circ$ pontos/ N° Quadrantes) como sendo a que melhor descreve o padrão pontual na área de estudo, na qual cada quadrante correspondente aproximadamente a 25 km².

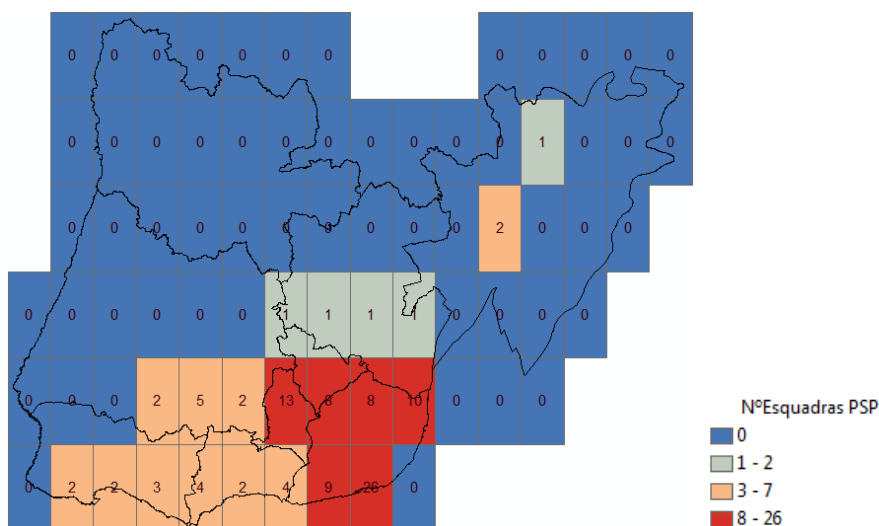


Figura 2.4 – Mapa do método de contagem por quadrantes (por Censos) do padrão pontual (classificação agrupada por intervalo geométrico)

Verifica-se facilmente pela figura anterior, a presença de um padrão fortemente aglomerado centrado no concelho de Lisboa, no qual os concelhos vizinhos (Oeiras, Sintra, Cascais e Odivelas) correspondem a zonas de transição para um padrão disperso, observando-se igualmente, que os quadrantes correspondentes a zonas de transição situam-se maioritariamente ao longo da costa litoral. É possível constatar que nos concelhos de Lisboa, Amadora e Odivelas (parcialmente), localizam-se mais de dois terços (74 esquadras) das 107 existentes em toda a área da grande Lisboa (Área Grande Lisboa $\approx 1376 \text{ km}^2$).

2.2.1.2 - Estimador de densidade (*Kernel density estimation*)

Este método permite verificar com melhor detalhe, parte do que já foi visível notar no método de contagem por quadrantes. A melhor representação da densidade do padrão pontual foi obtida com uma largura de banda de 2500 m, observável na figura seguinte. Repara-se que a densidade (Nº eventos/Unidade de Área) da variável esquadras PSP, é nitidamente alta no concelho de Lisboa, seguida do concelho da Amadora. Relativamente a uma densidade média-alta de esquadras, constata-se *hot-spots* situados na costa litoral dos concelhos de Oeiras e Cascais (já indiciados no método contagem por quadrantes) e em menor extensão no concelho de Sintra. Observa-se que a densidade não é constante na área de estudo, graças à existência de variações locais de intensidade (variabilidade espacial), sendo este processo consequentemente não estacionário de 1ª ordem (tendências espaciais presentes na região de estudo).

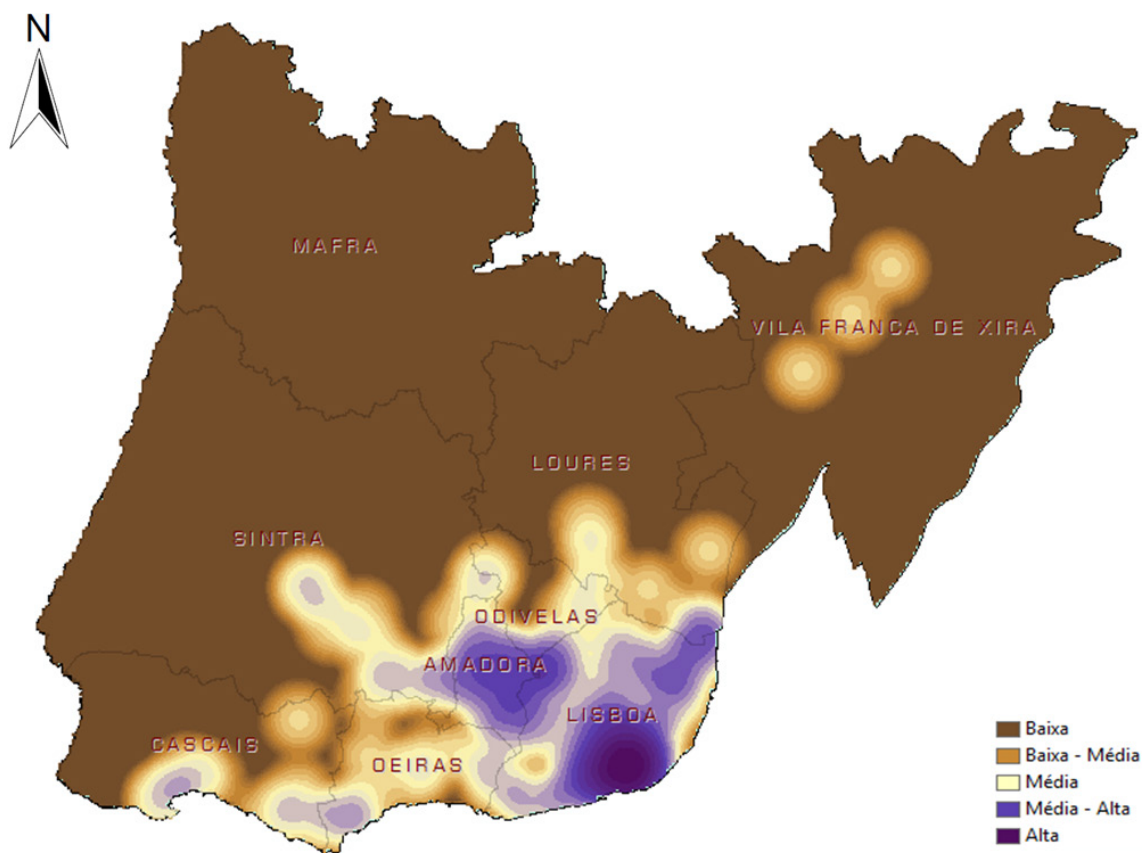


Figura 2.5 – Mapa do estimador de densidade *Kernel* dos eventos (Esquadras) com largura de banda de 2500 metros

2.2.2 - Técnicas de análise Exploratória de 2ª Ordem (Locais)

Foram utilizados os seguintes métodos baseados na medição de distâncias entre pontos (vizinhos mais próximos) para descrever as propriedades de 2ª ordem:

2.2.2.1 - Função G (Vizinho mais próximo)

A função G (avalia a significância dos desvios) baseia-se na distância medida entre cada ponto (evento) ao seu vizinho mais próximo, quantificando a forma como os eventos (esquadras) estão distribuídos espacialmente na região da grande Lisboa.

A distância máxima utilizada na função G (4000 m) foi obtida pela medição da maior distância entre 2 eventos vizinhos (aproximadamente 3800 m).

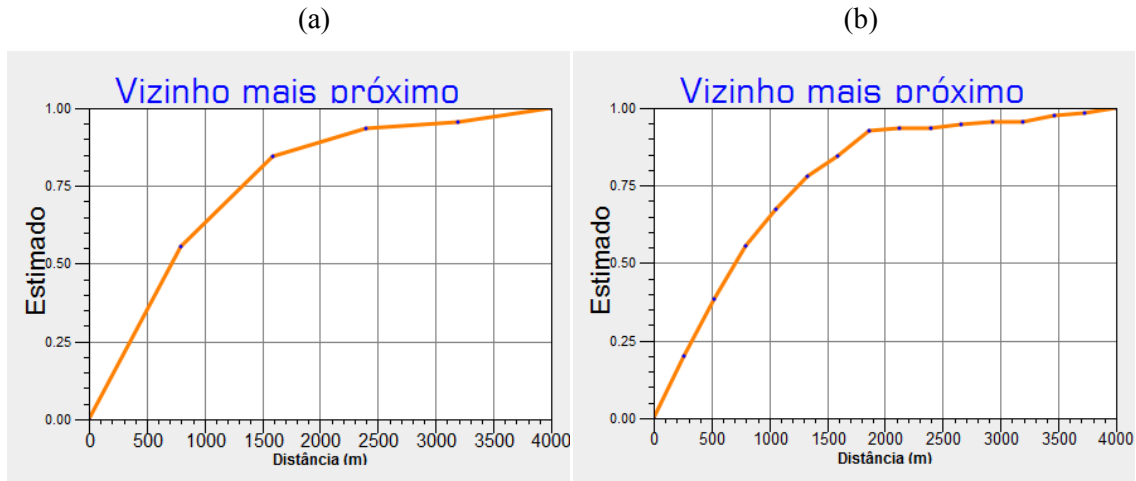


Figura 2.6 – (a) Gráfico da função G (vizinho mais próximo) com intervalos de 5 metros. (b) Gráfico da função G (vizinho mais próximo) com intervalos de 15 metros

É perceptível na figura 2.6, que aproximadamente até aos 1800 m (com intervalos de 15 m), a função aumenta rapidamente, o que aponta para a presença de eventos aglomerados nestas distâncias, especialmente até aos 1300 m de distância (distâncias para as quais o ângulo descrito pela função é superior a 45°).

Para distâncias no intervalo [1800; 4000] metros, a função G apresenta um ângulo bastante inferior a 45°, indicador da presença de eventos bastante dispersos neste intervalo de distâncias.

2.2.2.2 - Função L

A função L avalia todas as distâncias evento-evento (enquanto a função G que utiliza apenas a distância mínima entre 2 eventos). Como tal, a função L foi calculada para uma distância até 50000 metros (distância máxima medida entre os 2 eventos mais afastados é de 47000 m).

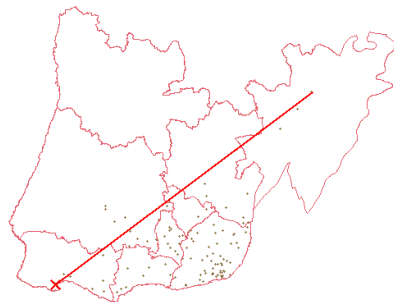


Figura 2.7 – Medição da distância dos 2 eventos mais afastados do padrão pontual

Na figura seguinte observa-se que os valores estimados ($\hat{L}(h)$) são positivos para distâncias até cerca de 33000 m, o que demonstra a presença de um padrão pontual aglomerado até aos 33000 m, sendo esse aglomerado máximo para distâncias perto dos 11000 m (extremo positivo).

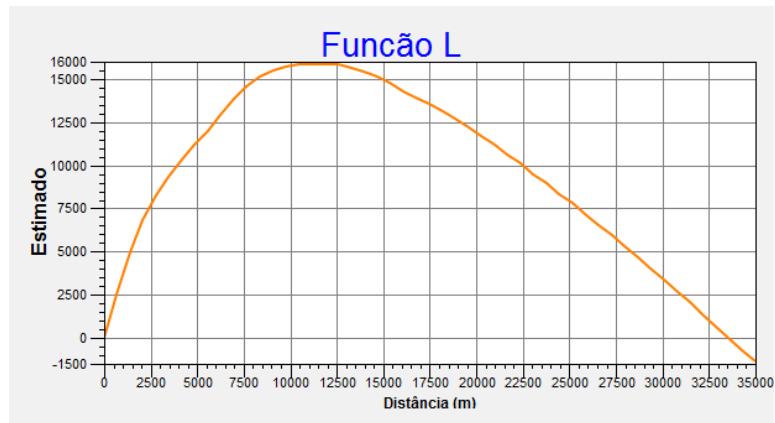


Figura 2.8 – Gráfico da função L para o padrão pontual Esquadras PSP

Para distâncias em redor dos 33000 m, onde os valores estimados são zero, o padrão pontual presente é aleatório. Após esta distância, os valores estimados são negativos e caminham para um extremo negativo, indicando existir uma ordenação regular (caminha para um extremo negativo) na distribuição espacial dos eventos, cuja regularidade pontual aumenta com o aumento da distância.

2.3 – Modelação dos dados espaciais

Neste subcapítulo procedeu-se à validação da significância estatística das técnicas de detecção de padrões de pontos (globais e locais) utilizadas na análise exploratória de dados, recorrendo a testes de significância estatística. Para tal, foi utilizado o modelo de referência CSR (*Complete Spatial Randomness*) que gera um intervalo de confiança através de simulações (envelopes superior e inferior) para obter uma base de comparação entre o comportamento de um processo aleatório e a distribuição de eventos medidos.

2.3.1 - Método de contagem por Quadrantes (censos)

2.3.1.1 - Interpretação de Variance to Mean Ratio (VMR)

Pelos resultados estatísticos obtidos na figura seguinte, $VMR=10,5583$ (*Variance to Mean Ratio*), o que se traduz na existência de alguns quadrantes terem mais pontos do que o estimado e outros terem menos (figura 2.4), pois sendo $VMR>1$, o padrão pontual é aglomerado (*cluster*) o que aponta (e confirma) para a existência de uma variabilidade espacial grande no padrão, ou seja, zonas de eventos aglomerados e contraste com zonas de eventos bastante dispersos.

Method		Census		
#Points (n)		107		
#Quadrats (x)		96		
μ		1,1146		
k	x	k- μ	(k- μ) ²	x(k- μ) ²
0	75	-1,1146	1,2423	93,1722
1	5	-0,1146	0,0131	0,0656
2	6	0,8854	0,7840	4,7038
3	1	1,8854	3,5548	3,5548
4	2	2,8854	8,3256	16,6513
5	1	3,8854	15,0965	15,0965
8	2	6,8854	47,4090	94,8179
9	1	7,8854	62,1798	62,1798
10	1	8,8854	78,9506	78,9506
13	1	11,8854	141,2631	141,2631
26	1	24,8854	619,2840	619,2840
Variance (s ²)		11,7681		
VMR		10,5583	χ^2	1013,60

Figura 2.9 – Dados estatísticos alcançados pelo método de contagem por quadrantes (Censos)

2.3.1.2 - Teste da significância estatística (Teste de hipóteses com teste Qui-Quadrado)

. Hipóteses para determinar se os padrões de pontos observados são significativamente diferentes de um padrão aleatório (i.e. VMR=1)

- H_0 : VMR \sim 1 \rightarrow Padrão pontual é aproximadamente aleatório
- H_A : VMR \neq 1 \rightarrow Padrão pontual disperso ou aglomerado

. Análise dos valores críticos para um nível de significância $\alpha=0.02$ (i.e. Valores na tabela do qui-quadrado para $p=0.99$ e $p=0.01$) pela comparação do valor de qui-quadrado com a distribuição teórica conhecida.

Tabela 2.1 – Distribuição teórica de Qui-quadrado ($p=0.99$ e $p=0.01$) e valor de nº de graus de liberdade (m)

F(x)	Number of Degrees of Freedom							
	40	50	60	70	80	90	100	> 100 (Approximation)
0.1	29.1	37.7	46.5	55.3	64.3	73.3	82.4	$\frac{1}{2}(h - 1.28)^2$
0.99	63.7	76.2	88.4	100.4	112.3	124.1	135.8	$\frac{1}{2}(h + 2.33)^2$

$h = \sqrt{2m - 1} = 14,53$; $m = n-1 = 106$

$1013,60 > 142,06$ ($=$) $\chi^2_{test} > \chi^2_{crit (high)}$ \rightarrow Então rejeita-se hipótese H_A e aceita-se H_0

Então pelos valores obtidos no teste de significância estatística, conclui-se que o padrão pontual é significativamente diferente de um padrão aleatório para um nível de significância $\alpha=0.02$ determinando-se, pelo teste de hipóteses, que o padrão observado é aglomerado (com uma probabilidade inferior a 2% de não ser um aglomerado).

2.3.2 - Métodos de análise do vizinho mais próximo (simulações)

2.3.2.1 - Função G com simulação (avalia a significância dos dados)

Esse método permite a comparação da função acumulada das amostras com as de dados gerados aleatoriamente (envelopes de simulação superior e inferior).

São observáveis nos dois gráficos seguintes, valores estimados e simulados (envelopes superior e inferior) acima da recta de referência (ângulo de 45°), nos quais os valores estimados estão compreendidos no intervalo de confiança dos valores simulados (envelopes superior e inferior). Estes dados indicam uma existência forte de aglomerados para as menores distâncias e com o aumento das distâncias entre pontos vizinhos, passa de um padrão pontual a um padrão proporcionalmente mais disperso.

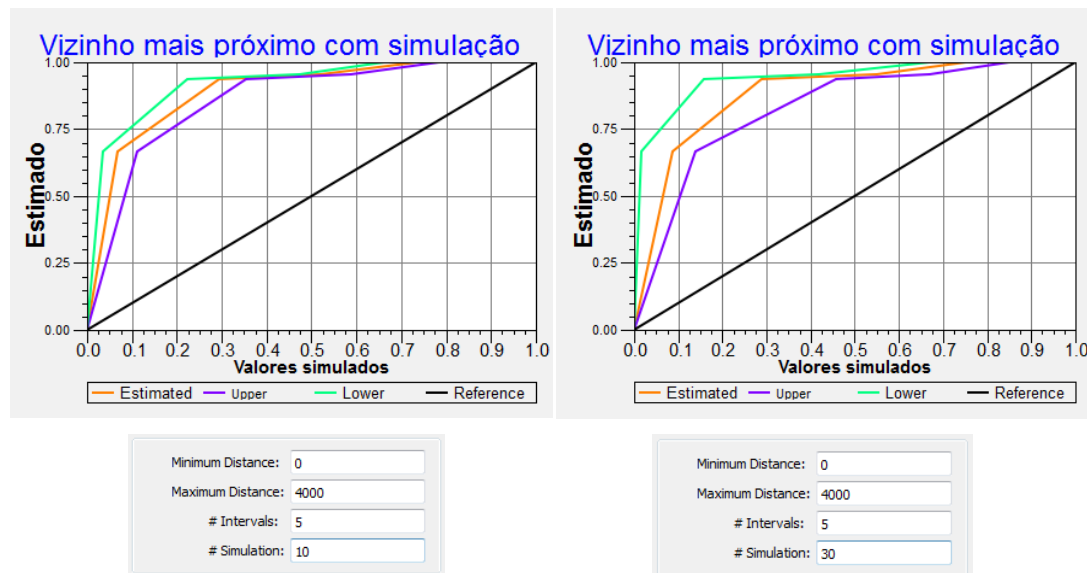


Figura 2.10 – Gráficos do vizinho mais próximo (intervalos de 5 metros)

2.3.2.2 - Função L com simulação (derivada da função K)

Esta função permite estimar da significância dos desvios da distribuição em relação à Completa Aleatoriedade Espacial, possibilitando aferir sobre a estatística do vizinho mais próximo ($R=R_O/R_E$).

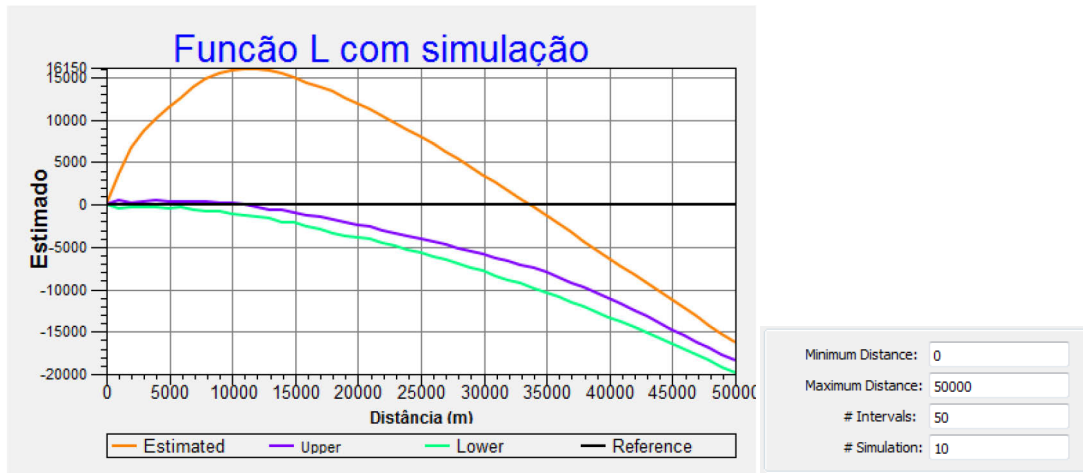


Figura 2.11 – Gráfico da função L com simulação para o padrão pontual Esquadras PSP

Pelos resultados alcançados na figura 2.11, repara-se que os valores estimados são superiores aos envelopes simulados ($R_E > R_O$) da hipótese CSR (*Complete Spatial Randomness*) em toda a distância considerada, revelador da existência de agrupamento no padrão pontual. O extremo positivo da função $\hat{L}(h)$, em torno dos 11000 m, é a distância para a qual a diferença entre valores estimados e simulados é maior, $R_E \gg R_O$, ou seja, onde o aglomerado é mais acentuado. Para o intervalo de distâncias [37900; 38100] metros, a estatística do vizinho mais próximo, $R = R_O/R_E \approx 2,1491$, indicador de que o padrão é completamente aleatório neste curto intervalo. A partir, aproximadamente dos 45000 m, os valores estimados e simulados tendem para valores idênticos, ou seja $R \approx 1$, o que se traduz num padrão aleatório. A conclusão destas observações é apresentada visualmente na figura seguinte.

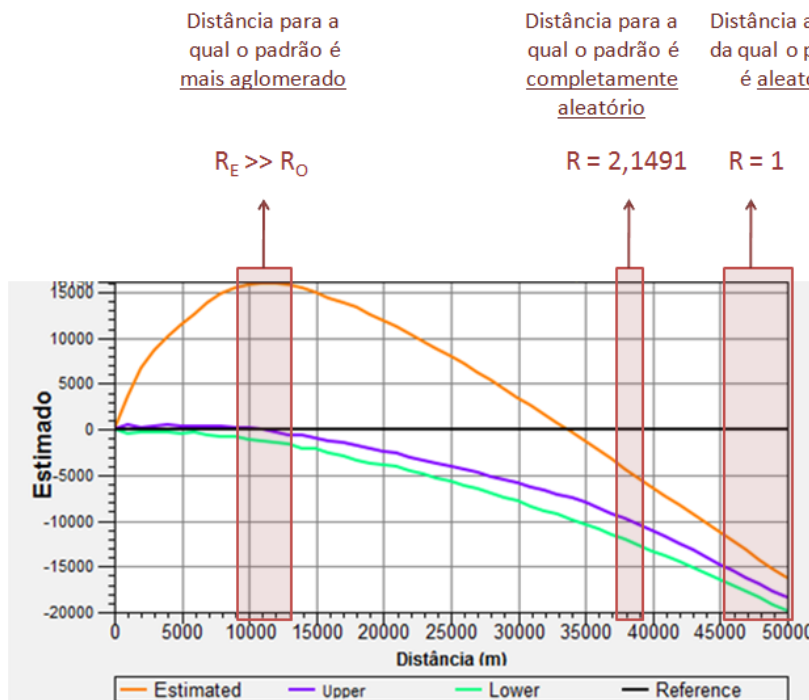


Figura 2.12 – Razão entre a distância observada (R_O) e a distância esperada (R_E) num intervalo de [0; 50000] m

Da figura 2.12 concluí-se que o padrão pontual da variável esquadras PSP atinge uma aglomeração máxima para um intervalo próximo dos 11000 metros de distância entre esquadras vizinhas, transformando-se progressivamente num padrão aleatório com o aumento de distância entre vizinhos.

2.3.2.3 - Média da distância do vizinho mais próximo

Este método compara a distância média do vizinho mais próximo com uma distância média do vizinho mais próximo gerada aleatoriamente (CSR).

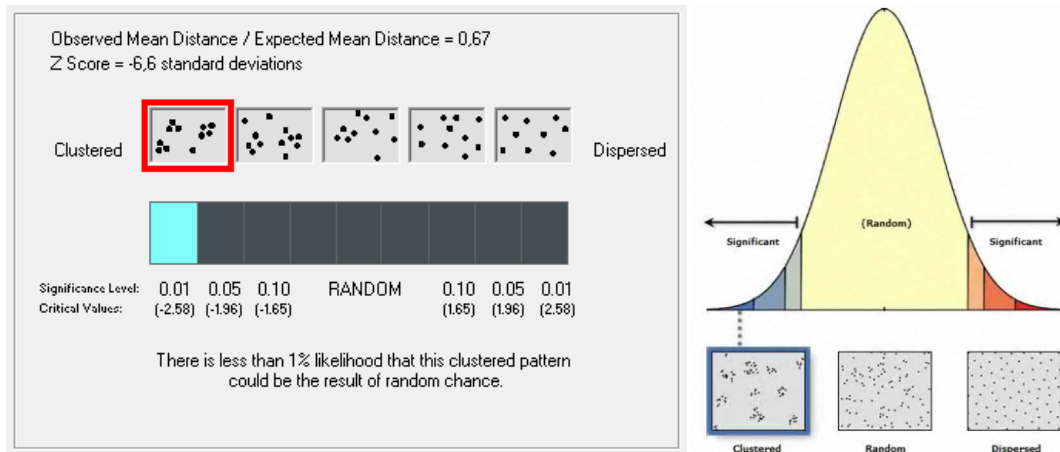


Figura 2.13 – Teste de pseudo-significância da média da distância do vizinho mais próximo

Os valores da média da distância do vizinho mais próximo são:

Distância média observada $\approx 925,593$ m

Distância média estimada $\approx 1388,840$ m

Pelos valores no teste de significância estatística ($Z\ score = 6,6$ desvios padrão) expressos na figura anterior, denota-se que para o nível de significância $\alpha=0,01$ (probabilidade de falsa rejeição da hipótese nula num teste estatístico), o padrão pontual é um aglomerado com uma significância estatística de 99%, (isto é, a probabilidade do padrão pontual não ser resultado de uma escolha aleatória é inferior a 1%).

3. Análise espacial de áreas

3.1 – Visualização

Na análise espacial de áreas, o padrão pontual esquadras PSP foi agregado em unidades territoriais (NUTS) correspondentes aos concelhos na área da grande Lisboa. Importante referir que, embora a unidade territorial mais apropriada seja a freguesia (maior resolução espacial, equivalente a dados mais desagregados), a escolha recaiu numa divisão por concelhos, face à limitação das variáveis relevantes para este trabalho que apresentassem dados actuais, possuírem como maior resolução espacial, a divisão por concelhos. Os dados das variáveis seguintes foram obtidos em Abril de 2011.

Variáveis em estudo:

- N° total de crimes registados pelas autoridades policiais (dados referentes ao ano 2010)
- Densidade populacional (N° habitantes/Km²; valores correspondentes ao ano 2009)
- N° desempregados registados nos centros de emprego (IEFP, dados de Junho de 2009)
- N° esquadras da PSP (informação referente a Maio de 2011)
- Taxa de retenção e abandono escolar (Valores referentes a 2007)

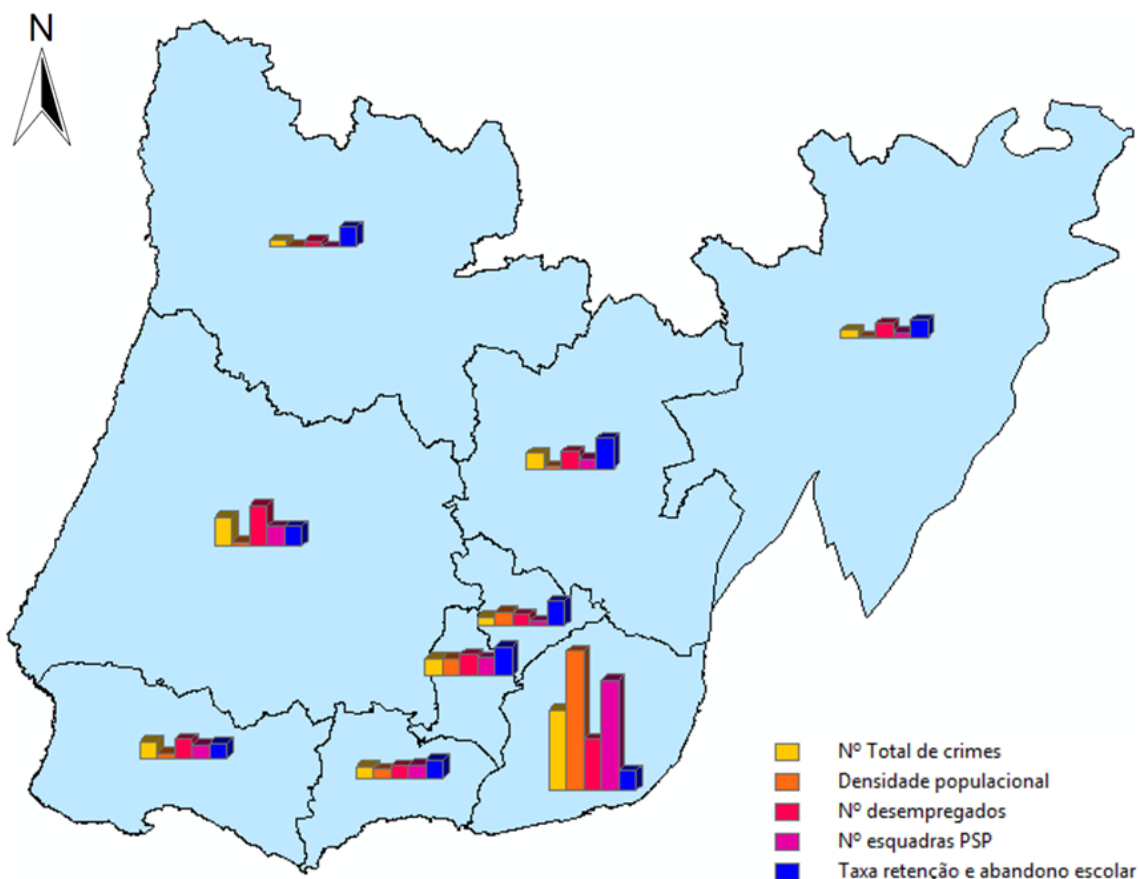


Figura 3.1 – Mapa com distribuição espacial por concelhos de todas as variáveis em análise (normalizadas pela percentagem do total)

De seguida são apresentados os mapas das distribuições espaciais destas variáveis (atributos), que tendo em conta o número reduzido de concelhos, foram classificados em três classes de quantis (cada classe apresenta o mesmo número de concelhos) possibilitando uma representação espacial mais adequada e objectiva.

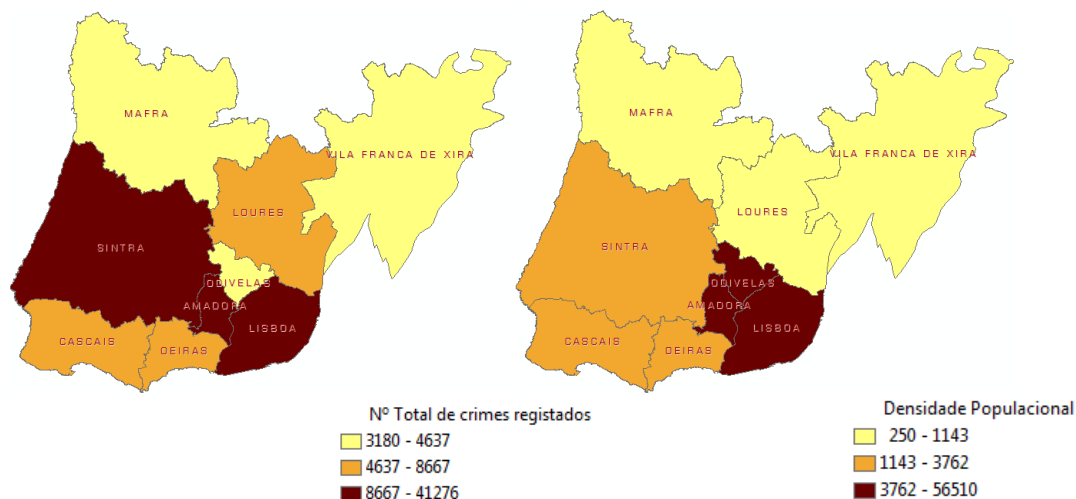


Figura 3.2 – Distribuição espacial das variáveis Nº Total de crimes registados e Densidade populacional (classificadas em 3 classes de quantis)

A variável densidade populacional ostenta uma distribuição espacial aglomerada no litoral, especialmente nos concelhos a sudeste, nomeadamente Lisboa, Amadora e Odivelas, contrastando com os baixos valores dos concelhos situados a norte da grande Lisboa.

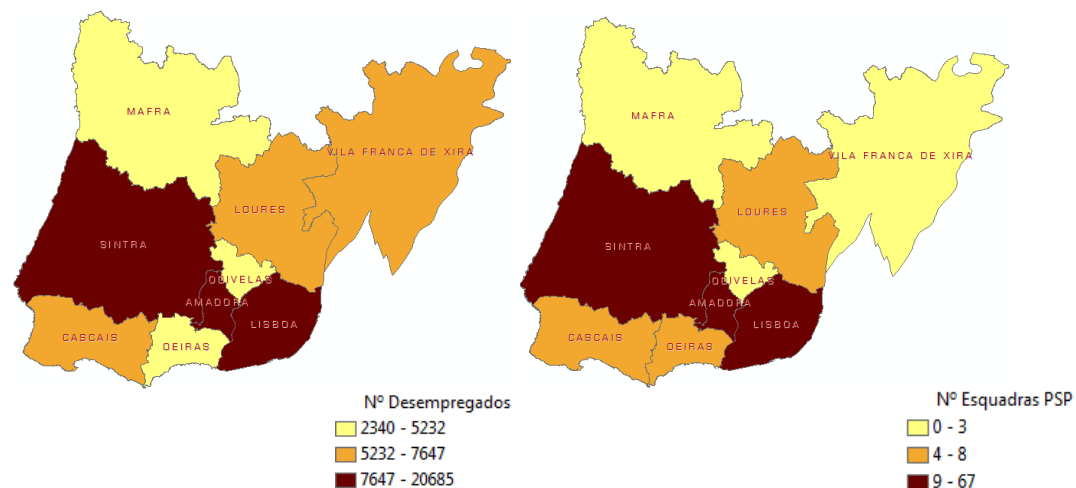


Figura 3.3 – Distribuição espacial das variáveis Nº desempregados e Nº esquadras PSP (classificadas em 3 classes de quantis)

As variáveis nº total de crimes e nº esquadras PSP exibem uma distribuição espacial idêntica, com aglomerado da classe de valores altos nos concelhos na direcção NO-SE (Sintra, Amadora e Lisboa) e classe de valores baixos a registarem-se tendencialmente nos concelhos a norte.

O nº de desempregados (figura 3.3) mostra um comportamento semelhante às duas variáveis anteriores em relação à classe de valores elevados, no entanto a classe intermédia passa de uma distribuição maioritariamente a sul para uma distribuição ao longo da região de estudo na direcção SO-NE.

O atributo taxa de retenção e abandono escolar na figura seguinte, exhibe um aglomerado de valores altos nos concelhos interiores da região de estudo (Amadora, Odivelas e Loures), verificando-se que os concelhos com valores de menor taxa de retenção e abandono se localizam nos extremos da grande Lisboa, na direcção SO-NE, assinalando que esta variável tem um comportamento distinto dos restantes atributos.

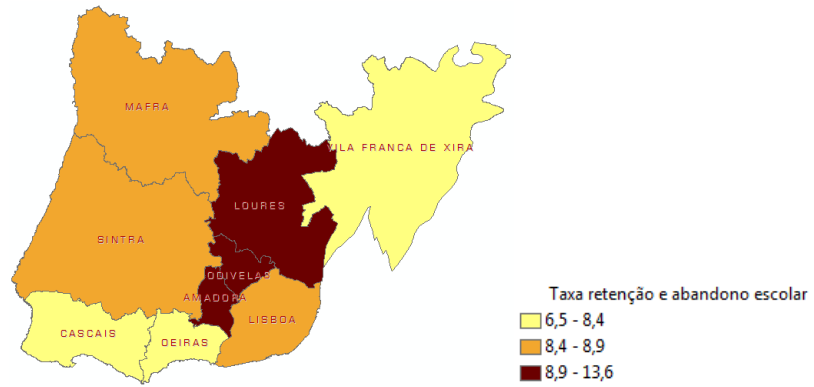


Figura 3.4 – Distribuição espacial da variável Taxa de retenção e abandono escolar (classificadas em 3 classes de quantis)

3.2 – Indicadores de associação espacial global

3.2.1 - Índice Global de Moran

De forma a quantificar o grau de autocorrelação espacial existente em cada uma das variáveis de estudo, foi calculado o índice global de Moran.

Na aquisição dos resultados ao longo deste capítulo, o método de relação espacial entre concelhos foi a contiguidade de 1ª ordem para criação da matriz de proximidade espacial.

Analisando os valores obtidos do índice global de Moran na tabela seguinte, é possível constatar que a nível global as variáveis densidade populacional e taxa de retenção e abandono escolar apresentam uma distribuição espacial aleatória (ausência de autocorrelação espacial) pelo que as variáveis são espacialmente independentes. A estas variáveis, incluem-se numa menor extensão, o número de esquadras e de crimes, respectivamente. No número de desempregados o resultado indica autocorrelação espacial repartida entre a aleatoriedade e a autocorrelação espacial negativa (I com valores próximos de -1), indicador de padrão dispersão.

Tabela 3.1 – Índice de Moran global das variáveis em análise

Variáveis	Nº crimes	Nº desempregados	Nº esquadras	Densidade populacional	Taxa retenção e abandono escolar
I Moran global	-0,1998	-0,430	-0,1186	-0,0452	0,0487

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

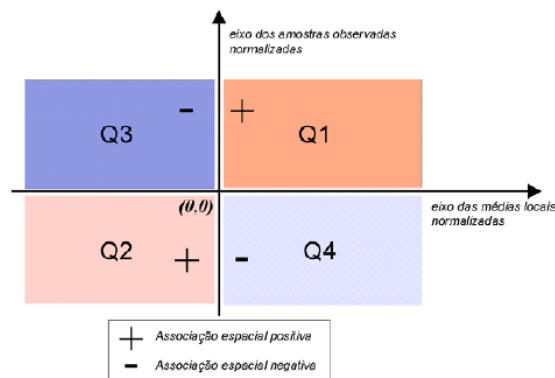


Figura 3.5 – Regimes espaciais no diagrama de dispersão de Moran

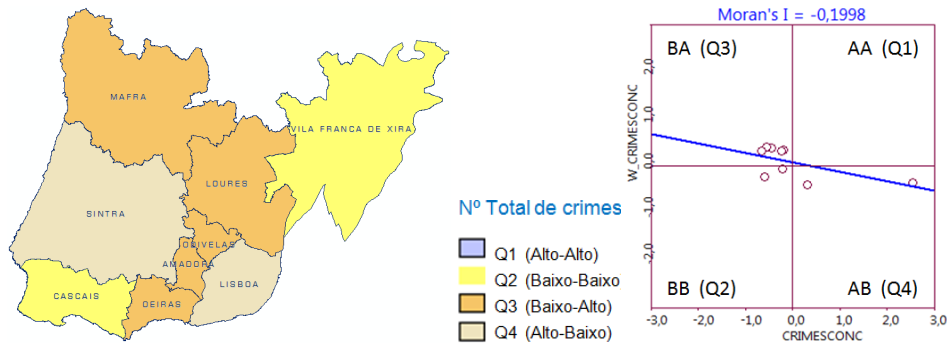


Figura 3.6 – Mapa de dispersão de Moran para a variável N° total de crimes

Na figura anterior observa-se no diagrama de dispersão a tendência global para uma associação espacial negativa (quadrantes 3 e 4) na variável número total de crimes, sinónimo de um padrão espacial com concelhos vizinhos menos similares que os mais distantes. Embora apresente um mapa de dispersão similar ao da variável número de desempregados, repara-se que os concelhos dos regimes espaciais Q3 e Q4 (baixo-baixo e baixo-alto) indiciam corresponder, levemente, a uma zona de transição de regimes espaciais inserida entre os concelhos de Cascais e Vila Franca. Estes são concelhos com vizinhos (Loures e Sintra mais Oeiras respectivamente) de valores semelhantes (regime baixo-baixo), representativos de uma autocorrelação espacial positiva.

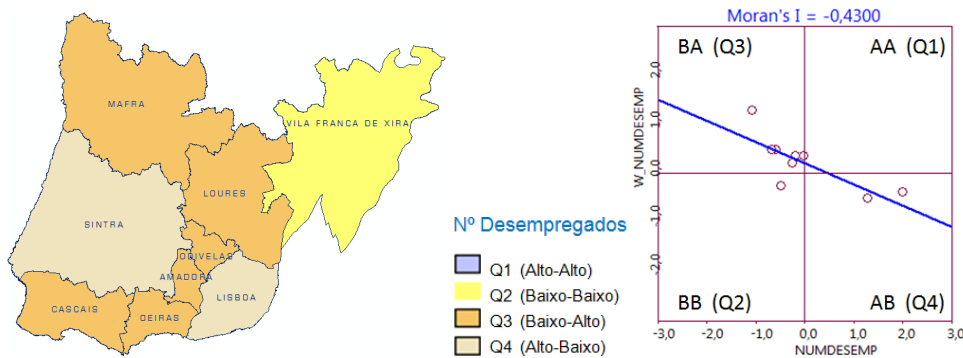


Figura 3.7 – Mapa e diagrama de dispersão de Moran para a variável N° desempregados

O número de desempregados apresenta uma tendência para associação espacial negativa (quadrantes 3 e 4), sinal de que estes concelhos têm vizinhos com valores distintos. Apenas o concelho de Vila Franca de Xira exibe tendência espacial positiva (quadrante 2) reveladora de que possui valor semelhante ao seu concelho vizinho, Loures. Observam-se na figura anterior os concelhos de Sintra e Lisboa apresentarem um regime espacial alto-baixo, fazendo transparecer estes concelhos estarem rodeados por vizinhos cujo valor do atributo desempregados está acima da média. De referir, que esta variável não apresenta nenhum concelho considerado *outlier*.

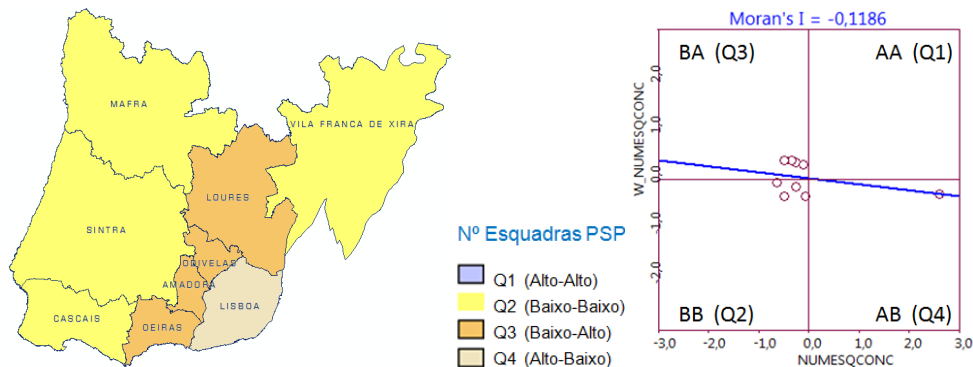


Figura 3.8 – Mapa e diagrama de dispersão de Moran para a variável N° esquadras PSP

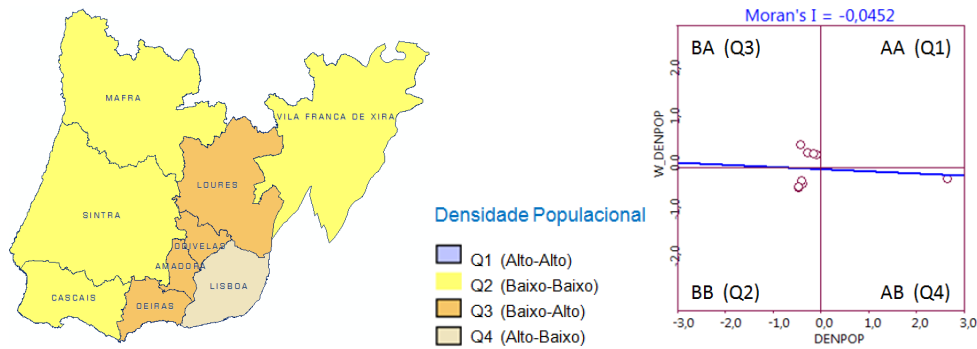


Figura 3.9 – Mapa e diagrama de dispersão de Moran para a variável densidade populacional

O índice de Moran dos atributos número de esquadras e densidade populacional (figuras 3.8 e 3.9 respectivamente) é quase zero, revelando a já referida ausência de autocorrelação espacial (aleatoriedade espacial) embora observe-se pelos respectivos diagramas de dispersão, tratar-se de uma aleatoriedade não absoluta, dada a nuvem de pontos não se encontrar dispersa pelos quatro quadrantes do diagrama, pelo contrário a nuvem de pontos (valores) é muito concentrada [Serrano, 2000]. Analisando a associação espacial dos concelhos nos mapas de dispersão de Moran, verifica-se uma predominância de associação espacial de ligeira tendência positiva (quadrante 2) nos concelhos periféricos à região sudeste da grande Lisboa, indicador de que os seus concelhos vizinhos (Oeiras, Amadora, Odivelas e Loures) apresentam valores mais semelhantes relativamente aos concelhos mais afastados (caso do concelho de Lisboa). Observa-se também que apenas o concelho de Lisboa (11,1% da totalidade dos concelhos) exhibe valores de médias locais positivas.

No concelho de Lisboa (quadrante 4) e seus concelhos vizinhos (quadrante 3) impera uma tendência para associação espacial negativa (dispersão), na qual o concelho de Lisboa tem densidade populacional alta e os concelhos vizinhos densidade baixa. Os concelhos do quadrante 3 (regime espacial baixo-alto) têm uma densidade populacional baixa com vizinhos de densidade alta. Denota-se que os concelhos do quadrante 3 representam uma zona de transição entre diferentes regimes espaciais (padrões baixo-alto e alto-baixo) ou pontos de não estacionariedade (estrutura e magnitude da correlação espacial) da densidade populacional.

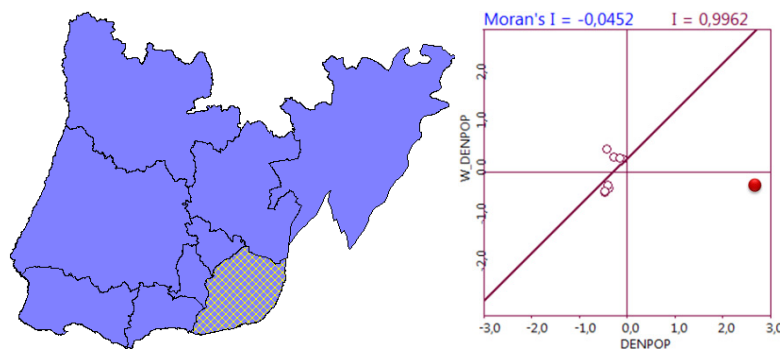


Figura 3.10 – Seleção e exclusão do leverage point (concelho de Lisboa) no índice global de Moran na variável densidade populacional

Pelo diagrama de dispersão de Moran na figura anterior denota-se que o concelho de Lisboa corresponde a um leverage point, isto é, uma localização que têm grande influência sobre a tendência central é um ponto de alavancagem [Varga, 1998]. Este concelho de regime espacial alto-baixo (que apresenta um grande desvio em relação à média da densidade populacional) apresenta uma influência de tal forma elevada na autocorrelação espacial, que a sua exclusão leva a densidade populacional a apresentar uma autocorrelação positiva ($I=0.9962$ indicador de agrupamento) em detrimento da aleatoriedade espacial ($I=-0.452$) exibida inicialmente com a sua inclusão (perante a nova recta de declive de Moran, Lisboa assemelha-se a um pseudo-outlier). Esta situação verifica-se também na variável número de esquadras (figura 3.8), dada a similaridade evidenciada no diagrama de dispersão de ambos.



Figura 3.11 – Mapa e diagrama de dispersão de Moran para a variável Taxa de retenção e abandono escolar

A variável taxa de retenção e abandono escolar (figura 3.11) apresenta a nuvem de pontos bastante dispersa relativamente às restantes variáveis, o que sugere, neste caso, uma aleatoriedade mais acentuada, embora não esteja dispersa nos quatro quadrantes do diagrama de dispersão [Serrano, 2000].

Observa-se um aglomerado com valores altos (quadrante 1) dos concelhos “interiores” (Amadora, Odivelas e Loures) correspondentes aos concelhos com maior taxa de retenção. Este *cluster* equivale a uma zona de transição entre os concelhos situados a oeste do regime espacial alto-alto (concelhos com valores observados e média dos vizinhos abaixo da média global) com autocorrelação espacial positiva (*cluster*) e os concelhos a este (Lisboa e Vila Franca de Xira) com regime espacial baixo-alto, os quais apresentam autocorrelação espacial de cariz negativo, sinónimo de um padrão espacial disperso.

Neste diagrama repara-se na presença de um *outlier* (valor extremo relativamente à recta de Moran na figura seguinte), Vila Franca de Xira, cujos vizinhos apresentam desvios elevados em relação à média (concelho de Vila Franca).

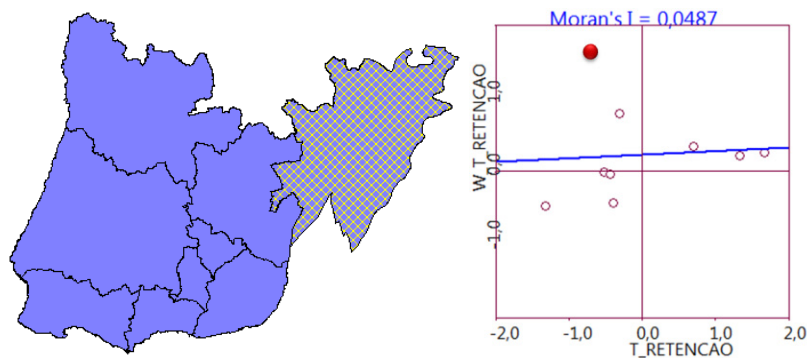


Figura 3.12 – Selecção do *outlier* excluído (Vila Franca de Xira) no índice global de Moran na variável Taxa de retenção e abandono escolar

Verificou-se que o índice de Moran com exclusão do *outlier* (figura 3.12) conferiu à variável uma maior tendência para autocorrelação espacial positiva (valor do índice de Moran apresentou um aumento ligeiro).

3.2.2 – Verificação da validade estatística

Uma vez calculado o índice global de Moran (índice de associação espacial global), há que verificar a sua validade estatística, para tal procedeu-se à aplicação de testes de pseudo-significância estatística (i.e. significância) pela geração de 999 permutações, para cada variável em análise.

A quantificação e interpretação da significância estatística foram realizadas respeitando os critérios (nível de significância, nível crítico, *Z score*) da distribuição normal demonstrados na figura seguinte.

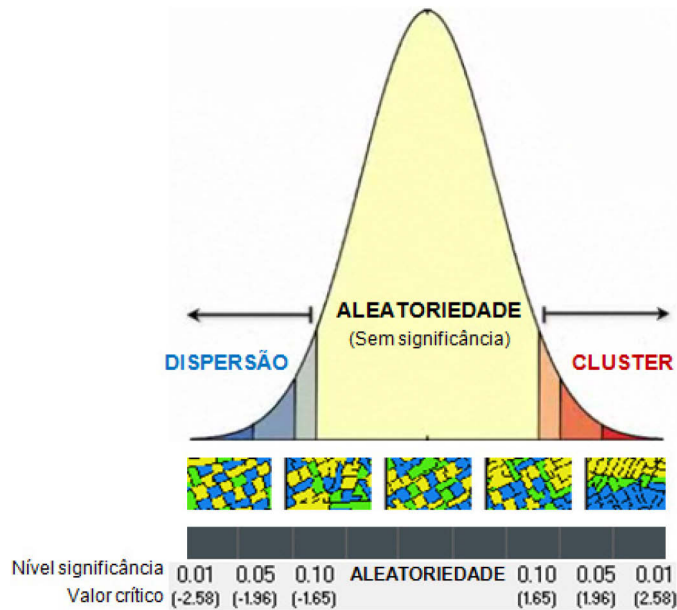


Figura 3.13 – Classificação da distribuição de valores nos testes de pseudo-significância

Os valores adquiridos são apresentados na tabela seguinte.

Tabela 3.2 – Verificação da validade estatística do índice global de *Moran* (testes de pseudo-significância com 999 permutações)

Variáveis	Testes de pseudo-significância						Autocorrelação espacial
	Z score	Média	Desvio padrão	Nível significância	p value	Valor crítico	
Nº crimes	-1,050	-0,129	0,121	Sem significância	0,233	-	<p>Embora relativamente disperso, o padrão pode ser resultado de uma escolha aleatória (CSR)</p>
Nº desempregados	-1,900	-0,132	0,207	0,1	0,049	1,65	<p>5-10% de probabilidade do padrão disperso ser resultado de uma escolha aleatória (CSR)</p>
Nº esquadras	-0,330	-0,124	0,095	Sem significância	0,527	-	<p>Independência espacial (Aleatoriedade)</p>
Densidade populacional	0,400	-0,124	0,091		0,843		
Retenção e abandono escolar	0,180	-0,123	0,213		0,213		

Verifica-se que a variável número de desempregados é a única variável a apresentar distintamente um resultado com significância estatística. Esta significância, na ordem dos 90% a 95%, permite validar a conclusão de que, para o índice global de *Moran*, o número de desempregados apresenta uma tendência para autocorrelação espacial negativa, ou seja um padrão disperso.

As restantes variáveis, em consequência do reduzido desvio padrão nos valores *Z score*, apresentam resultados correspondentes ao intervalo da distribuição normal sem significância estatística.

3.3 – Indicadores de associação espacial Local

3.3.1 - Local Indicators of Spatial Autocorrelation (LISA)

Para realizar uma análise de dependência espacial local foram obtidos os mapas LISA de significância e aglomerados, permitindo determinar se algum padrão de autocorrelação espacial foi omitido na obtenção dos indicadores globais.

Para a aquisição dos seguintes resultados, foi criada uma matriz de pesos em função das distâncias dos centróides de cada polígono.



Figura 3.14 – Mapas de significância e aglomerados locais LISA da variável Nº total de crimes após 999 permutações



Figura 3.15 – Mapas de significância e aglomerados LISA da variável Nº de desempregados após 999 permutações



Figura 3.16 – Mapas de significância e aglomerados LISA da variável Nº esquadras PSP após 999 permutações

Nas figuras 3.14 a 3.16 não é detectável a existência de padrões locais de autocorrelação espacial que não fossem identificáveis globalmente, como a presença de *outliers* ou *clusters* locais significantes nas respectivas variáveis, tal como a indicação de instabilidade local (desvios locais do padrão global de autocorrelação espacial).

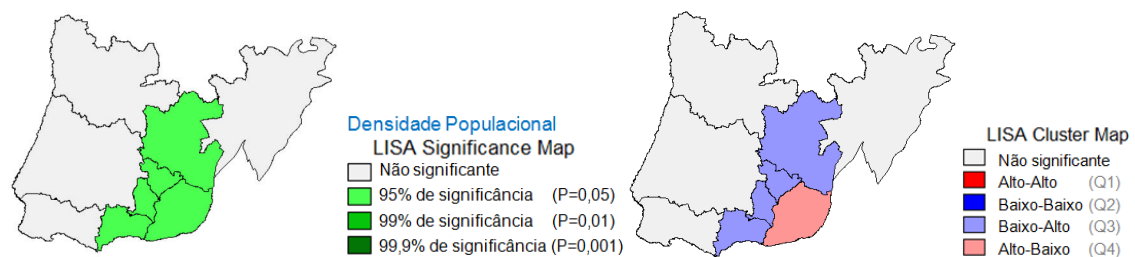


Figura 3.17 – Mapas de significância e aglomerados LISA da variável Densidade populacional após 999 permutações

Na análise de dependência espacial local na figura anterior (mapa de significância LISA) a autocorrelação espacial local do atributo densidade populacional detecta-se um aglomerado de concelhos com 95% de significância estatística composto pelos concelhos de Lisboa e seus vizinhos, áreas que possuem dependência espacial significativa (quadrantes 3 e 4) confirmando a ligeira tendência global da autocorrelação espacial negativa.

Pelo mapa LISA *Cluster* na figura 3.17 é visível que os concelhos do quadrante 3 (Oeiras, Amadora, Odivelas e Loures) correspondem a uma zona de transição entre diferentes regimes espaciais (ou pontos de não estacionariedade do atributo, isto é a existência de tendência espacial) de concelhos Q2 (médias locais negativas e observações negativas) para Q4 (médias locais negativas e observações positivas) confirmando-se localmente, os regimes espaciais obtidos a nível global.

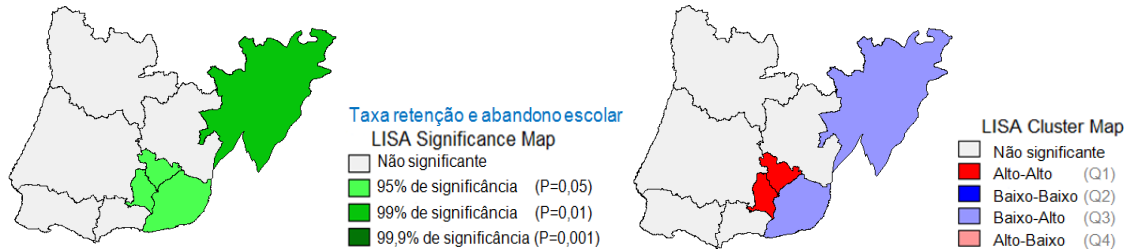


Figura 3.18 – Mapas de significância e aglomerados LISA da variável Taxa de retenção e abandono escolar após 999 permutações

3.3.2 - Índice Local de Moran

Índice utilizado de modo a medir a dependência espacial e a determinar, a nível local, a presença de *clusters* (concelhos de valores altos e positivos) e *outliers* (*anti-clusters*).

O parâmetro utilizado para relação espacial entre as áreas foi a contiguidade de polígonos.

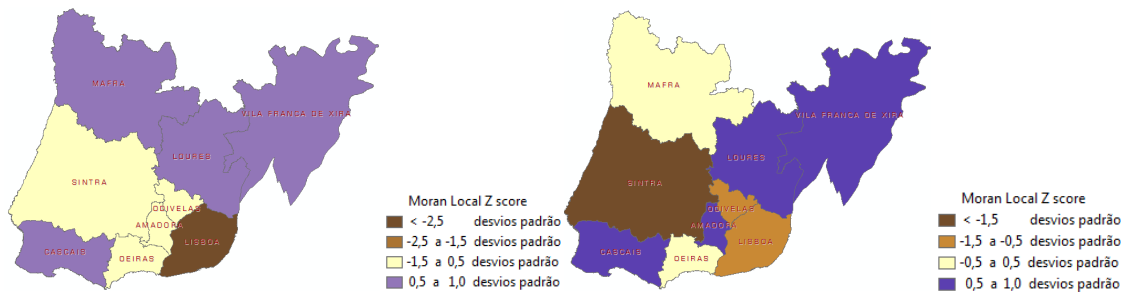


Figura 3.19 – Valores Z score do Índice Local de Moran das variáveis N° total de crimes e N° desempregados

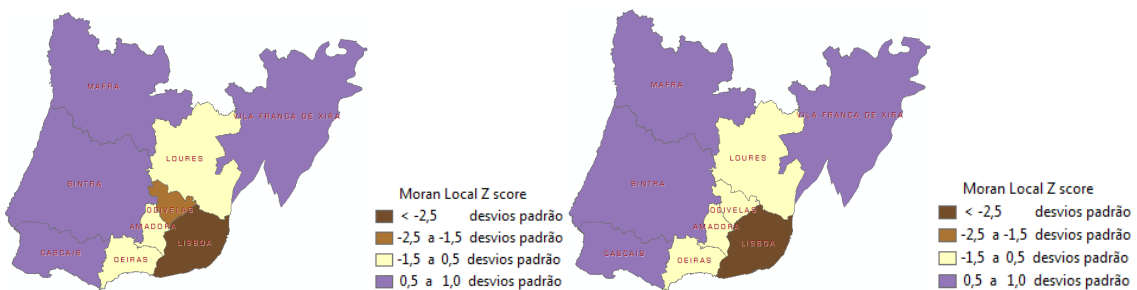


Figura 3.20 – Valores Z score do Índice Local de Moran das variáveis N° esquadras PSP e densidade populacional

As variáveis número de crimes, número de esquadras e densidade populacional apresentam um *outlier* espacial (*anti-cluster*) com uma significância de 99%, o concelho de Lisboa, visto este apresentar um acentuado valor negativo de Z score com desvio superior a dois desvios padrão [Queiroz, 2003]. Sendo este valor negativo, implica que concelho de Lisboa tem valores elevados e os seus vizinhos valores baixos [Esri, 2011]. A variável número de esquadras apresenta adicionalmente o concelho de Odivelas como *anti-cluster*, embora com significância de 90%-95%.

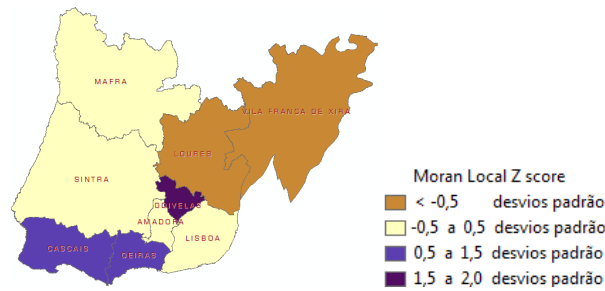


Figura 3.21 – Valores Z score do Índice Local de Moran da variável Taxa de retenção e abandono escolar

Relativamente às variáveis número de desempregados e taxa de retenção os concelhos, respectivamente, de Sintra e Odivelas são apresentados como *outliers*, embora com uma significância estatística reduzida, inferior a 90-95%.

3.3.3 - Estatística Getis-Ord General G (aglomerados de valores altos/baixos)

De forma a determinar o grau de aglomeração de valores altos e baixos nas variáveis, procedeu-se à determinação da estatística G. A avaliação, presente na tabela 3.3 foi efectuada segundo a figura de seguida apresentada.

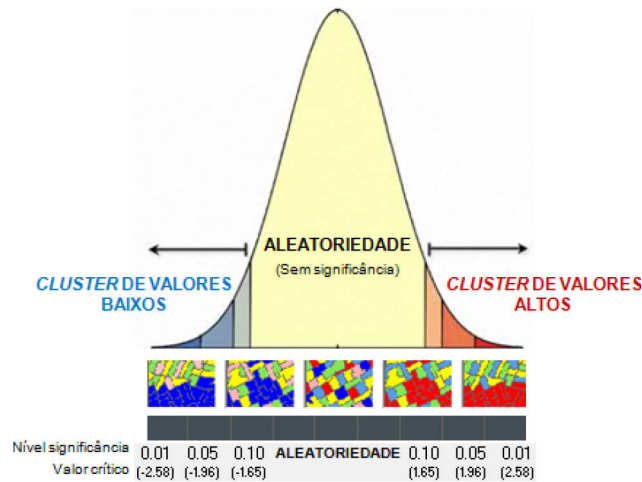


Figura 3.22 – Classificação da distribuição de valores na estatística *Getis-Ord General G*

Tabela 3.3 – Estatística G para medição do grau de aglomeração de altos/baixos nas variáveis em estudo

Variáveis	Getis-Ord General G				
	G	Z score	Nível significância	Valor crítico	Clusters altos/baixos
Nº crimes	0,49	0,57	Sem significância	-	
Nº desempregados	0,51	0,80			0,01 0,05 0,10 ALEATORIEDADE 0,10 0,05 0,01
Nº esquadras	0,56	0,75			Ausência de clusters (Aleatoriedade espacial)
Densidade populacional	0,81	1,65	0,1	1,65	
					0,01 0,05 0,10 ALEATORIEDADE 0,10 0,05 0,01 5-10% de probabilidade do cluster de valores altos ser resultado de uma escolha aleatória (CSR)
Retenção e abandono escolar	0,48	1,44	Sem significância	-	
					0,01 0,05 0,10 ALEATORIEDADE 0,10 0,05 0,01 Embora exista algum clustering, o padrão pode ser resultado de uma escolha aleatória (CSR)

3.3.4 – Estatística G_i^* (Estatística para análise de hot spots/cold spots)

Para a determinação dos concelhos correspondentes aos aglomerados de valores altos e baixos (*hot spots* e *cold spots*) levou-se a cabo a estatística G_i^* .

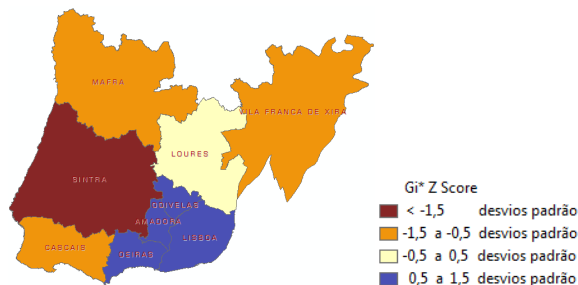


Figura 3.23 – Mapas G_i^* (valores Z score de G_i^*) classificados por desvios padrão das variáveis Nº total de crimes, densidade populacional e nº equipas PSP

Percebe-se na figura anterior, que o concelho de Sintra é um *cold spot* (*cluster* espacial de valores baixos), na medida em que apresenta um valor Z score baixo aliado a um p value baixo neste concelho [Esri, 2011]. Para o número de desempregados, na figura seguinte, foi verificada a mesma situação.

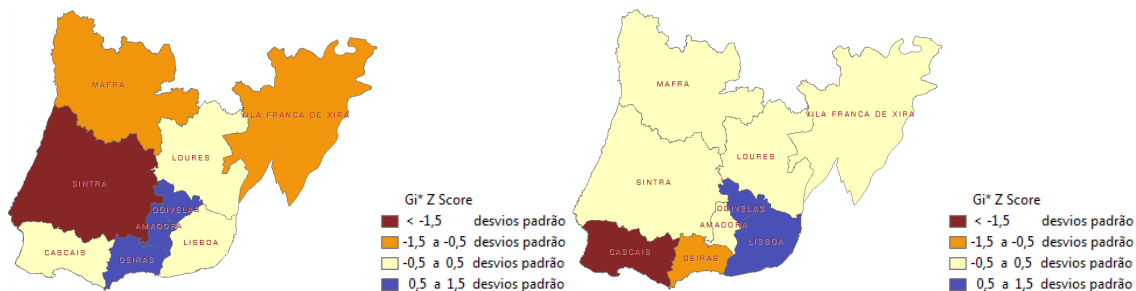


Figura 3.24 – Mapas G_i^* (valores Z score de G_i^*) classificados por desvios padrão das variáveis Nº desempregados e Taxa de retenção e abandono escolar

Na taxa de retenção, obteve-se Cascais como um *cold spot* e observa-se que em todas as variáveis não foram detectados *hot spots* (*cluster* espacial de valores altos) com significância estatística.

3.4 – Conclusão da análise espacial de áreas

Pelos resultados obtidos neste capítulo foi possível concluir que a distribuição espacial das equipas do Comando Metropolitano apresenta, principalmente pelos indicadores globais, um comportamento espacial bastante similar à densidade populacional. Esta semelhança é indicativa da densidade populacional ser um critério tido em conta na área da grande Lisboa, permitindo uma gestão mais eficiente de recursos através da distribuição espacial de equipas PSP por parte do Ministério da Administração Interna.

Ao confrontar a distribuição do número de equipas com o número total de crimes registados e o número de desempregados, verifica-se que estes atributos aparentam ter menor influência na gestão da distribuição espacial de equipas na grande Lisboa, como se observou com a densidade populacional. Um indicador socioeconómico como o número de desempregados, não aparenta apresentar relevância na distribuição espacial das equipas. Para uma análise mais aprofundada seria útil conhecer as localizações geográficas de zonas de risco e exclusão social, como bairros sociais e o número de crimes relativamente a áreas vizinhas.

Quanto à taxa de retenção, esta apresenta uma correlação espacial distinta do número de crimes, indiciando que insucesso escolar juvenil não apresenta um peso considerável na quantidade de crimes, embora sejam unicamente considerados os crimes registados pelas autoridades e não os ocorridos. Para estas variáveis, seria útil, por exemplo, dispor da informação relativa à faixa etária estimada dos infractores expressas nas queixas preenchidas pelas vítimas registadas pela PSP. A relação entre a taxa de retenção e os crimes será abordada no capítulo seguinte.

Importante destacar os resultados desta análise espacial de áreas, infelizmente, serem limitados pela impossibilidade de considerar freguesias como unidade territorial na agregação por áreas que permitiria dados mais desagregados (maior nível de resolução espacial). Esta situação do efeito de escala, é um problema da unidade de área modificável (*Modifiable Areal Unit Problem*), não permitiu lançar uma análise que seria sem dúvida mais interessante e exacta sobre a realidade espacial actual do tema em análise.

4. Regressão espacial

Neste capítulo, os objectivos principais tidos em conta a cumprir, passaram pela análise e determinação de qual o melhor método de regressão (modelo para estimar uma variável no espaço) e pela avaliação da presença ou não de efeitos espaciais (tendências espaciais). O primeiro objectivo, permite a obtenção do modelo que melhor estime a variável dependente (Y) através do conhecimento de quais as variáveis explicativas que exprimem melhor a variável dependente (X_i). O segundo, a verificação da existência de dependência espacial.

Ponderando as variáveis abordadas no capítulo dois, considerou-se fazer mais sentido utilizar a variável número de crimes registados como sendo a variável dependente em estudo, em detrimento da variável pontual número de esquadras da PSP.

Na criação dos modelos com efeitos espaciais (*Spatial Lag Model* e *Error Model*), a matriz de proximidade espacial (matriz de vizinhança) necessária foi criada pelo método *rook contiguity* e visto o número de áreas não ser elevado, o método foi de 1ª ordem [Anselin, 2005].

4.1 - Relação entre as variáveis independentes e dependente

Na medida em que modelos de regressão devem ser calculados preferencialmente, quando as suas variáveis explicativas (independentes) apresentem correlação significativa com a variável dependente (para obtenção de um modelo mais exacto) procedeu-se primeiramente a *scatter plots* apresentados nas figuras 4.1 e 4.2, nos quais é possível averiguar pela recta de declive estimado a existência de relação entre as variáveis, em particular, se a relação é linear.

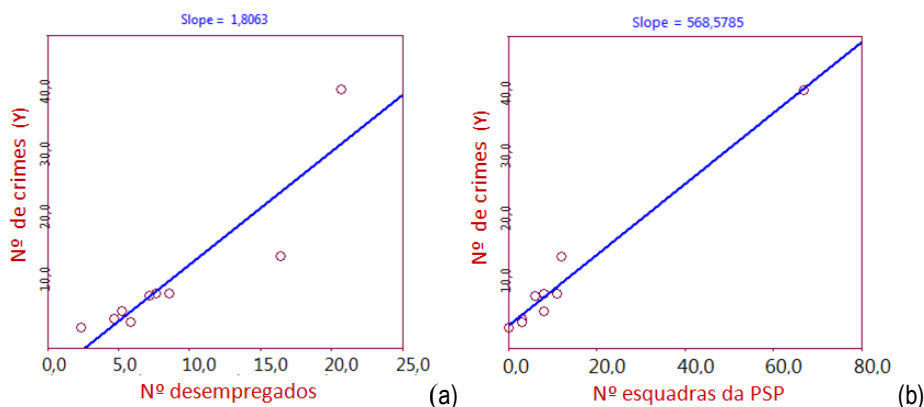


Figura 4.1 – Gráfico de dispersão da relação entre o número total de crimes (variável dependente) e o número de desempregados (a) e número de esquadras da PSP (b)

As variáveis do número de desempregados e principalmente o número de esquadras apresentam uma relação linear significativa com a variável dependente Y (número total de crimes registados), isto é, uma relação que explique a variável Y como se observa pelo declive estimado na figura anterior.

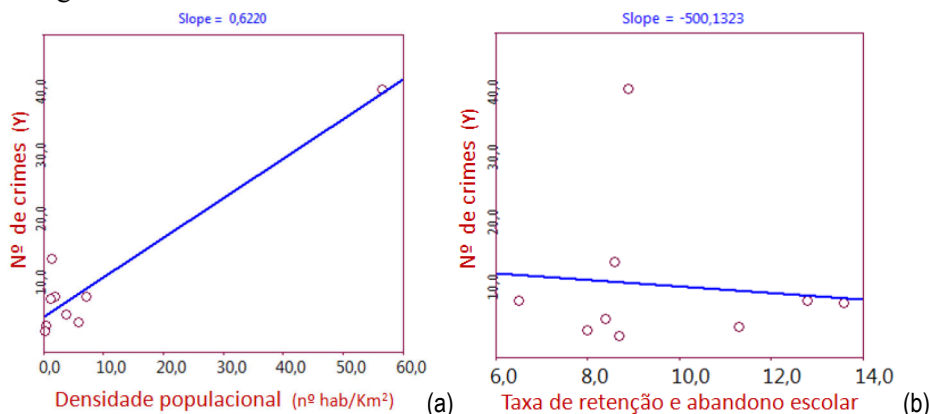


Figura 4.2 – Gráfico de dispersão da relação entre a variável independente (Y) e a densidade populacional (a) e taxa de retenção e abandono escolar (b)

Na figura 4.2 (a) observa-se que a relação também é linear, mas em menor extensão. Verifica-se que a concentração da nuvem de pontos aparenta uma distribuição menos linear e confirma-se a situação analisada no capítulo 3.2.1 (Índice Global de Moran), no qual o ponto afastado do aglomerado de valores é um *leverage point* (concelho de Lisboa) ponto este com um peso que permite a relação com o número de crimes ser linear.

No gráfico correspondente à taxa de retenção e abandono escolar (figura 4.2 b) é exibida uma nuvem de pontos dispersa de forma não linear, pelo que se conclui que não existe uma relação linear com a variável dependente, isto é, uma quase ausência de correlação entre as variáveis.

É identificável também um *outlier* assinalado na figura seguinte, para o qual se verificou a sua influência pela exclusão, na relação com a variável Y (número de crimes) observando-se uma relação ausente mais acentuada entre variáveis.

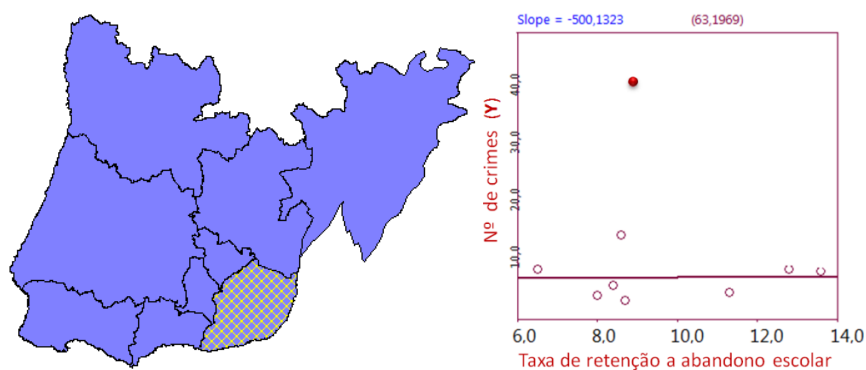


Figura 4.3 – Seleção de *outlier* (concelho de Lisboa) no gráfico de dispersão da relação entre a variável independente (Y) e a taxa de retenção e abandono escolar

Dada esta ausência de relação linear, isto é, ausência de correlação entre ambas, conclui-se que a taxa de retenção é um atributo que não explica significativamente a variável dependente número de crimes registados. Como tal, com vista à determinação de melhores estimativas dos modelos de regressão, esta variável foi excluída na determinação dos modelos de regressão com e sem efeitos espaciais [Neter, 1996].

4.2 - Avaliação da qualidade do ajustamento dos modelos de regressão

Os critérios estatísticos obtidos para análise da qualidade de ajustamento nos diferentes modelos de regressão são apresentados na próxima tabela.

Tabela 4.1 – Estatística e diagnóstico dos modelos de regressão

MODELOS DE REGRESSÃO	Sem efeitos espaciais				Com efeitos espaciais (Globais)		
	Regressão Linear				Regressão Múltipla	Spatial Lag Model (SAR)	Spatial Error Model (CAR)
VARIÁVEIS INDEPENDENTES	Densidade populacional	Nº de desempregados	Nº esquadras por concelho	Taxa de retenção e abandono escolar	.Nº Densidade populacional .Nº desempregados .Nº esquadras		
R ²	89,97%	82,06%	97,41%	1,00%	99,36%	99,41%	99,58%
R ² ajustado	88,54%	79,50%	97,04%	-13,14%	98,97%	-	-
AIC	176,64	181,87	164,44	197,24	155,89	157,18	154,49
σ ²	1,61E+07	2,88E+07	4,16E+06	1,59E+08	1,44E+06	7,41E+05	5,30E+05
I resíduos	-0,62	-0,31	-0,37	-0,02	-0,15	-0,13	-0,24
Outliers resíduos	Não			Sim	Não		
Avaliação da qualidade do ajustamento dos modelos de regressão (R ²)	Correlação elevada entre variáveis (Correlação intermédia a máxima)		Correlação quase máxima entre variáveis	Correlação entre variáveis quase nula	Correlação máxima entre variáveis		

Verificam-se que os melhores modelos de regressão são a regressão múltipla (sem efeitos espaciais) e os modelos com efeitos espaciais *Spatial Lag Model* e *Spatial Error Model*, todos eles com correlação máxima ($R^2 \cong 100\%$). Este facto manifesta que as três variáveis conjuntamente explicam melhor o comportamento do número de crimes (Y).

A diferença da qualidade de ajustamento entre estes modelos, tanto pelo coeficiente de determinação como pelo AIC (quanto menor o seu valor, melhor o modelo), é negligente. Desta igualdade conclui-se que não existe dependência espacial, na medida em que os modelos com efeitos espaciais (incorporam uma componente espacial introduzida pela matriz de proximidade espacial) apresentam valores idênticos (pelos critérios de avaliação R^2 e AIC) ao modelo sem efeitos espaciais (regressão múltipla), mostrando que a componente espacial não influencia a regressão, como tal basta considerar-se o modelo de regressão múltipla como o modelo que melhor ajustamento apresentou e melhor explicou a variável número de crimes através das variáveis dependentes [Catita, 2011].

Nos casos de regressão linear simples, confirmou-se a situação verificada na figura 4.2, na qual a taxa de retenção apresenta uma correlação praticamente nula com o número de crimes (Y), apresentando um coeficiente de determinação (R^2) de apenas 1,00% e como seria expectável a sua inclusão nos modelos de regressão com e sem efeitos espaciais influenciou negativamente a capacidade de explicação da variável dependente e consequentemente os valores de ajustamento destes modelos. É observável a correlação, quase máxima do número de esquadras ($R^2=97,41\%$) e elevada da densidade populacional e número de desempregados com a variável dependente.

De relatar, que o índice de Moran dos resíduos ($I_{Resíduos}$) é mais próximo da independência espacial para os modelos com efeitos espaciais, quando comparado com os modelos de regressão linear.

Tabela 4.2 - Coeficientes das variáveis independentes dos diferentes modelos de regressão

MODELOS DE REGRESSÃO		Sem efeitos espaciais					Com efeitos espaciais (Globais)	
		Regressão Linear				Regressão Múltipla	<i>Spatial Lag Model</i> (SAR)	<i>Spatial Error Model</i> (CAR)
Variáveis Dependentes		Densidade populacional	Nº de desempregados	Nº esquadras por concelho	Taxa de retenção e abandono escolar	.Densidade populacional .Nº desempregados .Nº esquadras		
Coeficientes	Densidade populacional	0,6220	-	-	-	0,0121	0,0313	0,1848
	Nº desempregados	-	1,8063	-	-	0,5250	0,5288	0,6530
	Nº esquadras	-	-	568,5785	-	430,4825	413,2387	229,8228
	Retenção e abandono escolar	-	-	-	-500,1323	-	-	-
	Constante	5608,3120	-4748,0440	3584,9710	15863,1600	700,9231	1038,5060	719,7506

Na tabela anterior estão apresentados os coeficientes das variáveis independentes, onde é possível quantificar a correlação das variáveis resultantes da regressão.

Observa-se na tabela 4.2 que o coeficiente da variável densidade populacional exibe um valor bastante distinto no modelo de regressão linear comparativamente aos restantes modelos. Esta situação dever-se-á possivelmente, ao facto de este modelo atribuir a correlação aos resíduos, aliada ao facto de na tabela 4.1, se verificar no modelo linear, a densidade populacional ser a única variável a apresentar um índice de Moran dos resíduos indicador de tendência de autocorrelação espacial negativa.

Sendo o modelo de regressão múltipla o que melhor qualidade de ajustamento apresentou, a sua expressão para explicação da variável número de crimes registados, em função das variáveis independentes consideradas é a seguinte:

$$Y = a + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \quad (=)$$

$$Y = 700,9231 + 0,0121\beta_1 + 0,5250\beta_2 + 430,4825\beta_3$$

4.3 - Análise da presença de efeitos espaciais

4.3.1 - Mapeamento de resíduos dos modelos de regressão

Para realização de uma análise espacial de resíduos (variância) com vista à análise da presença de tendências espaciais, procedeu-se ao mapeamento de resíduos dos modelos de regressão, dos quais se consideraram o modelo de regressão múltipla e os modelos com efeitos espaciais para finalidades de comparação, através dos gráficos de dispersão de resíduos apresentados na figura seguinte.

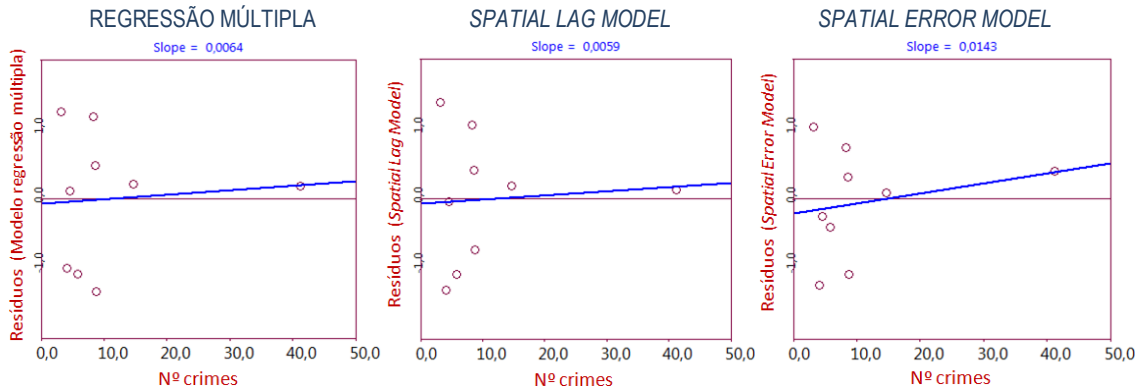


Figura 4.4 – Gráficos de dispersão dos resíduos dos modelos de regressão múltipla, *Spatial Lag Model* e *Spatial Error Model* versus variável dependente (número de crimes)

Nos gráficos de dispersão é observável, principalmente nos modelos de regressão espaciais, um claro afunilamento dos resíduos, representativo de uma variância não constante que se traduz na presença de variabilidade espacial nos modelos de regressão. Esta situação reflecte um processo não estacionário de 1ª ordem (existência de tendência espacial na região de estudo) podendo apresentar associação espacial onde não exista.

4.3.2 - Medição da autocorrelação espacial dos resíduos de regressão

Pela tabela 4.1, verifica-se que o índice de Moran dos resíduos ($I_{Resíduos}$) exibe valores a tender essencialmente para aleatoriedade espacial mas com ligeira autocorrelação espacial negativa dos resíduos, o que mostra os resíduos serem independentes.

Na figura seguinte é possível observar a significância estatística da distribuição espacial dos resíduos.

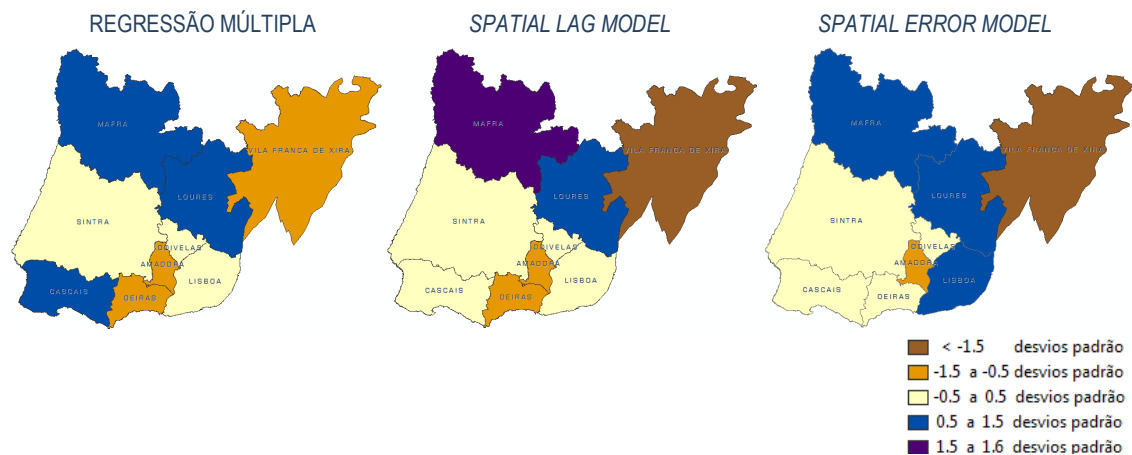


Figura 4.5 - Mapa dos resíduos padronizados dos modelos de regressão *Spatial Lag Model* e *Spatial Error Model* (classificação por desvio-padrão)

5. Análise de Redes

5.1 – Análise de redes

Neste capítulo foi realizada uma análise através da utilização de algoritmos de redes, cuja aplicação permite a simulação de alguns casos práticos aplicáveis à temática da variável pontual número de esquadras da PSP, usando para tal a rede viária da área de estudo.

5.2 – Elaboração da rede

Na produção da rede viária a sua componente geométrica foi, após a aquisição da rede, editada para eliminação de pontos redundantes (*edges*), pela prof.^a Cristina Catita em *Microstation*.

Com o objectivo de poder aplicar os algoritmos de redes não apenas em função da distância percorrida mas também em função do tempo de percurso, foram introduzidos três novos atributos na tabela de atributos da rede, permitindo ter mais uma opção de impedância (custo do movimento numa rede) na aplicação dos algoritmos, isto é, o cálculo em função da duração de um percurso (tempo) além do cálculo em função da distância de um percurso.

Após introduzir os atributos comprimento de um troço (*length*), velocidade limite de um troço (*Speed_Lim*) e duração do percurso de um troço (*TravelTime*) na tabela. Depois do calculado o atributo comprimento em cada troço, a velocidade limite de cada troço presente na rede viária, foi considerada mediante os valores em vigor no actual código da estrada presentes no actual Plano Rodoviário Nacional (PRN2000), conforme expresso na tabela seguinte.

Tabela 5.1 - Atribuição dos valores de limite de velocidade de cada troço presente na rede viária criada

Rede viária			
Tipo de troço	Rede Rodoviária Nacional indicada	Limite de velocidade (Km/h)	
steps	circulação temporária	5 (em corrida)	
pedestrian		16 (em corrida)	
footway		16 (em corrida)	
ciclovía		25	
construction		30	
living street		50	
path		50	
residential		50	
road		50	
service		50	
terciary		N	50
primary		IC, N	80
primary link		IC, N	80
secondary		IC, N	80
trunk		IP, IC	90
trunk link		IP, IC	90
motorway		AE, IC	110 (100-120)
motorway link	AE, IC	110 (100-120)	

De seguida, foi determinado o tempo de percurso (minutos) de cada tipo de troço da rede mediante a velocidade associada:

$$[\text{Tempo percurso}_{\text{troço}}] = [\text{Comprimento}_{\text{troço}}] * \frac{60}{[\text{Limite Velocidade}_{\text{troço}}]}$$

Por fim, obtiveram-se os valores da distância percorrida (em minutos), exemplificado na tabela 5.2 permitindo ter os avaliadores em ambas as impedâncias, a distância (*Length*) para o comprimento (em metros) e *TravelTime* para o tempo de percurso (em minutos).

Tabela 5.2 - Atributos introduzidos na rede viária para determinação do tempo de percurso em cada tipo de troço

length	Speed Lim	TravelTime
1.698	50	2.04
1.674	50	2.01
9.519	50	11.42
1.216	50	1.46
3.255	50	3.91
1.281	50	1.54
2.929	50	3.51
1.109	80	0.83
6.703	80	5.03
6.460	80	4.85
2.477	80	1.86

A rede viária final é apresentada na figura 5.1.

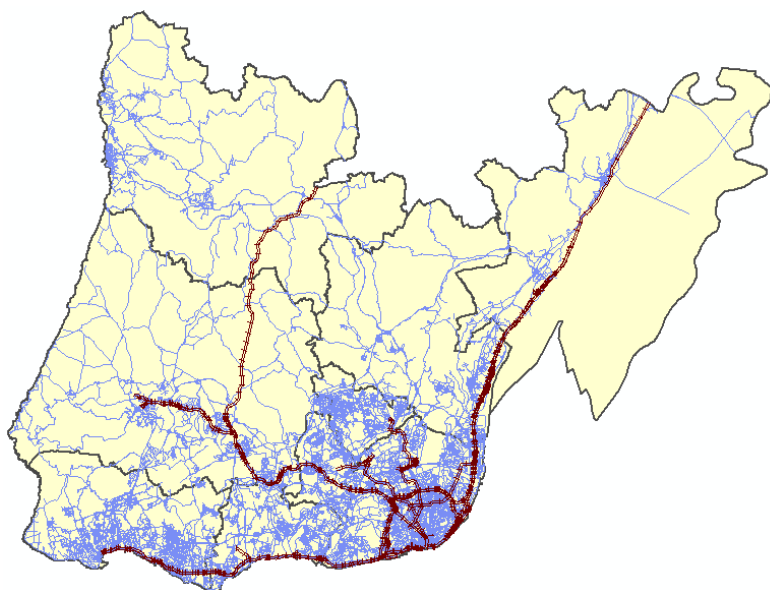


Figura 5.1 - Rede viária na região da grande Lisboa

5.3 – Algoritmos de análise de redes

5.3.1 – Acessibilidade (*Closest facility*)

Na aplicação deste algoritmo foram excluídos da rede os troços não passíveis de circulação rodoviária (passeios, escadas, ciclovias) dado o objectivo ser a avaliação do melhor tempo de resposta da PSP (por menor duração e menor distância) a uma dada ocorrência de cariz policial no concelho de Mafra (evento), por meio de uma viatura, partindo de cada uma das quatro esquadras mais próximas do concelho de Mafra (situadas em Sintra, Odivelas, Loures e Vila Franca de Xira).

Na medida em que se verificou no capítulo 2 (figura 2.5) o concelho de Mafra não apenas ser o único na região da grande Lisboa a não apresentar esquadras da PSP, mas também que as esquadras mais próximas nos concelhos vizinhos se situarem longe das fronteiras do concelho, o local da ocorrência foi o escolhido por representar o local de menor cobertura policial relativamente a infra-estruturas na área de estudo.

As figuras 5.2 e 5.3 contêm os caminhos determinados unicamente no sentido esquadras PSP (*facilities*) para a ocorrência (evento), respeitando os sentidos permitidos pela rede viária, respectivamente, dos caminhos de menor distância (comprimento como impedância) e de menor duração (tempo como impedância).

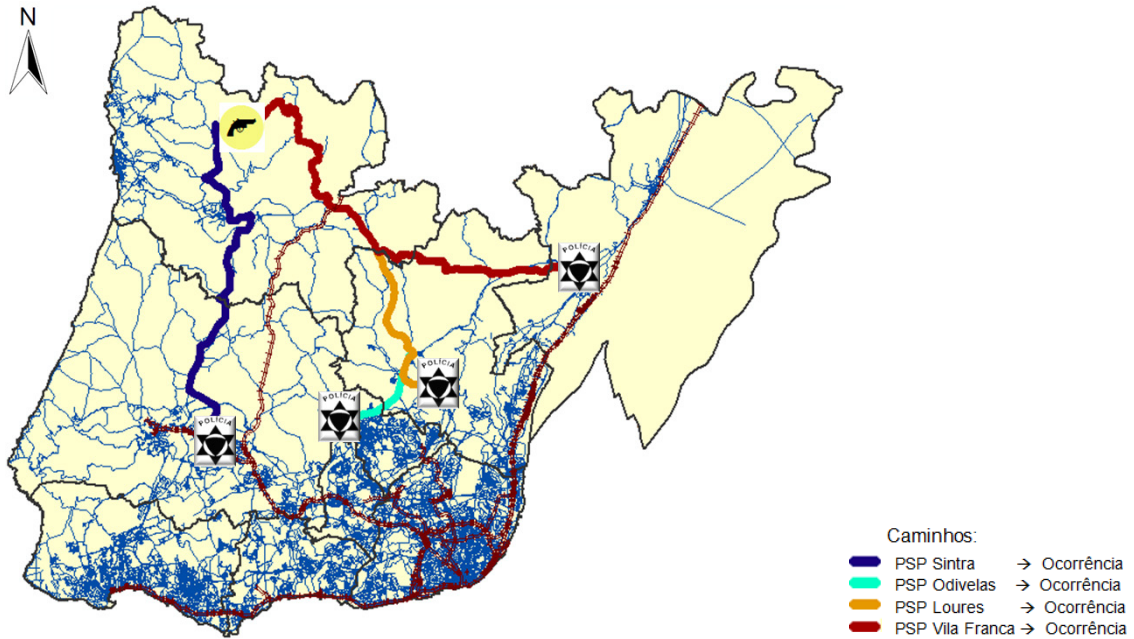


Figura 5.2 - Acessibilidade (*Closest facility*) de cada uma das 4 esquadras mais próximas de Mafra (origem) a uma dada ocorrência de cariz policial (destino) mediante o caminho de menor distância (impedância: comprimento)

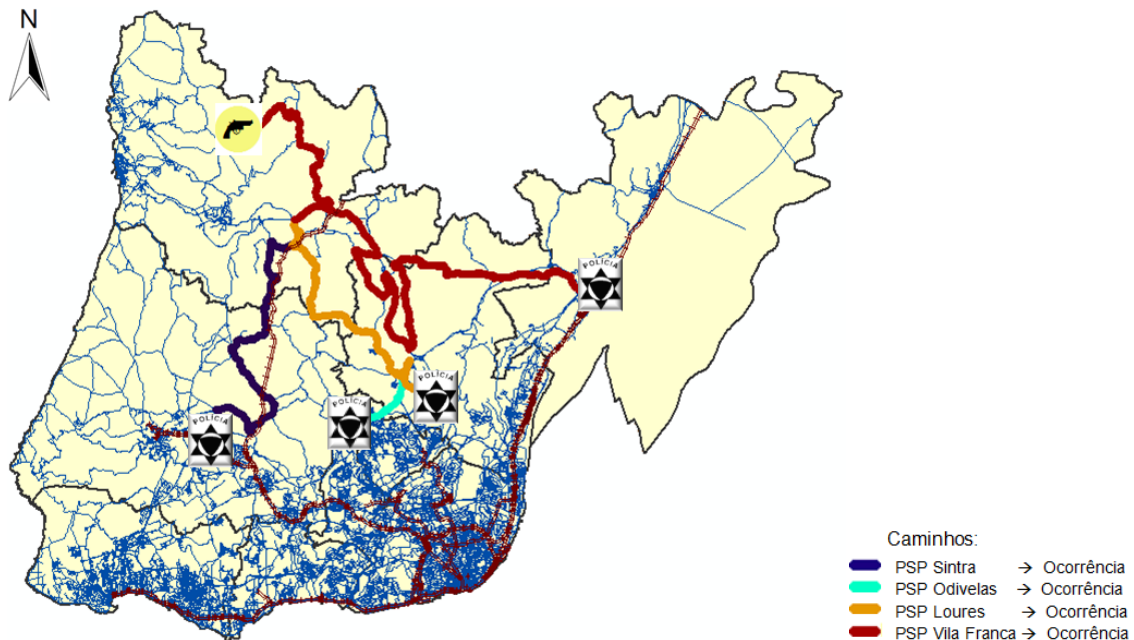


Figura 5.3 - Acessibilidade (*Closest facility*) de cada uma das 4 esquadras mais próximas de Mafra (origem) a uma dada ocorrência de cariz policial (destino) mediante o caminho de menor duração (impedância: tempo)

Verifica-se nas figuras 5.2 e 5.3 que em qualquer dos quatro casos, os trajectos são diferentes consoante se considere o caminho mais rápido ou o caminho mais curto. Os valores obtidos estão dispostos na tabela seguinte.

Tabela 5.3 - Valores obtidos de menor tempo e menor distância dos 4 percursos mediante a impedância escolhida (tempo e distância)

Percursos	Impedância	Tempo (Menor tempo)		Distância (Menor distância)
		Duração (minutos)	Distância (m)	Distância (m)
Percursos	Esquadra PSP Sintra -----> Ocorrência Mafra	127	45063	33368
	Esquadra PSP Odivelas -----> Ocorrência Mafra	104	42565	34269
	Esquadra PSP Loures -----> Ocorrência Mafra	102	38954	31087
	Esquadra PSP Vila Franca --> Ocorrência Mafra	113	62314	36727

Pela tabela 5.3 observa-se que o percurso mais rápido (impedância tempo) e o mais curto (impedância comprimento) são ambos a partir da esquadra de Loures. No caso inverso, avaliando pela duração do percurso, o de maior duração é a partir de Sintra e avaliando pelo comprimento, a maior distância tem origem em Vila Franca de Xira.

Outra particularidade observável na tabela é a duração elevada de cada percurso ao definir a impedância pelo tempo, independentemente do facto de, a rede não rodoviária ter sido excluída, de ter sido respeitado o sentido único e dos valores *TravelTime* terem sido confirmados. Relativamente a este problema foram detectados nos detalhes de cada percurso, identificando os problemas presentes, como assinalado na tabela seguinte.

Tabela 5.4 - Detalhes do percurso de menor duração Esquadra PSP Odivelas → Ocorrência no concelho de Mafra

[-] Route: Esquadra PSP (Odivelas) → Ocorrência em Mafra	38953.9 m	1 hr 43 min
1: Start at Graphic Pick 5		
2: Go northwest on fixme residencial	106.4 m	< 1 min
3: Bear right	70.9 m	4 min
4: Turn left on Avenida de Paris residencial	25.8 m	2 min
5: Make sharp right	1713.2 m	8 min
6: Continue on Rua da Estrela residencial	40 m	7 min
7: Continue	160.7 m	2 min
8: Turn right on Acesso de Sete Rios trunk_link	1333.3 m	< 1 min
9: Turn left on Rua Helena Vaz da Silva residencial	20 m	5 min
10: Continue on Eixo Norte-Sul trunk_link	1478 m	< 1 min
11: Turn right	359.5 m	5 min
12: Make sharp right	3679.1 m	10 min
13: Turn left on Rua Dr Domingos Machado Pereira residencial	4907.3 m	7 min
14: Turn right	39.7 m	6 min
15: Continue on Avenida Marginal residencial	4271.7 m	2 min
16: Turn left	1992.3 m	9 min
17: Turn right	1093.5 m	7 min
18: Turn left on N249-3 trunk	1239 m	1 min
19: Continue on Rua do Com+@rcio residencial	15.5 m	4 min
20: Turn right on Av. General Norton de Matos / 2-- Circular trunk	2634.4 m	3 min
21: Make sharp left	137.4 m	10 min
22: Continue on Cril trunk	6936.8 m	4 min
23: Continue	1939.9 m	< 1 min
24: Turn left on Rua Dia Mundial da Crian+ªa residencial	2081.5 m	1 min
25: Turn left on Avenida do Santo Condest+ível trunk	989.6 m	1 min
26: Continue	1688.5 m	3 min
27: Finish at Graphic Pick 6, on the right		
Total time: 1 hr 43 min		
Total distance: 38953.9 m		

Observou-se que os problemas se focaram nas curvas (intersecções, ou seja, os nós na rede) de forma parcial, visto algumas apresentarem tempos correctos (dependendo da velocidade em vigor num dado troço) assinaladas na verde na tabela, e outras, assinaladas a vermelho, representarem exemplos de casos erráticos mais flagrantes. Na figura seguinte está presente este erro identificado como mais comum na causa para a duração elevada em certos locais da rede.

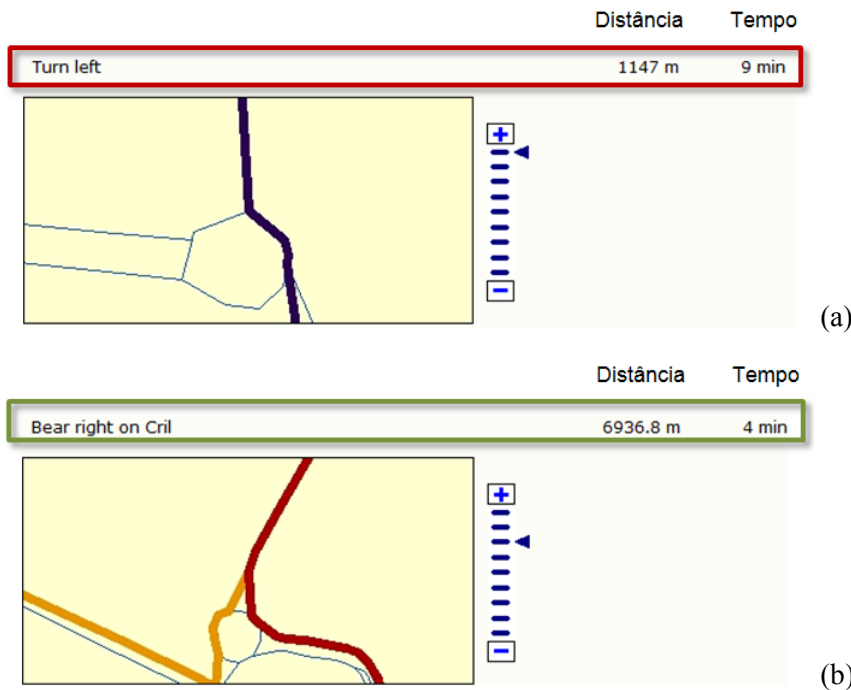


Figura 5.4 - (a) Caso particular de uma rotunda que não interfere erradamente no valor temporal em função do valor de velocidade atribuído a um dado segmento do trajecto. (b) Caso particular de uma rotunda que interfere erradamente no valor temporal em função do valor de velocidade atribuído a um dado segmento do trajecto

Na figura 5.4 é possível analisar dois casos em tudo semelhantes, isto é, retratando rotundas, nas quais a figura (a) representa um caso incorrecto e a (b) um caso correcto. Desta forma consideram-se apenas os trajectos de menor distância como correctos, ou seja, utilizando apenas a impedância tempo, neste e nos restantes algoritmos até ao final deste capítulo. A consequência desta ocorrência poderá ser fruto de uma limitação inerente à topologia arco-nó presente na rede viária utilizada conjugada com o algoritmo utilizado.

Pode-se então concluir que, de acordo com os dados da tabela 5.3 relativamente à impedância distância, a acessibilidade do local da ocorrência é de diferença quase negligenciável para cada uma das quatro esquadras consideradas, na medida em que a diferença entre a de maior e a menor distância percorrida é de apenas cerca de 3000 metros.

5.3.2 – Origin-Destination Cost Matrix

Sendo *Origin-Destination Cost Matrix* um algoritmo que permite determinar a menor distância de um dado número de entidades (origem, representa por cada esquadra PSP) a eventos (destino), primeiro procedeu-se à escolha de uma variável pontual, relevante no contexto deste trabalho. Essa variável a escolher foi a localização das zonas consideradas criminalmente mais problemáticas em toda a região da grande Lisboa, pela Confederação Europeia de Polícia (EuroCOP) mediante dados de 2009 (criminalidade variada, urbanizações clandestinas entre outros parâmetros), cujas zonas são identificadas na figura seguinte.

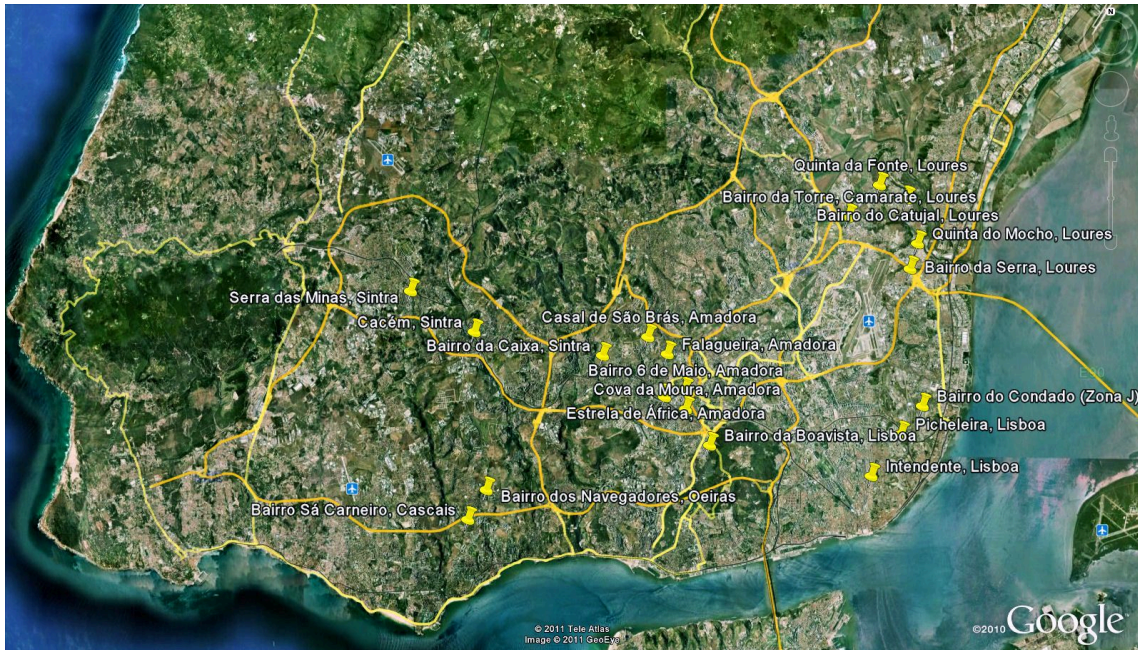


Figura 5.5 - Distribuição e identificação das zonas criminalmente mais problemáticas na região da grande Lisboa

Antes de aplicar o algoritmo foi efectuada uma estimação de densidade *Kernel*, presente na figura 5.6 na qual foram identificadas duas grandes concentrações zonas problemáticas, na Amadora e em Loures. Comparando com o estimador de densidade Kernel das esquadras PSP (figura 2.5) é observável uma relação comum entre as variáveis, na medida em que ambas apresentam o concelho da Amadora como um foco comum de densidade elevada.

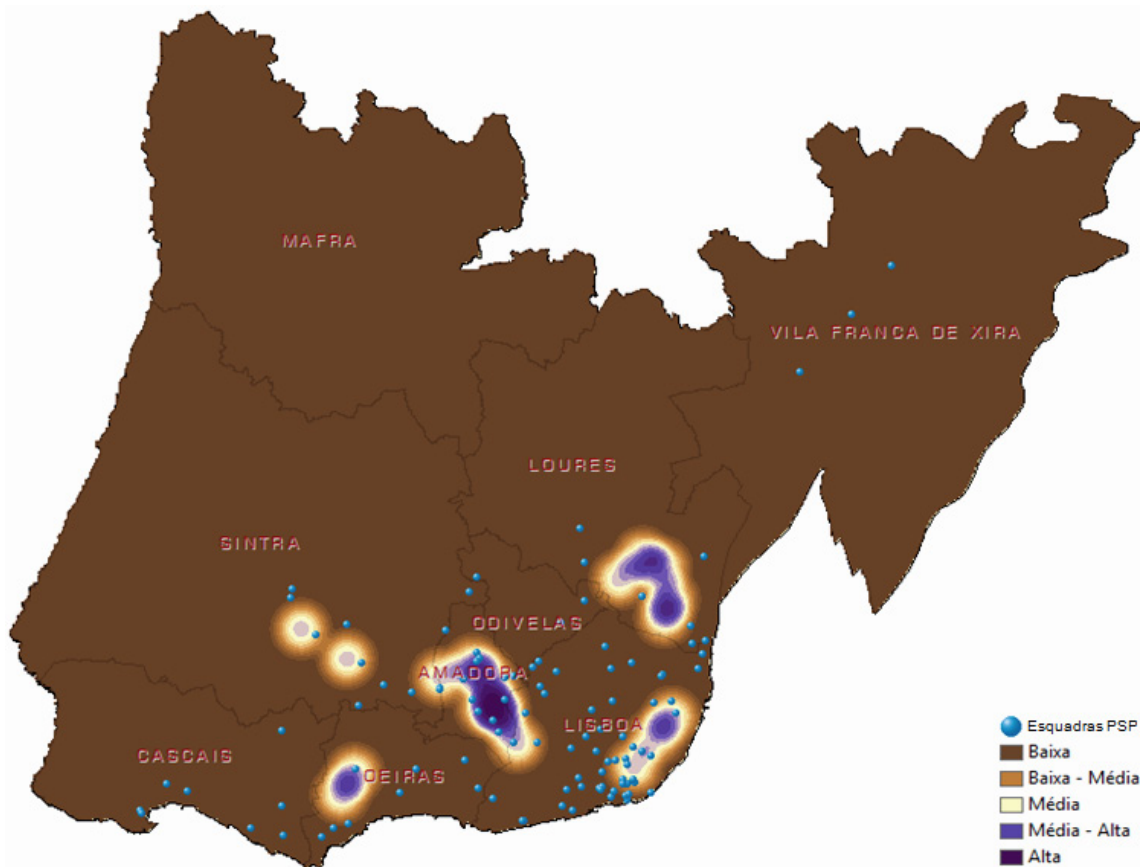


Figura 5.6 - Mapa do estimador de densidade *Kernel* das zonas mais problemáticas na região da grande Lisboa com uma largura de banda de 1800 m, juntamente com a localização pontual das esquadras PSP

Na medida em que o algoritmo de acessibilidade (*closest facility*) a nível global da região de estudo não permite uma visão discernível (representação de uma multitude de pares *facilities*-eventos com *true shapes*) este representa os pares com linhas rectas permitindo uma visão indicadora da relação entre esquadras e os eventos, possibilitando visualizar sobre quais os bairros problemáticos com mais esquadras na sua proximidade (melhor tempo de resposta pela PSP) e vice-versa.

Na figura seguinte são apresentados os pares, nos quais cada esquadra forma um par com o evento mais próximo aliados à acessibilidade de cada esquadra para distâncias de 1000 e 2000 metros.

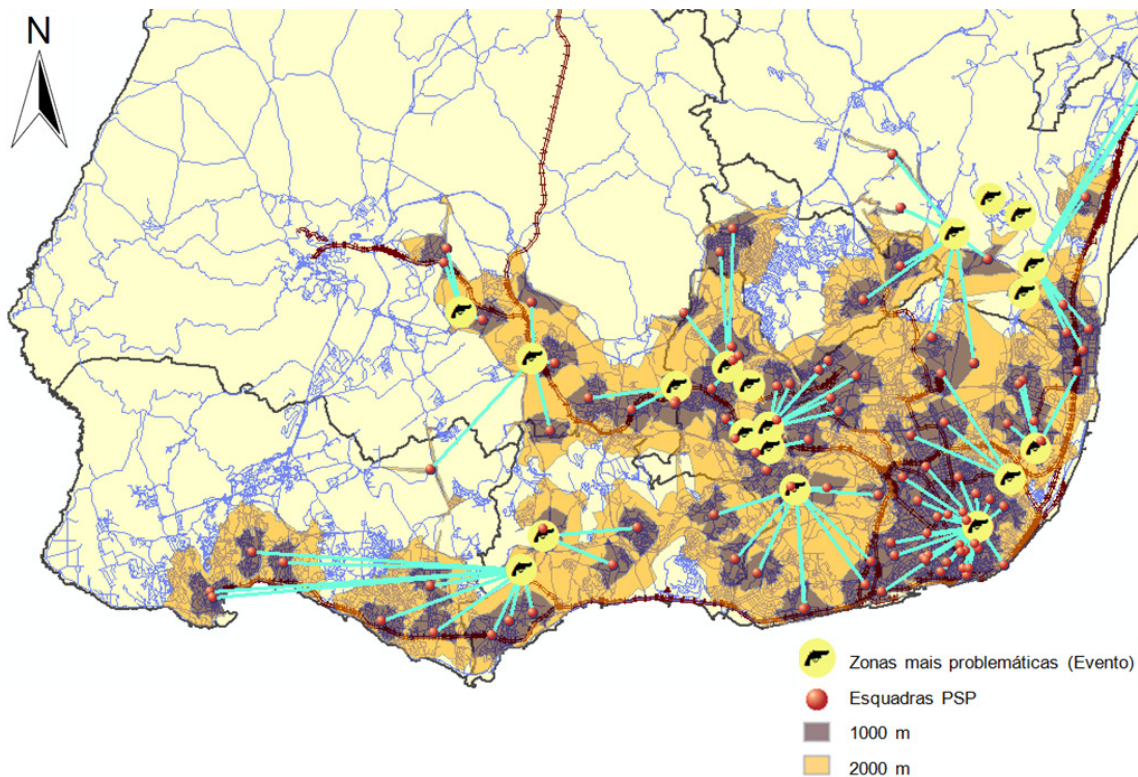


Figura 5.7 – *Origin-destination cost matrix* e acessibilidade (*Service area*) das esquadras PSP (*facilities*) para as zonas mais problemáticas da grande Lisboa (eventos) com impedância definida pela distância, representada por polígonos de 1000 m e 2000 m

É possível reparar na figura 5.7, que a zona de risco correspondente ao Intendente no concelho de Lisboa (figura 5.5) indicia ser a mais “sob controlo” no caso de um eventual motim por exemplo, dado o grande número de esquadras circundantes e a indicação de que, em redor deste bairro, a acessibilidade das esquadras ser exclusivamente de 1000 metros.

5.3.3 – Localização/Alocação (*Service area*)

A implementação deste algoritmo foi efectuada nas duas variáveis pontuais em análise, a referente às esquadras PSP e a respeitante às zonas de risco mais problemáticas, retratada no algoritmo de rede *Origin-Destination Cost Matrix*.

Relativamente às esquadras, a *service area* de cada esquadra foi processada para polígonos de intervalos 1000 e 2000 metros, para as quais são encontrados todos os percursos possíveis na rede viária até estas distâncias. Nesta situação, dado ter-se considerado a direcção do movimento com origem das esquadras foi tida em conta apenas a rede rodoviária.

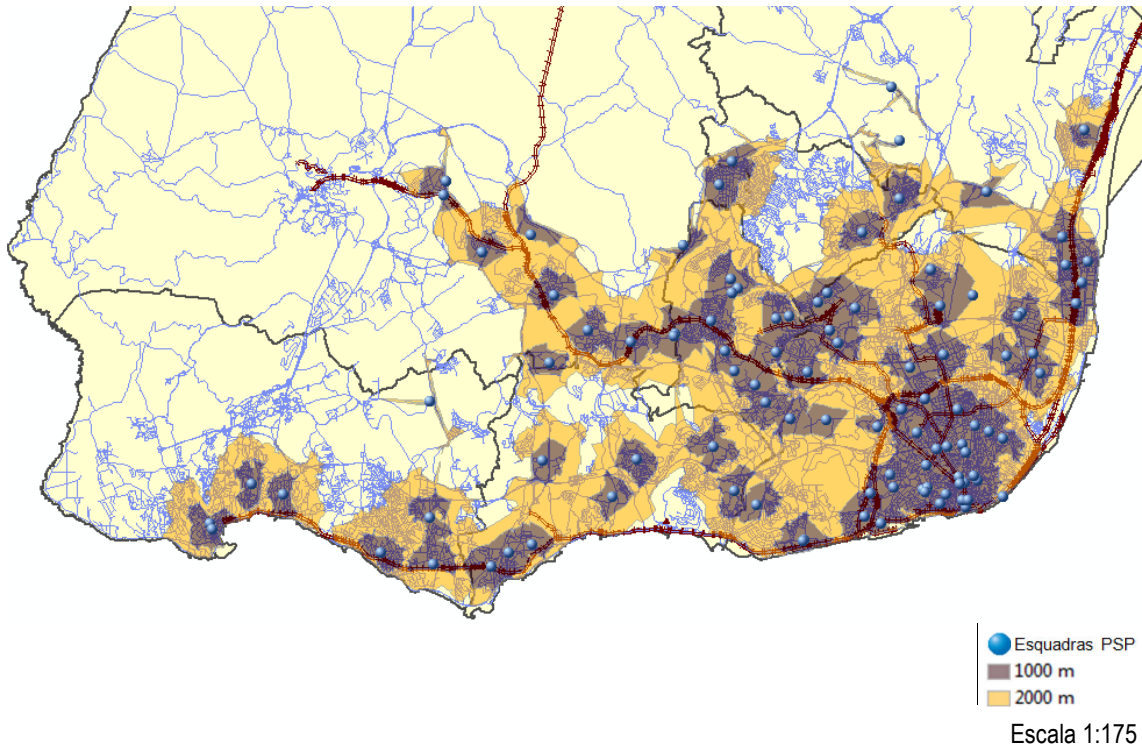


Figura 5.8 - Localização/Alocação (*Service area*) das esquadras da PSP definidas por intervalos de impedância de 1000m e 2000 m

Pela figura 5.8 é conclusivo que é no concelho de Lisboa que a PSP apresenta uma grande área de abrangência (na ordem dos 1000 metros), quando considerando a distância como custo do trajecto (impedância), apresentando os restantes concelhos uma similaridade aos resultados já evidenciados na análise pontual do capítulo 2 (figura 2.5).

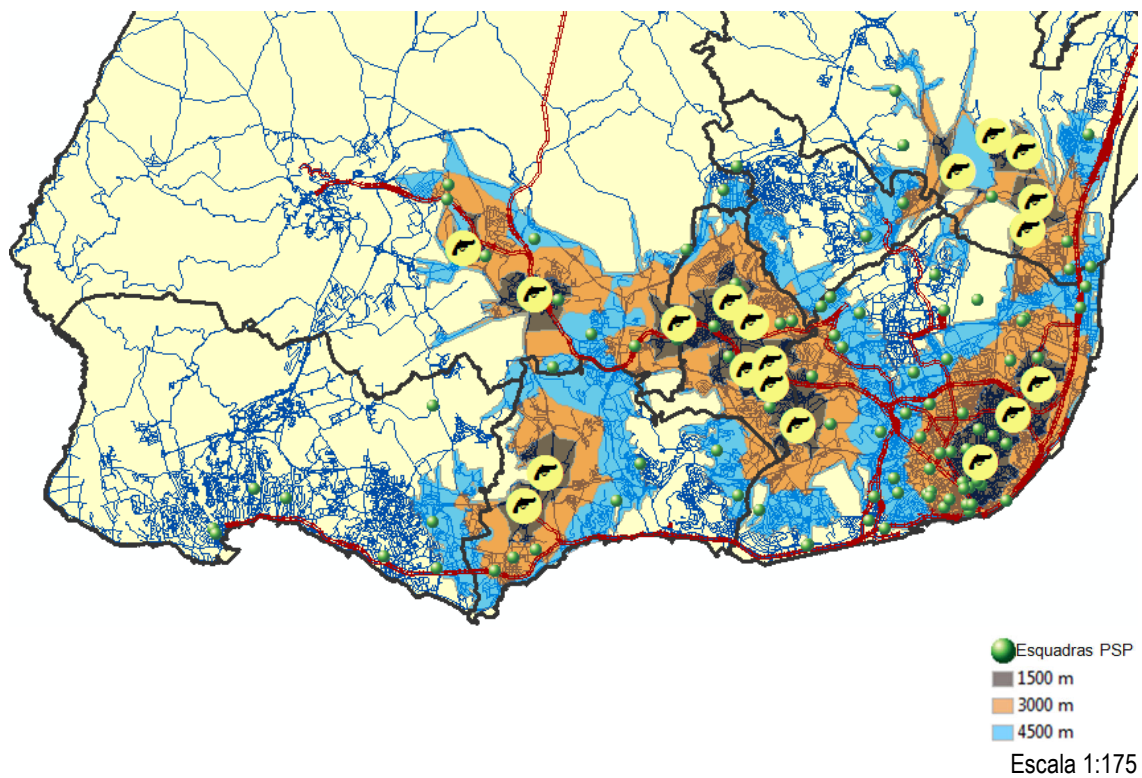


Figura 5.9 - *Service area* (Localização/Alocação) dos bairros mais problemáticos da região definida por intervalos (*break lines*) do atributo distância de 1000 m, 3000 m e 4500 m

A figura 5.9 fornece um indicador quanto à área de abrangência de população residente nos bairros, com foco população e actividades criminais em toda a região da grande Lisboa para distâncias de 1500, 3000 e 4500 metros. De forma a poder analisar de situação prática e menos geral, foi escolhido concelho da Amadora, que como se verificou na figura 5.6 corresponde ao maior foco de zonas de risco, expressa na figura seguinte com o objectivo de verificar as áreas de abrangência destes focos de criminalidade variada, permitindo também uma gestão e optimização percursos de patrulhas da PSP (a pé e em viaturas) nas vizinhanças destes bairros e antevisão de possíveis locais de propagação de actividade criminosa (estações de transportes públicos como CP ou metro nas proximidades destes bairros, etc) ou de movimentações gerais de gangues, por exemplo.

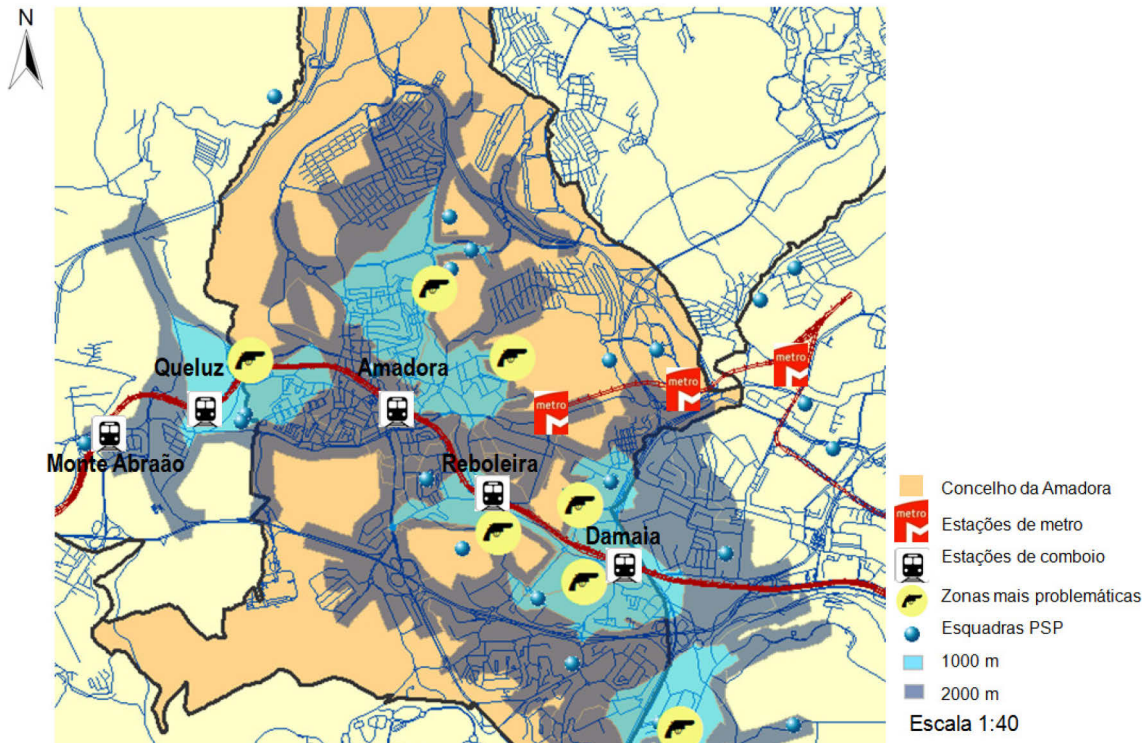


Figura 5.10 - *Service area* (Localização/Alocação) dos bairros mais problemáticos no concelho da Amadora definida por intervalos do atributo distância de 1000 m e 2000 m

Como se verifica pela figura, os bairros mais problemáticos da Amadora localizam-se nas proximidades dos transportes públicos ferroviários, em particular as estações da CP de Queluz, Reboleira e Amadora, todas inseridas numa *service area* de 1000 metros (no caso da Damaia inferior em 300 metros). Para 2000 metros incluem-se as duas estações do metro no concelho da Amadora. Sabendo-se que, os dois bairros mais próximos da estação de comboios da Damaia são precisamente dois dos bairros perigosos mais numerosos na grande Lisboa (aproximadamente 9000 habitantes) e que a esquadra da PSP da estação de comboios de Monte Abraão é a única responsável pela vigilância em toda a linha ferroviária da linha de Sintra com cerca de 50 efectivos [Diário de Notícias, 2010] pode-se concluir que esta estação em particular pode constituir um foco de criminalidade e servir de transporte destas para outras zonas, devendo receber uma maior vigilância por parte das patrulhas das esquadras circundantes relativamente a outras zonas potencialmente menos perigosas.

De referir ainda que, ao contrário do caso prático considerado anteriormente com esquadras, nesta situação foi ignorada a obrigação de respeitar a sentidos dos troços e foi tida em conta toda a rede viária (não apenas rodoviárias), permitindo a consideração de todo o tipo de movimentações, razão pela qual as velocidades adoptadas na tabela 5.1 para troços não rodoviários são em corrida, simulando com maior veracidade uma dada ocorrência criminal.

6. Considerações finais

6.1 – Resultados gerais

O estudo levado a cabo, da análise espacial à distribuição espacial das esquadras da PSP do Comando Metropolitano de Lisboa, permitiu a obtenção de resultados conclusivos.

Foi observado que as esquadras da PSP apresentam um padrão aglomerado centrado no concelho de Lisboa, com uma expansão preferencial ao longo da costa litoral atingindo uma aglomeração máxima para distâncias entre esquadras vizinhas na ordem dos 11000 metros, convertendo-se num padrão aleatório para distâncias entre vizinhas progressivamente superiores, reflectindo a aleatoriedade da localização das poucas esquadras existentes mais afastadas do concelho de Lisboa.

Pelos valores adquiridos resultantes de uma investigação aos dados do tipo área, correspondentes a concelhos, foi demonstrada a distribuição espacial de esquadras da PSP por parte do Ministério da Administração Interna ter como critério principal a densidade populacional, dada a semelhança existente entre ambas as variáveis. De realçar que a localização das zonas categoricamente mais problemáticas, retratadas na análise à rede viária na região de estudo, indiciam ser um factor secundário na localização de esquadras quando na vizinhança destes focos de risco.

Relativamente à investigação e explicação da variável número de crimes por concelho em função das restantes variáveis, foi mostrada a não influência de efeitos espaciais na obtenção do melhor modelo de regressão, a regressão múltipla, que melhor ajustamento apresentou e explicou a variável em causa através do número de esquadras, do número de desempregados e da densidade populacional.

7. Referências Bibliográficas

- AGUIAR, D.; WAGNER S. (2004), *Análise espacial da cana-de-açúcar no estado de São Paulo*, Instituto Nacional de Pesquisas Espaciais, Brasil
- ANSELIN, L. (2005), *Exploring spatial data with GeoDaTM: a workbook*, University of Illinois.
- ANSELIN, L. et al. (2001), *Strutural covariates of U. S. county homicide rates: incorporating spatial effects*. *Criminology*, p.201-232.
- CARNEIRO, E., et al (1999), *Análise especial aplicada na determinação de áreas de risco para algumas doenças endêmicas: O uso de técnicas de geoprocessamento na saúde pública*, Brasil
- CATITA, C. (2012), *Análise Espacial de Informação Geográfica*, Faculdade de Ciências da Universidade de Lisboa.
- DIGGLE, P. (2003), *Statistical Analysis of Spatial Point Patterns, Second Edition*, Oxford University Press, London
- DRUCK, S. et al (2004), *Análise Espacial de Dados Geográficos*, Embrapa, Brasília
- ESRI, (2008), *ArcGis Desktop 9.3*, ArcGis Tutorials.
Disponível em: <http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Tutorials>
- GRIMES, A.; Yun L. (2008), *Spatial Determinants of Land Prices: Does Auckland's Metropolitan Urban Limit Have an Effect?*, Motu Economic & Public Policy Research, Nova Zelândia
- MONGRAIN, J.; JESSICA, L. (2009) *Spatial point pattern analysis applied to bubble nucleation insilicate melts*. *Computers & Geosciences*.
- NETER, J. et al (1996), *Applied Linear Statistical Models*, 3ªEd.
- PONCIANO, P.; JOÃO S. (2010), *Spatial analysis of the dairy yield using a conditional Autoregressive model*, Semina: Ciências Agrárias, Londrina
- PEIXOTO, B., et al (2002), *Criminalidade na região metropolitana de Belo Horizonte: Uma análise espacial*, Brasil
- PRADO, F.; ADILSON, B.; VILMA T.; NILTON I. (2010), *Aplicação e análise de modelos de regressão clássica e espacial para os distritos da cidade de São Paulo*, UNESP, FCT. Brasil
- QUEIROZ M. (2003), *Análise espacial dos acidentes de trânsito do município de Fortaleza*, Brasil
- RAMOS, R, et al (2004), *A comparison of two methods for the definition of Regional Metropolitan Areas through am application in the North of Portugal*, Universidade do Minho, Braga.
- SCALON, C. e CANO, I. (2007), *Desigualdades de classe e desigualdades Espaciais*. XIII Congresso Brasileiro de Sociologia, Brasil.
- SERRANO, R., et al (2000), *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*. 1 ed. Barcelona: Ediciones de la Universitat de Barcelona.
- STARCHER, J. (2007), *Drive-Time Analysis of West Virginia Hospitals using Network Analyst*. West Virginia Health Care Authority