

## **Técnicas de Análise Espacial:**

### **Concentração de nitratos nas águas subterrâneas e Análise de variáveis sociais nos municípios Portugueses**

Paulo Nunes, Aluno nº45444

(fc45444@alunos.fc.ul.pt)

---

#### **RESUMO**

No âmbito da disciplina de Análise Espacial de Informação Geográfica (AEIG), foi desenvolvido o presente estudo destinado à exploração dos algoritmos de geoestatística, análise espacial e de redes disponíveis nas aplicações de Sistemas de Informação Geográfica (SIG).

O presente relatório pretende descrever o trabalho efetuado na análise de padrões pontuais e de variáveis agregadas em áreas, a aplicação de técnicas estatísticas de regressão e correlação espacial, interpolação de dados espaciais e geração de superfícies contínuas, e a aplicação de algoritmos de análise de redes.

Os dados utilizados nas análises de áreas e regressão espacial foram recolhidos no sítio do Instituto Nacional de Estatística, são relativos ao ano 2011, foram trabalhados ao nível geográfico do município. Para a análise de padrões pontuais utilizaram-se dados da rede de monitorização da qualidade das águas subterrâneas, os dados foram recolhidos no sítio da Agência Portuguesa do Ambiente e dizem respeito à série temporal de 1996 a 2012 de um só parâmetro: Nitrato total dissolvido. Para os processos de interpolação utilizaram-se dados de modelos digitais de elevação, dados de gravidade e precipitação para a zona da bacia hidrográfica do Vouga. A análise de redes foi realizada com dados ao nível da freguesia para o município de Lisboa, sobre a densidade de população envelhecida e a localização dos serviços de emergência médica.

Na fase 1 do trabalho, foi realizada a análise de padrões pontuais, no caso, concentrações máximas de nitratos nas reservas subterrâneas de água. Os métodos aplicados descreveram os dados como sendo de aglomerados.

Na segunda fase foram utilizadas técnicas de análise de dados agregados de padrão social ao nível do município. Os métodos aplicados indicaram a presença de aglomerados, de zonas em que as variáveis apresentavam as mesmas características que a sua vizinhança.

Na terceira fase aplicaram-se técnicas de elaboração de modelos simples de regressão clássica e espacial. No fenómeno analisado de crimes de violência doméstica foi possível verificar que os resíduos não apresentavam uma correlação espacial. O teste de Moran permitiu concluir que não é necessário inserir o fator espaço no modelo de regressão.

Na quarta fase realizaram-se operações de interpolação utilizando métodos deterministas e probabilísticos para gerar superfícies contínuas que permitam obter o valor esperado da variável em estudo em toda a área. Nesta fase conclui-se a importância da análise exploratória para a escolha do método a utilizar.

Na quinta e última fase realizou-se a análise da localização de serviços de emergência médica com base em algoritmos de redes que permitem verificar se a localização dos hospitais e corporações de bombeiros correspondem às necessidades das freguesias de Lisboa com população mais envelhecida.

Em todas as fases dos trabalhos concluímos que as aplicações SIG oferecem um conjunto de ferramentas que permitem efetuar análises de apoio a estudos sociais e das características físicas fundamentais para a gestão das atividades humanas.

---

#### **1. Introdução**

A análise espacial é uma disciplina que utiliza as capacidades de computação e análise instaladas nos sistemas de informação geográfica (SIG) para extrair informação sobre os fenómenos com referência espacial.

Os estudos descritos neste relatório pretendem explorar técnicas e algoritmos de tratamento de informação geoespacial e sua distribuição para produção de informação sobre os fenómenos.

Este trabalho irá desenvolver-se em dois grupos. No primeiro, pretende-se efetuar uma análise de padrões pontuais relativos à distribuição de contaminantes nos lençóis freáticos (reservatórios de água subterrânea). Neste caso, analisam-se os níveis de concentração de nitratos (NO<sub>3</sub> em mg/l) na

rede de monitorização de águas subterrâneas da Agência Portuguesa do Ambiente. Os Nitratos são contaminantes resultantes da atividade agrícola e da utilização de adubos químicos. As suas concentrações são objeto de legislação europeia, nomeadamente a Diretiva Quadro da Água. (Parlamento Europeu, 2000). Nessa diretiva são apresentados os limites máximos admissíveis para a concentração de nitratos nas águas subterrâneas. No caso deste poluente os limites admissíveis da sua concentração são de 50 mg/l de acordo com Anexo 1 da Diretiva 2006/118/CE (Parlamento Europeu, 2006). No segundo grupo, pretende-se realizar a análise de variáveis agregadas por áreas, desenvolver estudos de correlação entre várias variáveis e autocorrelação espacial. Nos diversos métodos vão ser utilizados dados sociais do Instituto Nacional de Estatística e na Base de Dados Portugal Contemporâneo. O nível geográfico de agregação será o município. Os dados das variáveis sociais são associados às áreas do município presente na carta administrativa oficial de Portugal. O ano de referência utilizado no estudo é 2011, data em que ocorreram os últimos censos, por uma questão de uniformidade metodológica utilizou-se a CAOP 2011 referente ao mesmo ano.

O relatório tem por objetivo demonstrar resultados e documentar as técnicas de análise espacial em ambiente SIG, pelo que, optou-se pela apresentação dos mapas, gráficos e tabelas nos locais do texto onde se descrevem as técnicas. Por uma questão puramente metodológica o relatório não apresenta um capítulo individualizado de discussão dos resultados sendo a análise crítica realizada no decurso do texto.

## **2. Metodologia**

### **2.1 Fluxo de trabalho**

A análise de dados científicos (com referência espacial e temporal) e abstratos é uma atividade desempenhada por todos os domínios científicos constituindo a base da fundamentação dos fenómenos que ocorrem. A estatística é uma das principais ciências que se ocupa do tratamento de dados, porém, as técnicas utilizadas não preveem uma análise dos dados num enquadramento do espaço, e muitas vezes desprezar a sua influência conduz a análises incompletas.

A análise de dados espaciais aparece para colmatar esta limitação tendo desenvolvido metodologias próprias apoiadas em sistemas de informação geográfica que permitem a análise de dados com referência espacial (coordenadas num sistema de referência) e utiliza algoritmos de geoestatística para extrair informação de padrões de distribuição espacial numa vertente global e local. Exemplos deste tipo de análise: a análise de surtos de doenças contagiosas, análise dos focos de criminologia, análise do padrão de dispersão de substâncias contaminantes, etc...

A análise espacial de dados baseia-se na lei de Waldo Tobler que pressupõe que «que todas as coisas são parecidas, mas coisas mais próximas parecem-se mais do que coisas mais distantes» (Catita, 2014). Ou seja, a maioria dos fenómenos, sejam eles sociais ou naturais, tem uma ocorrência motivada pelas características do espaço e suas variações.

A análise espacial distingue-se das análise designadas não espaciais porque prevê a observação das variáveis no contexto do espaço geográfico, procurando encontrar a dependência entre as variáveis e os diversos fatores que justificam o seu padrão (Fischer e Wang, 2011).

O principal objetivo da análise de dados espaciais é a construção de modelos através dos quais se possam inferir os valores das variáveis a partir de outras conhecidas e exercer uma ação preditiva sobre os fenómenos. Por exemplo: predição da posição de ocorrência de tempestades e eventos extremos, e a sua relação com a ocorrência e distribuição de determinadas variáveis climáticas.

No âmbito deste trabalho vamos adotar a classificação dos dados espaciais seguida pelo Instituto Espacial de Pesquisas Espaciais (Druck, Carvalho *et al.*, 2004):

- Eventos ou Padrões pontuais: Fenómenos expressos através de ocorrências identificadas como pontos localizados no espaço. São exemplos: localização de crimes, localização de acidentes, localização de espécies vegetais, localização de doenças.
- Áreas contínuas: Estimadas a partir de um conjunto de amostras de campo que podem estar regular ou irregularmente distribuídas. Este tipo de dados resultam de levantamentos de recursos que incluem mapas geológicos, topográficos, ecológicos, etc...
- Áreas com contagens e taxas agregadas: Representam variáveis agregadas ao nível de uma determinada fronteira, normalmente são dados associados a levantamentos populacionais, como censos e estatísticas de saúde, e que originalmente se referem a indivíduos localizados no espaço. Por razões de confidencialidade, os dados são agregados e apresentados como atributos dos polígonos que refletem uma determinada organização administrativa.

No trabalho pretende-se analisar dados do tipo pontual e variáveis agregadas por áreas, fica para uma fase posterior a análise espacial de superfícies criadas a partir de processos de interpolação. No estudo vão ser exploradas as diferentes técnicas e algoritmos que atualmente estão disponíveis nas aplicações SIG. Nas análises vão ser utilizadas as seguintes ferramentas: ArcGIS 10.2 (ESRI), Quantum GIS (QGIS ORG) e Geoda (Arizon State University).

A análise de dados espaciais segue o fluxo genérico que é apresentado na Fig. 1.

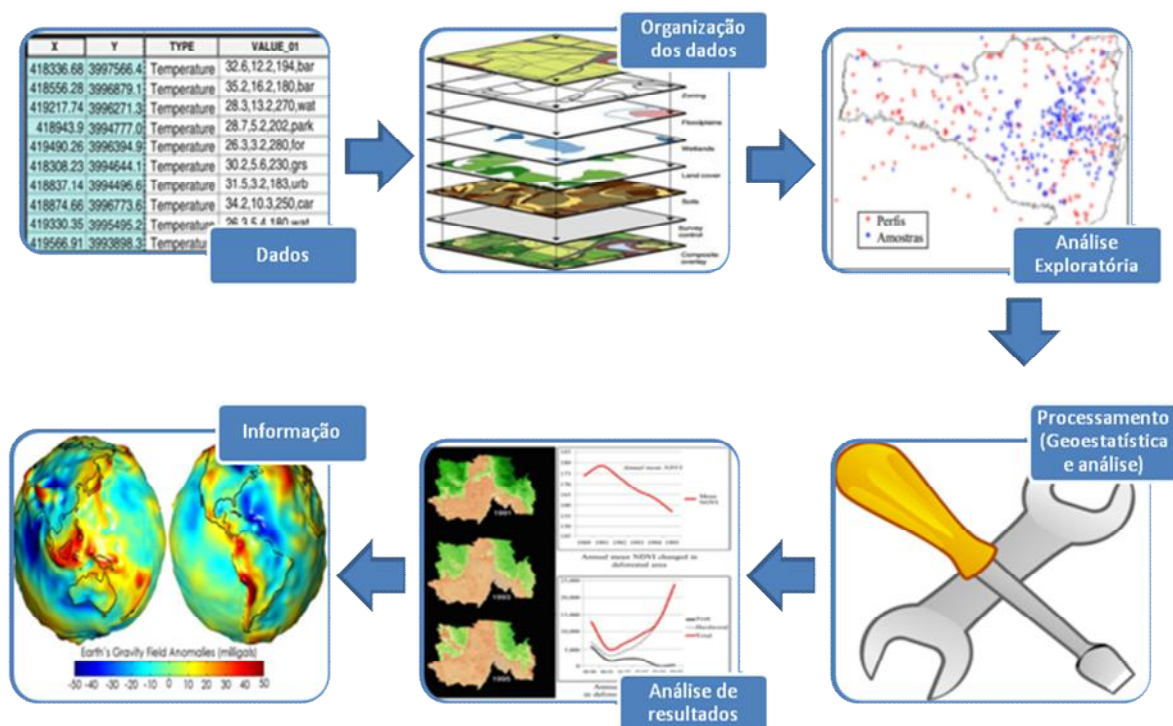


Fig. 1 – Esquema genérico do fluxo de trabalho em análise espacial

## 2.2. Fontes de dados

As fontes de dados utilizadas nas análises descritas neste trabalho encontram-se listadas na Tabela I.

Tabela I – Fontes de dados utilizadas na análise espacial

Dados	Fonte de dados	Data de recolha	Observações
Dados da concentração de nitratos nas águas subterrâneas do continente.	Sistema Nacional de Informação de Recursos Hídricos da Agência Portuguesa do Ambiente: <a href="http://snirh.apambiente.pt">http://snirh.apambiente.pt</a>	Março de 2014	A rede compreende 1060 estações, tendo sido possível compilar informação de valores médios e máximos de 963 pontos de amostragem situados em Portugal Continental
Estatísticas relativas a variáveis sociais para os municípios de Portugal Continental	Instituto Nacional de Estatística <a href="http://www.ine.pt/">http://www.ine.pt/</a> PORDATA <a href="http://www.pordata.pt/">http://www.pordata.pt/</a>	Março de 2014	
Carta administrativa Oficial de Portugal 2011	Direção-Geral do Território <a href="http://www.igeo.pt/produtos/cadastro/caop/versao2011.htm">http://www.igeo.pt/produtos/cadastro/caop/versao2011.htm</a>	Março de 2014	
CORINE (mapa de ocupação do solo)	Direção-Geral do Território <a href="http://www.igeo.pt/e-IGEO/egeo_downloads.htm">http://www.igeo.pt/e-IGEO/egeo_downloads.htm</a>	Março de 2014	CORINE referente a 2006

### 3. Análise padrões pontuais, dados agregados e regressão espacial

#### 3.1. Fase 1 - Padrões Pontuais

Nesta fase irão ser aplicadas técnicas de análise espacial sobre eventos discretos. Estes dados apresentam fenómenos com ocorrência não contínua no espaço. São caracterizados por uma distribuição de pontos com referência espacial (coordenadas num determinado sistema de referência). São exemplos deste tipo de eventos a localização de crimes, ocorrências de doenças e localização de espécies vegetais ou amostras geológicas. O objetivo destas técnicas de análise de padrão pontual é estudar a distribuição espacial da amostra de pontos, inferindo sobre a aleatoriedade do padrão pontual. Para tal, testa-se a hipótese estatística de uma distribuição aleatória (normal ou gaussiana) *versus* um padrão aglomerado ( *et al.*, 2004)<sup>1</sup>.

##### 3.1.1 – Recolha e tratamento dos dados

Os dados pontuais utilizados nesta etapa dizem respeito às medições de concentrações de nitratos (mg/l) nas águas subterrâneas recolhidas em 963 estações que fazem parte do sistema de monitorização. Os dados estão disponíveis em linha no sítio da Agência Portuguesa do Ambiente e fazem parte do Sistema Nacional de Informação de Recursos Hídricos (Agência Portuguesa do Ambiente, 1995).

De acordo com a informação recolhida na legislação europeia, o signatário verificou que os limites máximos admissíveis para a concentração de nitratos em águas subterrâneas é de 50 mg/l. Os dados dizem respeito a parâmetros estatísticos de uma série temporal no período entre 1996 e 2012. Na fase de exploração dos dados verificou-se que a metodologia de recolha de informação não prevê uma amostragem em todas as estações no mesmo momento temporal, e a frequência de amostragem é diferente de estação para estação. Verificando essa questão, existiam dois caminhos possíveis para a análise dos dados: (1) Optar por tratar os parâmetros estatísticos da série temporal; (2) Selecionar a última amostra em cada estação. O signatário optou pela primeira abordagem de forma a poder analisar todas as estações em conjunto. O parâmetro estatístico que foi utilizado foi o valor máximo da série em cada estação. O signatário parte do princípio empírico de que tendo sido registados valores superiores aos admissíveis em determinada estação, isto constitui um indício de que naquele local existem condições propícias à contaminação das águas subterrâneas, o que justifica uma monitorização e acompanhamento mais atento da qualidade da água.

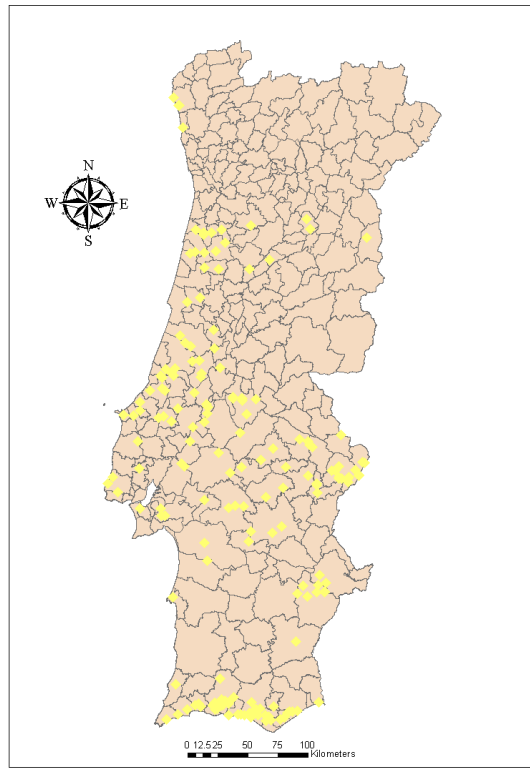
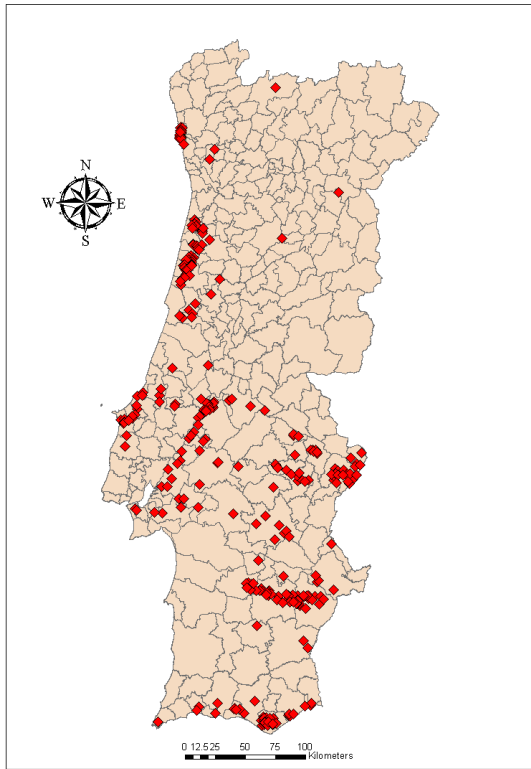
De acordo com a informação que foi possível consultar (Ministério da Agricultura do Desenvolvimento Rural e das Pescas, 1997), a contaminação da água por nitratos resulta em grande parte de adoção de más práticas na atividade agrícola. A título de exemplo: a adubação incorreta dos terrenos, o controlo incorreto dos resíduos orgânicos provenientes das explorações agrícolas de produção agro pecuária e má gestão dos resíduos provenientes dos detritos vegetais em culturas de regadio como o arroz.

##### 3.1.2 – Visualização de dados

Para análise do padrão pontual optou-se por separar a amostra em três grupos: estações que apresentaram concentrações máximas maiores ou iguais a 50 mg/l (limiar da legislação europeia e nacional), valores entre 25 e 50 mg/l e valores inferiores a 25 mg/l.

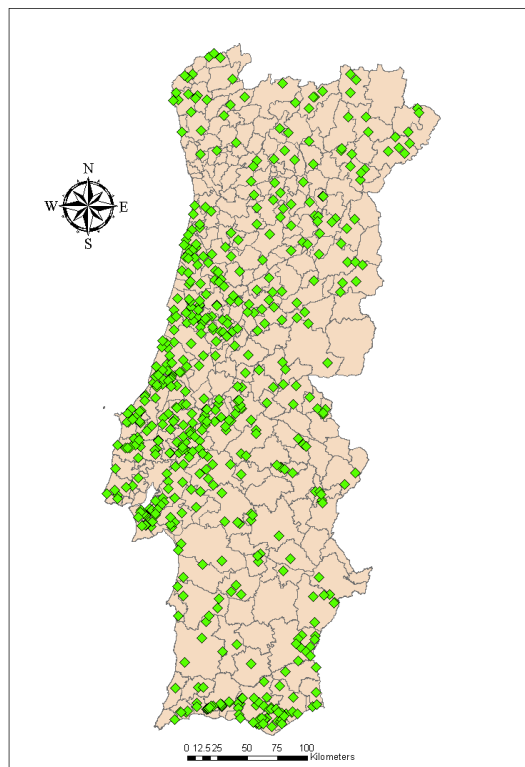
---

<sup>1</sup> O capítulo do livro relativo à análise de padrões pontuais pode ser consultado em linha no sítio do Instituto Nacional de Pesquisas Espaciais em: <http://www.dpi.inpe.br/gilberto/livro/analise/cap2-eventos.pdf> (acedido em abril de 2014).



a) Posições com concentração máxima superior ou igual a 50 mg/l

b) Posições com concentração máxima 25-50 mg/l



c) Posições com concentração máxima inferior a 25 mg/l

Fig. 2 – Visualização dos padrões pontuais relativos à distribuição da concentração de nitratos nas águas subterrâneas

Através dos mapas resultados da projeção dos dados pontuais visualmente verifica-se que existe uma distribuição espacial que indicia a existência de *clusters* da contaminação de águas subterrâneas por nitratos.

Tendo em conta o que é referido no Código de boas práticas, foi adicionada a informação de zonas agrárias identificadas no mapa de ocupação de solos CORINE do ano 2006<sup>2</sup>. Este tipo de visualização permite verificar se as amostras contaminadas correspondem a zonas agrárias de determinado tipo.

Visualmente constata-se que existe uma distribuição espacial do padrão pontual das estações com colheitas de amostras superiores ao limítar de 50 mg / l coincidente com áreas de culturas temporárias (referentes aos seguintes níveis CORINE: 211- Culturas temporárias de sequeiro, 212 – Culturas temporárias de regadio e 213 – Arrozais.) As zonas de culturas temporárias encontram-se representadas a verde na **Error! Reference source not found.**

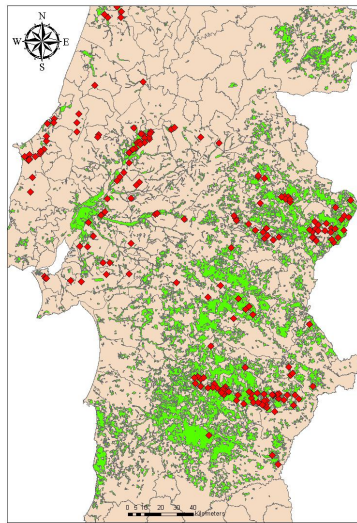


Fig. 3 – Mapa das zonas de culturas temporárias e pontos de concentração máxima de nitratos superior a 50 mg/l

Após a visualização da informação realizou-se a análise exploratória recorrendo a descritores<sup>3</sup> e detetores<sup>4</sup> estatísticos. A análise de padrão pontual baseia-se na identificação de padrões de aglomerados ou a aleatoriedade espacial dos fenómenos em estudo. O objetivo é utilizar os sistemas de informação geográfica para encontrar o motivo que justifica uma determinada distribuição.

Nas análises efetuadas foram utilizadas as seguintes técnicas de análise de padrão pontual: de 1.<sup>a</sup> ordem (análise global) e de 2.<sup>a</sup> ordem (análise local).

---

<sup>2</sup> Foi utilizado o mapa CORINE 2006 porque apresenta informação mais detalhada em três níveis de informação e o de 2007 apresenta dois. Para esta análise, tendo em conta que é um trabalho académica e que não existem mapas de ocupação do solo de acesso livre mais recente, considera-se que a ocupação do solo existente em 2006 é descritiva das variáveis para o período da análise.

<sup>3</sup> Os descritores de padrões de pontos destinam-se a apresentar uma perceção rápida de que forma os pontos se distribuem no espaço.

<sup>4</sup> Os detetores de padrões de pontos permitem o reconhecimento e medição do padrão de uma distribuição de pontos.

1. Técnicas de análise exploratória:
  - Elipses de distribuição espacial;
2. Técnicas de análise exploratória de 1.<sup>a</sup> Ordem (globais ou de grande escala):
  - Contagem por quadrantes;
  - Kernel density;
3. Técnicas de análise dos efeitos de 2.<sup>a</sup> Ordem (locais ou de pequena escala):
  - Análise do vizinho mais próximo;
  - Função K

A análise exploratória dos dados através da elaboração de elipses padrão de distribuição espacial utilizou-se a ferramenta *Spatial Statistics Toolbox* do *ArcToolbox*<sup>5</sup> (ESRI, 2013).

Para o caso da concentração de nitratos na água subterrânea, a visualização do padrão pontual apresenta aglomerados distribuídos espacialmente, pelo que se o cálculo da elipse de distribuição padrão fosse criada com toda a informação estaríamos a tirar conclusões incorretas. Este é um exemplo do cuidado que deve existir na aplicação das técnicas de análise espacial. Para separar os diferentes focos de interesse utilizámos a informação de que a concentração de nitratos resulta da atividade agrícola e da existência de culturas. Dessa forma, criaram-se as três elipses de distribuição espacial que são apresentadas na Fig. 4. Não faz parte dos objetivos deste trabalho apresentar o formulário que serve de base ao cálculo das elipses, porém fica o registo das principais quantidades calculadas: coordenadas X e Y do centro geométrico da distribuição, desvio padrão nos eixos e o ângulo de rotação. A forma da elipse resultante indica se estamos na presença de uma distribuição uniforme em todas as direções ou com uma distribuição espacial com uma tendência numa direção definida.

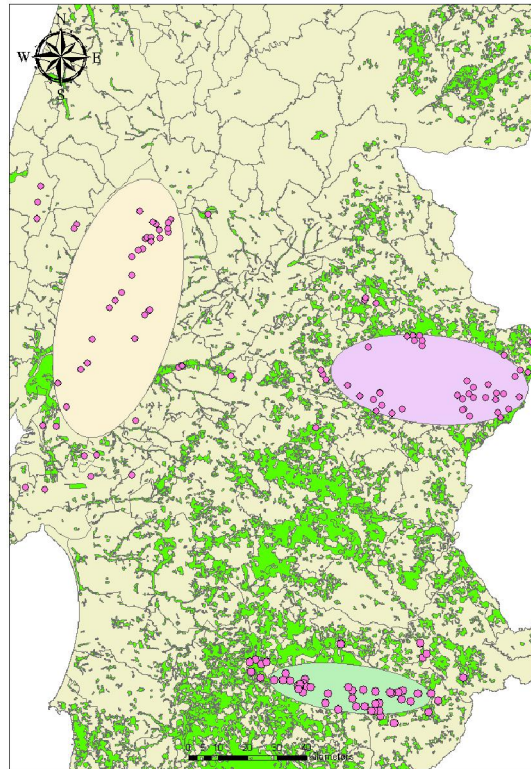


Fig. 4 – Elipses de distribuição padrão espacial (1 sigma)

Após fazer a análise exploratória recorrendo às elipses padrão de distribuição espacial foram aplicados métodos de análise globais.

### 3.1.3 – Métodos de análise de efeitos de primeira ordem

O primeiro método utilizado foi a contagem por quadrantes. Esta metodologia permite inferir a natureza da amostra através de dois descritores estatísticos.

<sup>5</sup> Para mais informação sobre o funcionamento do algoritmo do ArcGIS 10.2 ver: [http://resources.arcgis.com/en/help/main/10.1/index.html/](http://resources.arcgis.com/en/help/main/10.1/index.html#/) e [http://resources.arcgis.com/en/help/main/10.1/index.html#/How\\_Directional\\_Distribution\\_Standard\\_Deviational\\_Ellipse\\_works/005p0000001q000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Directional_Distribution_Standard_Deviational_Ellipse_works/005p0000001q000000/) (acedido em abril de 2014)



O método de contagem por quadrantes baseia-se na determinação de basicamente duas quantidades descritivas a média e a variância (*Variance to mean ratio*).

$$Média = \frac{NumeroPontos}{NumeroQuadrantes} \qquad VTMR = \frac{Variância}{Média}$$

Através da *Variance To Mean Ratio* (VTMR) é possível inferir a natureza do padrão (Catita, 2014):

VTMR > 1 : O padrão apresenta uma natureza de aglomerado. Este valor indica que existem conjuntos de dados agrupados e existem áreas extensas sem pontos (variabilidade espacial grande).

VTMR < 1 : O padrão apresenta uma natureza de dispersão regular. Este valor indica que os eventos estão distribuídos de forma regular por toda a região.

VTMR = 1 : O padrão apresenta uma natureza aleatória. Este valor indica que os eventos não tendem para aglomerados nem para dispersão.

Após calcular o VTMR torna-se necessário testar a hipótese de estarmos na presença de uma amostra aleatória utilizando a distribuição de qui-quadrado.

Para o teste utilizam-se os seguintes parâmetros:

$N_{gl}$  : Número de graus de liberdade (número de quadrantes - 1).

$H_0$  : Existir uma amostra aleatória

$H_1$  : Existir uma amostra de aglomerados

Variável de teste =  $N_{gl} * VTMR$

No caso da amostra em questão foi elaborada uma grelha com células de 20 km X 20 km. O retângulo que serviu de base à elaboração da grelha apresenta a seguinte configuração:

Coordenadas PTTM06 do canto inferior esquerdo: X= -150 000 / Y= -300 000.

Largura do retângulo = 300 000 m / Altura do retângulo = 600 000 m

Número de linhas = 30 / número de colunas = 15

Número de quadrantes com 20 km X 20 km = 450 células (quadrantes).

No caso da amostra dos padrões pontuais referentes às três distribuições da concentração de águas subterrâneas obtiveram-se os seguintes resultados:

Fig. 5 – Tabela Resumo dos resultados do método de contagem por quadrantes

<b>Amostra</b>	<b>Nr. de pontos</b>	<b>Nr. de quadrantes</b>	<b>VTMR</b>
Valores de concentração máxima de nitratos maior ou igual a 50 mg / l	316	450	9.9940
Valores de concentração máxima de nitratos maior que 25 mg / l e menor que 50 mg / l	179	450	3.7754
Valores de concentração máxima de nitratos menor que 25 mg / l	565	450	4.6329

Para demonstrar o método iremos realizar o teste de significância à segunda amostra (padrão pontual correspondente a valores de concentração máxima de nitratos maior que 25 mg / l e menor que 50 mg / l.

Número de graus de liberdade: 449

Valor da variável de teste =  $449 * 3.7754 = 1\ 695.15$

Vamos considerar um nível de significância de 0.05

Valor da tabela de qui-quadrado (Calculado no excel) = 400.87

Aceitar a hipótese nula corresponde a ter a variável teste inferior ao valor da distribuição qui-quadrado com 449 graus de liberdade e nível de significância de 0.05. Neste caso, com base no teste, podemos dizer que rejeitamos a hipótese nula com uma confiança de 95 %. Ou seja, estamos na presença de uma amostra de aglomerados.

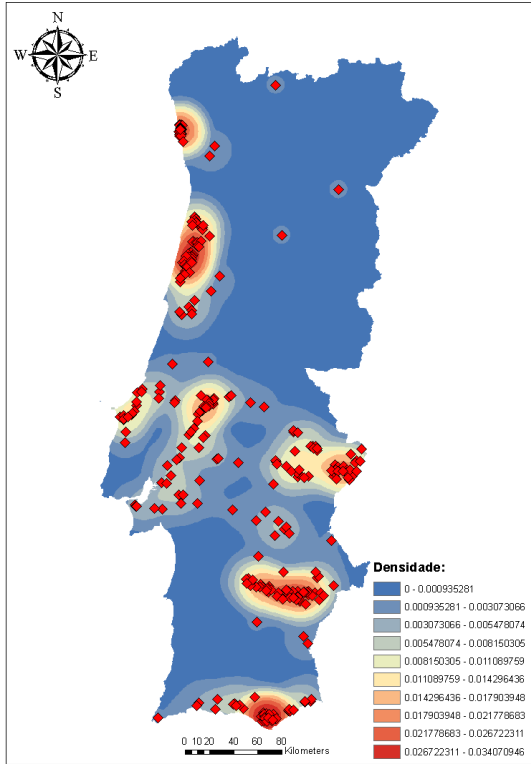
A segunda metodologia de análise dos efeitos de primeira ordem foi a elaboração de *Kernel Density* utilizando a ferramenta *Spatial Analyst Toolbox > Kernel Density*<sup>6</sup>. O mapa de densidade resultante reflete a densidade na vizinhança de uma célula. De acordo com o manual do ArcGIS a ferramenta Kernel density segue a formulação matemática proposta em (Silverman, 1986, p. 76). A informação resultante do mapa de densidade gerado são as áreas onde um determinado conjunto de pontos ou objetos estão concentrados. Do mapa de densidade deduzimos a distribuição do fenómeno por uma determinada região. Estes mapas são particularmente úteis para gerar um modelo simples de distribuição dos fenómenos.

As imagens apresentadas na Fig. 6 foram geradas com o *Kernel Density Tool* configurado para originar um *raster* com células de 500 metros e para um raios de busca de 30 000 metros.

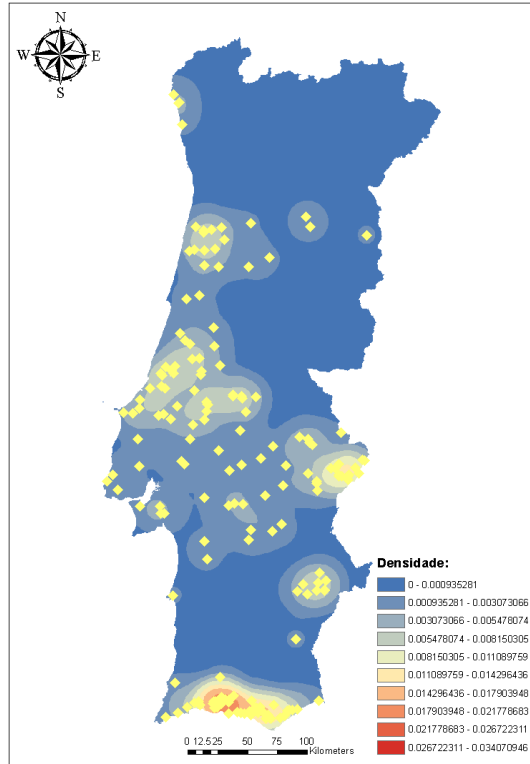
Os resultados obtidos com esta metodologia são os apresentados de seguida:

---

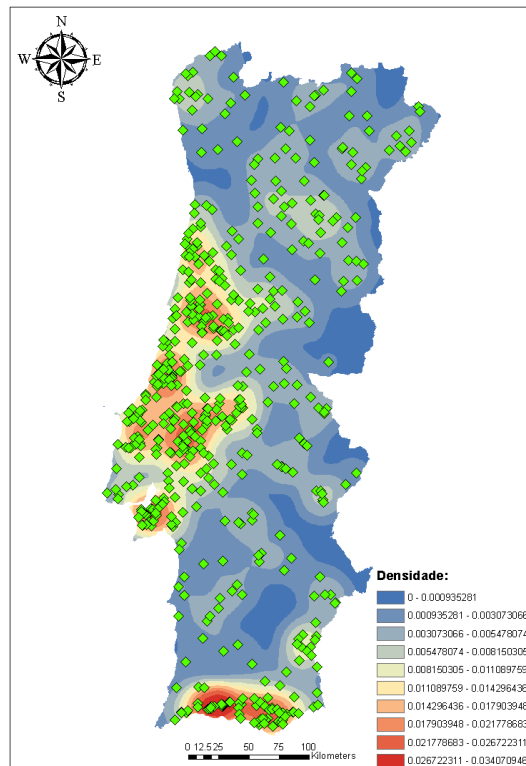
<sup>6</sup> Para mais informação sobre o Kernel Density ver: [http://resources.arcgis.com/en/help/main/10.1/index.html#/How\\_Kernel\\_Density\\_works/009z0000011000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Kernel_Density_works/009z0000011000000/) (acedido em abril de 2014)



a) Posições com concentração máxima igual ou superior a 50 mg/l



b) Posições com concentração máxima 25-50 mg/l



c) Posições com concentração máxima inferior a 25 mg/l

Fig. 6 – Raster com o resultado da aplicação do Kernel Density Tool aos pontos com as três classes de concentração de nitratos

Após aplicar os dois métodos de análise de efeitos de primeira ordem apresentados na teoria passou-se à aplicação de métodos de análise de efeitos de segunda ordem, ou locais. Estes métodos são baseados na distância entre os eventos. Os métodos que foram aplicados foram: o método do vizinho mais próximo, a função G, a função F e a função K.

### 3.1.4 – Métodos de análise de efeitos de segunda ordem

Os métodos de análise de efeitos de segunda ordem baseiam-se na distância entre os diferentes pontos procurando inferir a natureza da sua distribuição local.

O método do vizinho mais próximo baseia-se na estimação da distribuição dos eventos através da distância entre os pontos na região de interesse. Os métodos normalmente utilizam a distância euclideana entre os pontos e a determinação é feita de ponto a ponto.

O Método do vizinho mais próximo baseia-se na análise de função G determinada de forma cumulativa. O gráfico da função G dá informação sobre a forma como os eventos estão distribuídos na região de interesse. Um gráfico que aumenta rapidamente em pequenas distância indica um padrão de aglomerados. Um gráfico que aumenta lentamente até aproximadamente à distância entre eventos e depois aumenta rapidamente indica que os eventos apresentam um padrão de dispersão.

A função G no ArcGIS está implementada através da *Spatial Statistics Toolbox* > *Average Nearest Neighbor*<sup>7</sup>. Após o algoritmo correr sobre os dados é elaborado um relatório que apresenta dois descritores estatísticos que devem ser analisados: o *z-score* e o *p-value*.

O *z-score* e o *p-value* apresentam a medida de significância estatística que indica se a hipótese nula deve ou não ser rejeitada. A hipótese nula neste algoritmo é de estarmos na presença de uma amostra com uma distribuição aleatória.

O *p-value* indica a probabilidade de que o padrão pontual seja aleatório. Um valor elevado indica um padrão aleatório e um valor pequeno indica um padrão não aleatório. O *z-score* é o desvio padrão da amostra. Os dois descritores indicados estão associados à distribuição gaussiana.

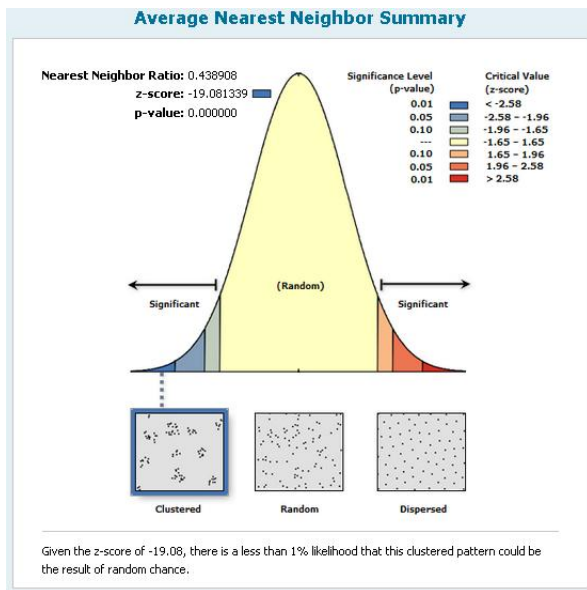
O manual do ArcGIS apresenta uma tabela que ajuda a interpretação dos resultados do teste estatístico de hipóteses com a hipótese zero sendo da Complete Spatial Randomness (CSR):

Tabela II – Interpretação dos resultados do teste estatístico através do algoritmo *Average Nearest Neighbor*<sup>8</sup>

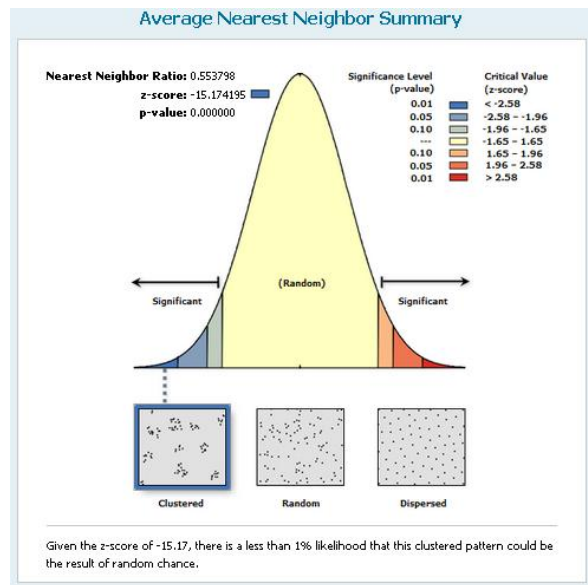
z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%

<sup>7</sup> Informação sobre o algoritmo de Average Nearest Neighbor pode ser encontrada em linha no sítio: <http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000008000000> (acedido em abril de 2014).

<sup>8</sup> Para mais informação sobre o *z-score* e *p-value* ver: <http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000006000000> (acedido em abril de 2014)



a) Distância Euclideana



a) Distância Manhattan

Fig. 7 – Resultados do algoritmo *Average Nearest Neighbor* à amostra de pontos que apresentam concentração máxima de nitratos superior ou igual a 50 mg/l.

Os resultados obtidos (Fig. 7 – z-score de -19.08 e p-value de 0.00) indicam que a hipótese nula (CSR) pode ser rejeitada com um nível de confiança de 99%. Neste caso estamos na presença de um padrão de agregado.

A função F é igual à função G mas usa um subconjunto de pontos. Não iremos analisar a função F que apresenta resultados semelhantes aos da função G.

A função K distingue-se das anteriores porque não analisa só a distância mínima entre dois eventos mas a soma da distância entre todos os eventos situados numa determinada região.

A função K no ArcGIS está implementada através do algoritmo *Spatial Statistics Toolbox > Multi-Distance Spatial Cluster Analysis (Ripleys K Function)*<sup>9</sup>.

Para entender um pouco melhor os resultados que apresentamos de seguida recorreremos à explicação dos resultados do algoritmo que é apresentada no manual do ArcGIS

<sup>9</sup> Para mais informação sobre a Função K implementada no ArcGIS ver: <http://resources.arcgis.com/en/help/main/10.1/index.html#//005p000000m00000> (acedido em abril de 2014).

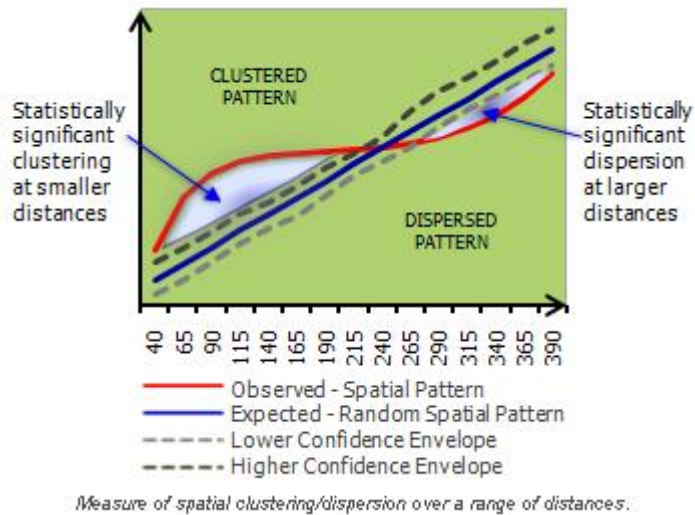
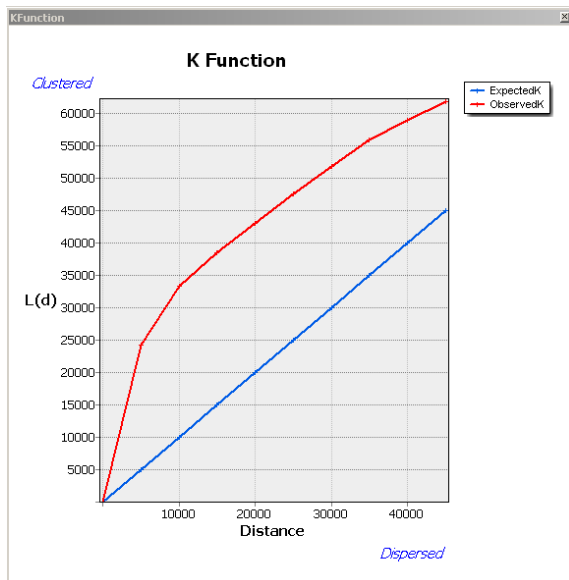


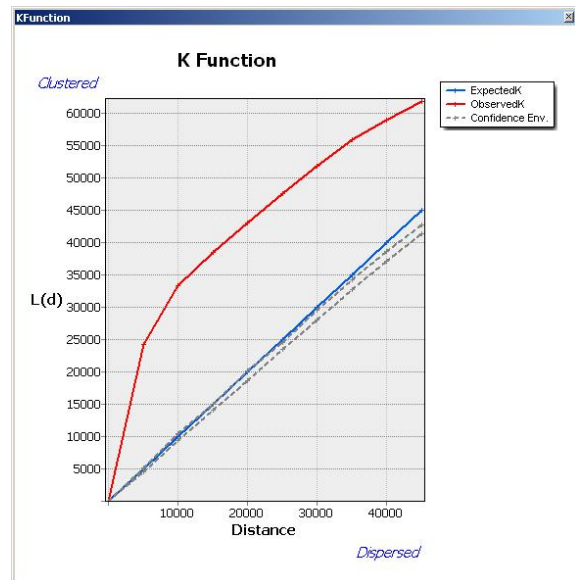
Fig. 8 – Diagram explicativo dos resultados da função  $K^{10}$

O algoritmo da Função K baseia-se na análise comparativo entre o gráfico teórico de uma amostra com uma distribuição aleatória. Essa amostra teórica é comparada com o padrão pontual resultante da amostra que se pretende estudar. Quando o gráfico da amostra a analisar cai na área superior, é considerado um padrão com tendência para o aglomerado quando cai na área inferior estamos na presença de uma padrão com tendência para a dispersão.

Todos os algoritmos de análise espacial e estatística são sensíveis à configuração que é adotada. De seguida apresentamos os resultados referentes a várias configurações do algoritmo da função K para demonstrar essa sensibilidade à configuração. Nos primeiros dois gráficos (Fig. 9 a) e b)) verificamos que até à distância de 50 km estamos na presença de um padrão de aglomerado. Nos dois gráficos seguintes (Fig. 9 c) e d)) verificamos que até à distância de 75 km estamos na presença de um padrão aglomerado, sendo de dispersão a partir dessa distância.



a) Sem permutações / 10 bandas / Distância de



b) 9 permutações / 10 bandas / Distância de

<sup>10</sup> Ver: <http://resources.arcgis.com/en/help/main/10.1/index.html#//005p000000m00000> (acedido em abril de 2014).

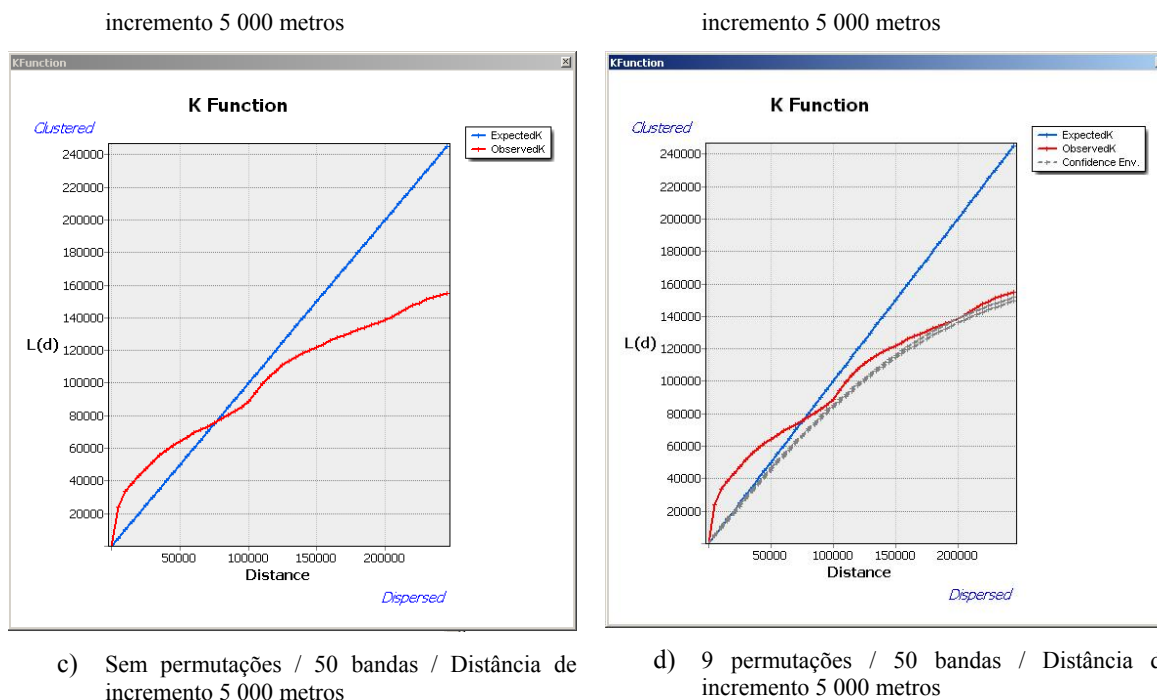


Fig. 9 – Resultados da função K

Com a função K conclui-se a demonstração de técnicas de análise espacial de padrões pontuais. As técnicas utilizadas mostraram o padrão de aglomerados com uma distribuição espacial, comprovando o que tinha sido observado visualmente na fase de prospeção dos dados.

### 3.2. Fase 2 – Análise de Áreas

Nesta fase existe o objetivo de efetuar análise de dados espaciais cuja localização está associada a áreas delimitadas por polígonos. Este caso é frequente quando se pretende uma análise de dados agregados por uma determinada unidade de estudo ou administrativa (distrito, município, freguesia, NUT, zona agrícola, etc...). Uma das aplicações deste tipo de análise é o estudo de dados que requerem agregação para manter a confidencialidade dos mesmos. Neste caso não existe o conhecimento da localização exata dos eventos, mas sim da sua medida por área. ( *et al.*, 2004)<sup>11</sup>

Neste e nos próximos sub-capítulos iremos explorar as técnicas de análise da autocorrelação espacial e regressão espacial. Os dados utilizados na fase inicial não eram adequados à análise de áreas pelo que iremos utilizar dados de padrões sociais da população portuguesa analisando como unidade geográfica de base o município (organização administrativa de Portugal). Os dados foram recolhidos no Instituto Nacional de Estatística e base de dados Portugal Contemporâneo (PORDATA). Os dados utilizados dizem respeito aos últimos censos realizados em Portugal em 2011. Por uma questão de uniformidade temporal da metodologia optou-se por utilizar a CAOP2011 embora existam cartas administrativas mais recentes. Se a análise tivesse sido feita ao nível da freguesia seria obrigatório utilizar a CAOP 2011 visto terem existido alterações da organização administrativa de Portugal após essa data. Nesse caso a hierarquia da organização administrativa seria um exemplo do problema de unidade de área modificável (*Modifiable area*

<sup>11</sup> O capítulo do livro relativo à análise de padrões pontuais pode ser consultado em linha no sítio do Instituto Nacional de Pesquisas Espaciais em: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf> (acedido em abril de 2014).

*unit problem - MAUP*). O problema de unidade de área modificável surge quando para a mesma amostra (População) obtemos resultados condicionados pela escolha das fronteiras da área. Embora reconhecendo o problema, no âmbito deste trabalho não iremos mostrar metodologias para resolver o problema MAUP e otimizar a análise.

Na CAOP foram organizadas as seguintes variáveis ao nível do município para fazer as análises apresentadas de seguida:

Média de número de desempregados [2011]

Número de Crimes [2011]

Número de crimes de violência doméstica [2011]

Número de crimes de furto de veículos [2011]

Número de crimes de furto em residências [2011]

Número de crimes de furto em espaços comerciais [2011]

Densidade populacional (Nº/km<sup>2</sup>) [2011]

Consumo de gás

Consumo de energia elétrica [2011]

Número de divórcios [2011]

Salário médio [2011]

Saldo migratório total [2011]

Saldo migratório natural

Saldo migratório [2011]

Distribuição de água [2011]

Número de habitantes por escalão etário [2011]

Número de habitantes por sexo [2011]

Na fase 2 iremos aplicar as seguintes técnicas à análise das variáveis anteriores:

Visualização de dados espaciais agregados por área

Métodos de corte

Gráficos de médias e mediana

Análise exploratória de dados (AED)

Indicadores globais de autocorrelação espacial

Indicadores locais de autocorrelação espacial (LISA)

Significância estatística

Associar aos índices da AED distribuições teóricas conhecidas

Teste da pseudo-significância (geração de permutações)

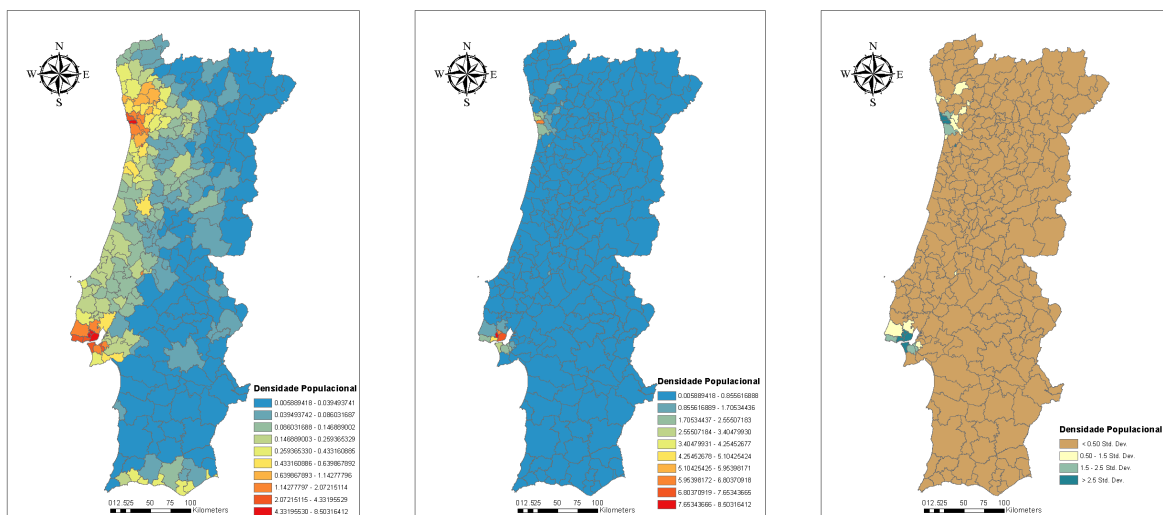
### **3.2.1 – Visualização de dados espaciais agregados por área:**



As técnicas de análise exploratória destinam-se a perscrutar os dados em busca de valores extremos, para fazer a busca utilizam ferramentas gráficas (histogramas e *boxplots*) e parâmetros estatísticos simples. A análise segue duas abordagens. (1) cada polígono é considerado uma unidade única limitada pelas suas fronteiras e não relacionada com a sua vizinhança; (2) analisa-se a autocorrelação espacial de uma determinada variável, neste caso analisam-se os polígonos e a sua vizinhança.

Iremos começar a análise exploratória dos dados utilizando métodos de corte de variáveis: por intervalos iguais, percentis e desvios padrão. Os métodos de exploração de dados através da visualização baseiam-se na construção de mapas que mostrem a variação da variável numa determinada área. No caso em que é necessário analisar várias variáveis é importante adotar processos de normalização, por uma questão metodológica de manutenção da proporção dos valores das variáveis adotou-se a normalização por percentagem do valor máximo (Razão entre o valor do polígono e o valor máximo da amostra).

Para efeitos de demonstração das possibilidades associadas à visualização enquanto método de análise iremos mostrar as possibilidades de visualização de uma variável e alguns parâmetros estatísticos básicos relativos à variável densidade populacional ( $n^\circ / km^2$ ):



a) Visualização da densidade populacional por município (Método de classificação por intervalos naturais)

b) Visualização da densidade populacional por município (Método de classificação por intervalos iguais)

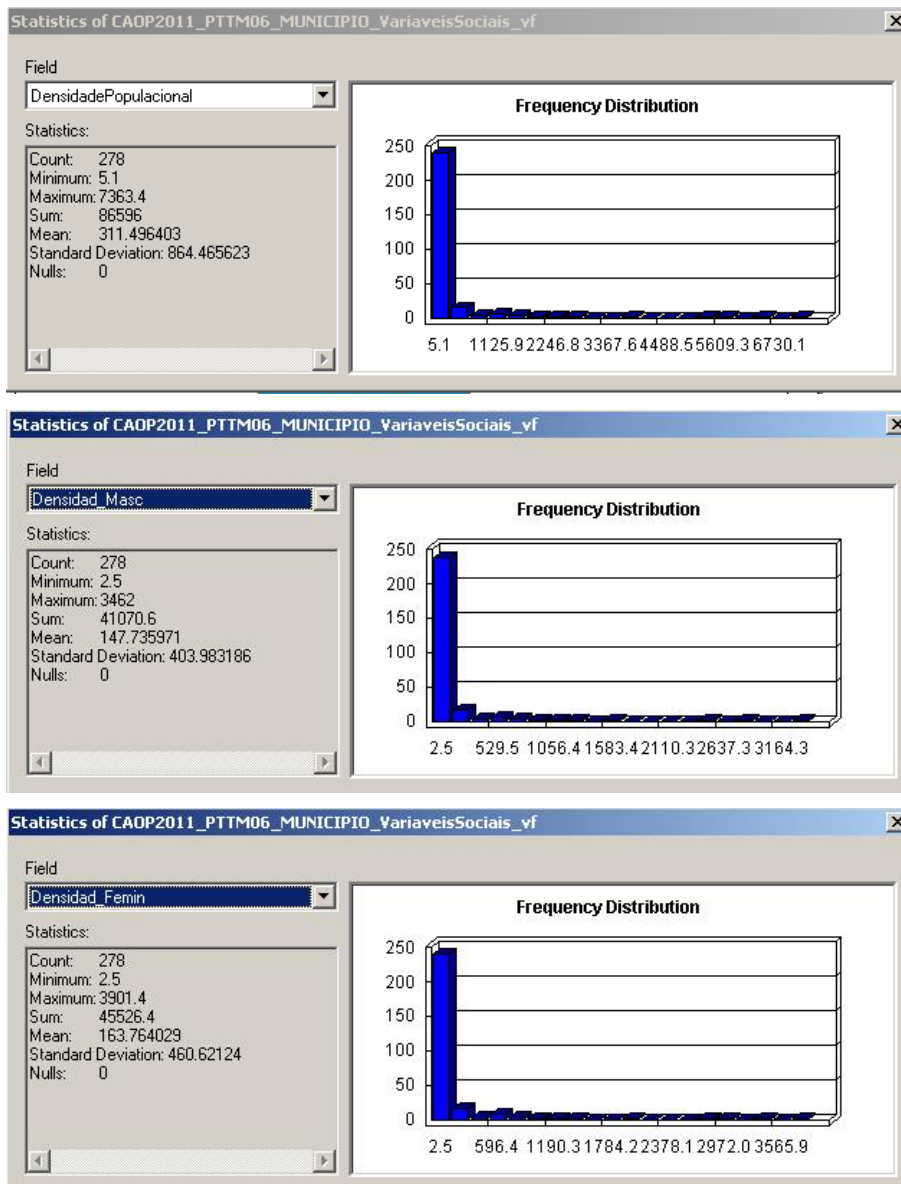
c) Visualização da densidade populacional por município (Método de classificação por Desvio Padrão)

Fig. 10 – Visualização da variável densidade populacional ao nível do município (ano 2011)

Como é possível observar pelos diferentes resultados que são obtidos, uma escolha e configuração criteriosa da visualização dos dados permite destacar diferentes características e detetar visualmente a concentração e distribuição espacial dos valores da variável. Para fazer este tipo de análise é importante ter em conta as diferenças de área entre os polígonos. Para fazer uma análise correta é necessário ter uma área de referência que possa normalizar a amostra. Neste caso estamos a utilizar o número de habitantes por quilómetro quadrado e não o número de habitantes por área do município.

Da análise dos mapas anteriores verificamos que tal como esperado a maior densidade populacional situa-se na zona litoral e mostra o contraste existente em Portugal Continental entre o Litoral e o Interior.

De seguida são apresentados alguns dos parâmetros estatísticos da variável densidade populacional ao nível do município:



a) Densidade populacional total (n.º / km<sup>2</sup>)

b) Densidade populacional de habitantes masculinos (n.º / km<sup>2</sup>)

c) Densidade populacional de habitantes femininos (n.º / km<sup>2</sup>)

Fig. 11 – Parâmetros estatísticos relativos à densidade populacional (n.º / Km<sup>2</sup>)

Os parâmetros estatísticos anteriores foram colocados para demonstrar um tipo de análise estatística que nada permite inferir sobre a natureza espacial da amostra. Este tipo de análise é simplista e ao contrário das ferramentas de visualização dos dados em SIG nada permitem concluir sobre a distribuição espacial da variável.

### 3.2.2 – Análise exploratória dos dados (AED):

No sentido de avaliar a autocorrelação espacial global, foi utilizado o índices de Moran I, que compara o valor da variável com o valor da mesma variável na sua vizinhança.

Para fazer a análise utilizando a autocorrelação espacial criou-se uma tabela de atributos com as diversas variáveis de estudo. A tabela foi associada à CAOP2011 de forma a criar uma *shapefile* que pudesse ser utilizada na análise utilizando as ferramentas disponíveis na aplicação GeoDa.

O primeiro passo para fazer a análise consiste em gerar uma matriz de proximidade espacial ou de vizinhança. A matriz de vizinhança neste caso foi criada no GeoDa. A aplicação apresenta dois critérios para a elaboração da matriz. O critério de contiguidade e da distância ao centróide do

polígono. Os municípios apresentam formas complexas e áreas muito diferentes, esse fator foi o motivo que levou o signatário a optar pela criação de uma matriz de proximidade baseada na contiguidade. Neste caso, a matriz de vizinhança considera vizinhos os polígonos que partilham as fronteiras (vizinhança de primeira ordem).

Depois de criada a matriz de vizinhança efetuaram-se as análises de autocorrelação espacial. Sendo este um trabalho focado na aplicação de técnicas de análise espacial iremos analisar várias variáveis na aplicação dos métodos.

A autocorrelação espacial global da variável (crimes de violência doméstica) foi analisada com a ferramenta *univariate Moran's I* do GeoDa e obtiveram-se os seguintes resultados:

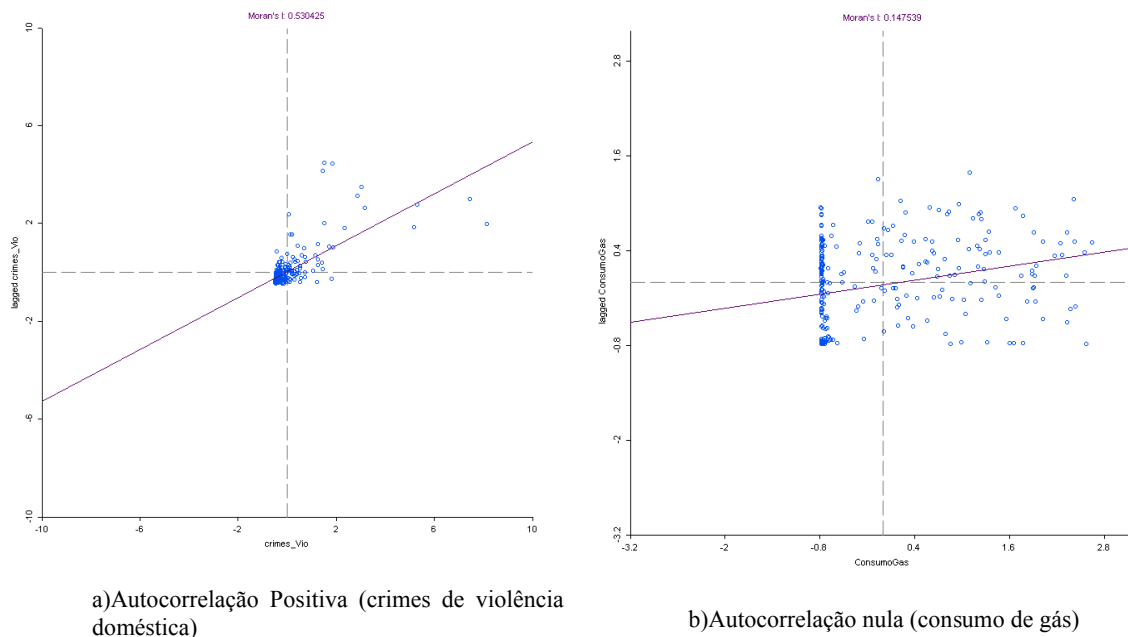


Fig. 12 – Resultados do algoritmo de autocorrelação espacial global (Moran's I do GeoDa)

Dos resultados apresentados na Fig. 12 conclui-se que para a variável do número de crimes de violência doméstica estamos numa situação de autocorrelação positiva e na variável de consumo de gás estamos na presença de uma situação de aleatoriedade (embora o coeficiente de Moran seja 0.15 o gráfico apresenta uma grande dispersão, ou seja, poderiam ser passadas muitas retas com diferentes declives na amostra).

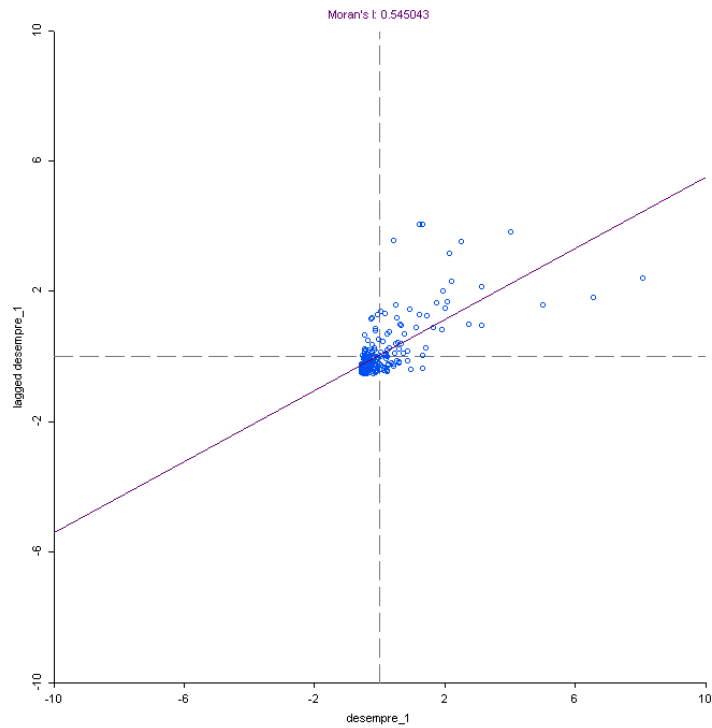
A tabela seguinte ajuda a interpretar os resultados do coeficiente de Moran (Catita, 2014):

Tabela III – Significado do coeficiente de Moran

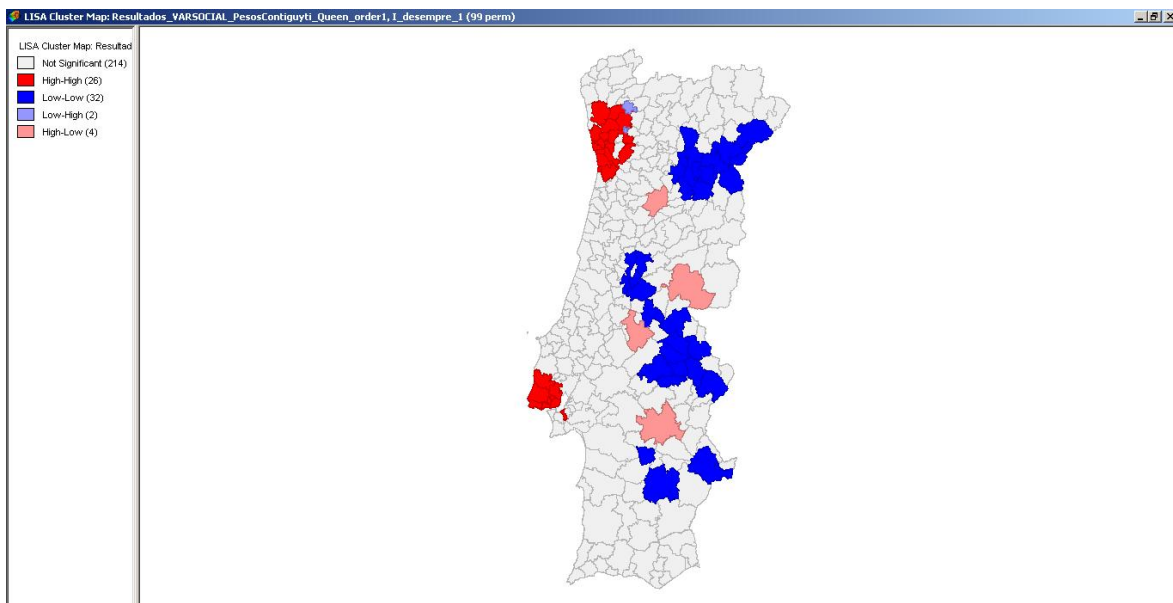
Valor do Coeficiente de Moran	Conclusões a tirar do coeficiente:
1	Autocorrelação espacial positivo ou direta (valores altos (baixos) tendem a estar na vizinhança de valores altos (baixos). Valores próximos de 1 indicam agrupamentos.
0	Significa que estamos perante a aleatoriedade da distribuição da variável.
-1	Autocorrelação espacial negativa ou inversa (valores altos (baixos) tendem a estar na vizinhança de valores baixos (altos). Valores próximos de -1 indicam uma situação de dispersão.

Depois de analisada a autocorrelação espacial global efetuou-se a aplicação do algoritmo do índice local de Moran (à variável de número médio de desempregados). Com esta ferramenta são calculados os *Local Indicators of Spatial Autocorrelation* (LISA) baseados na covariância do valor da variável nos polígonos vizinhos.

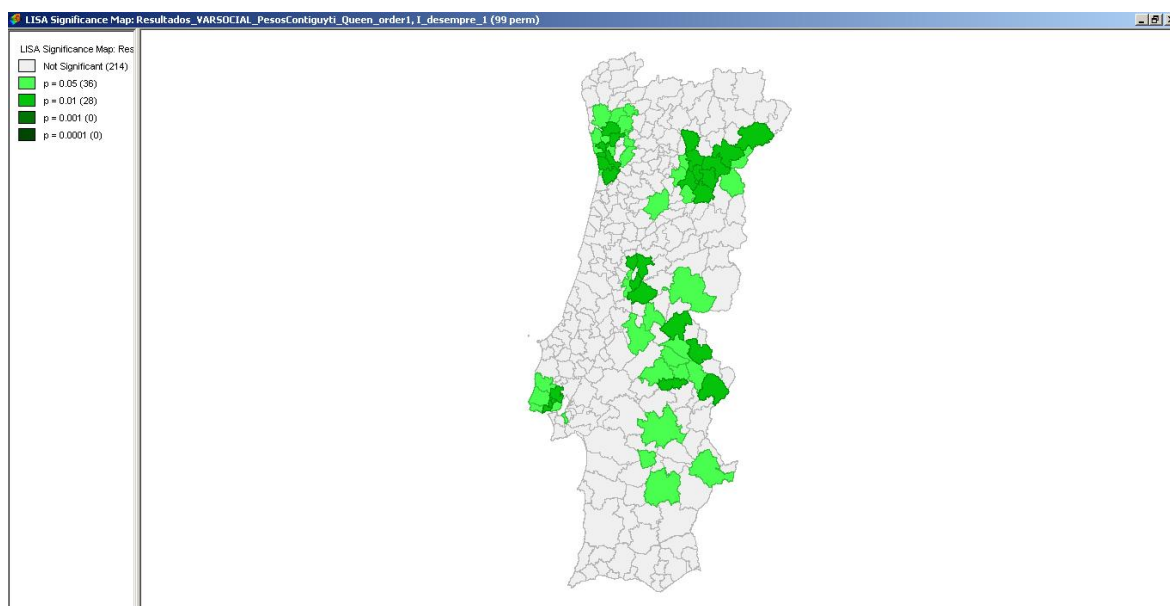
Os resultados são apresentados de seguida:



a) Coeficiente de Moran (autocorrelação positiva)



b) LISA Mapa de Clusters



c) LISA Mapa de significância

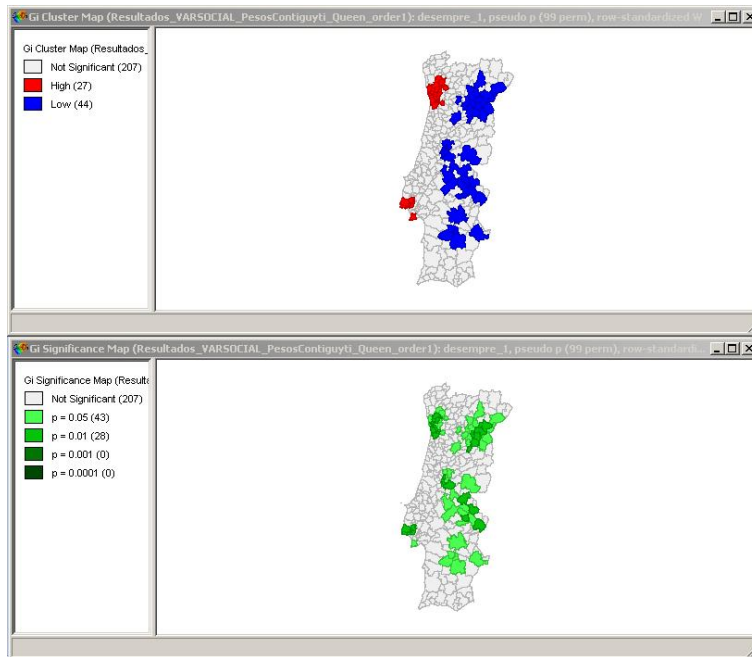
Fig. 13 – Resultados da aplicação do univariate local Moran's I (GeoDa)

Estes mapas indicam-nos onde estão situados os *clusters* (conjunto de polígonos que apresentam tendência semelhante) e com que nível de significância podemos afirmar essa hipótese (no mapa temos duas tonalidades de verde, o verde claro apresenta uma confiança de 95%, o verde mais escuro apresenta uma confiança de 99%). Neste caso o algoritmo calcula o coeficiente de Moran e efetua o teste estatístico considerando 99 permutações, indicando o resultado do teste e seu respetivo nível de confiança.

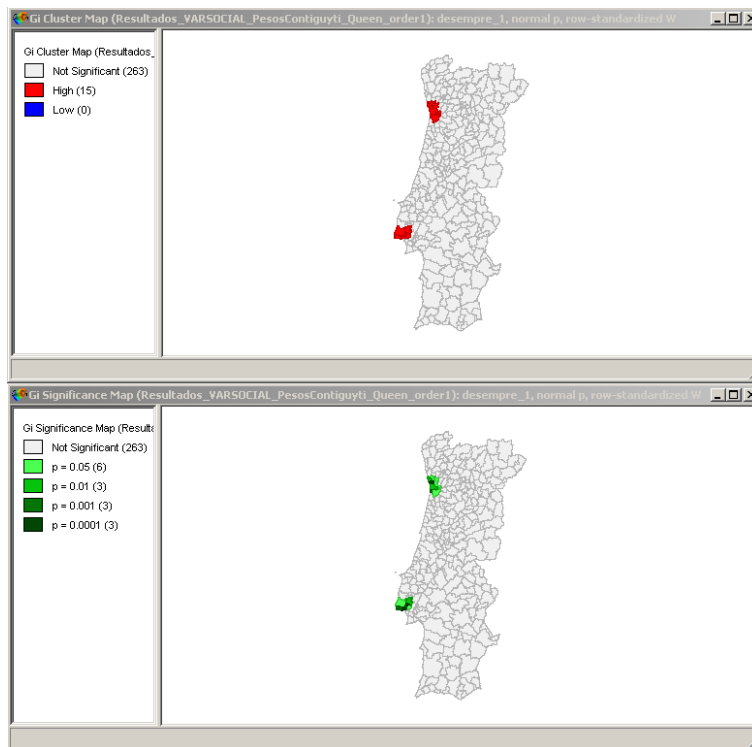
O teste de pseudo-significância resulta da comparação dos valores observados com uma amostra gerada aleatoriamente por forma a obter uma distribuição empírica para o índice que se pretende. Basicamente, são testes estatísticos analisando distribuições empíricas em que se testa a hipótese nula de aleatoriedade da amostra, ou seja, de não existirem aglomerados em que uma variável assume valores semelhantes.

O índice de  $G_i$  apresenta outro método de análise da autocorrelação espacial. O índice de Moran considera a covariância entre pares o índice  $G_i$  considera a soma dos valores das variáveis na vizinhança de acordo com a matriz de vizinhança. A função  $G_i$  e  $G_i^*$  são iguais na metodologia de cálculo. A diferença reside no facto de a determinação do índice de  $G_i$  não considerar o valor da variável no próprio polígono e o índice  $G_i^*$  considera o valor da variável no próprio polígono para determinação do índice.

Os valores altos de  $G_i$  e  $G_i^*$  apontam para a existência de altos índices de ocorrência deste atributo. Valores baixos de  $G_i$  e  $G_i^*$  apontam para uma situação de não ocorrência de valores similares ou seja de não existência de situações de aglomerados.

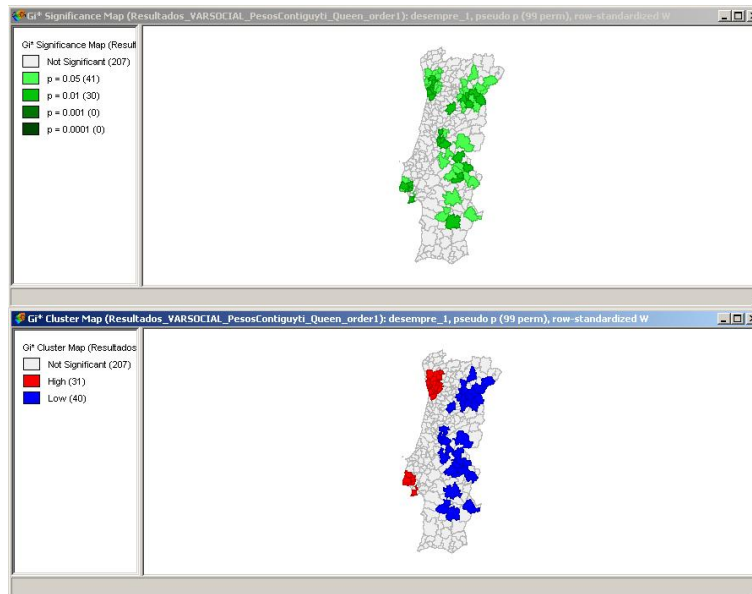


a) Resultados da função Gi sem normalização do valor p (GeoDa)

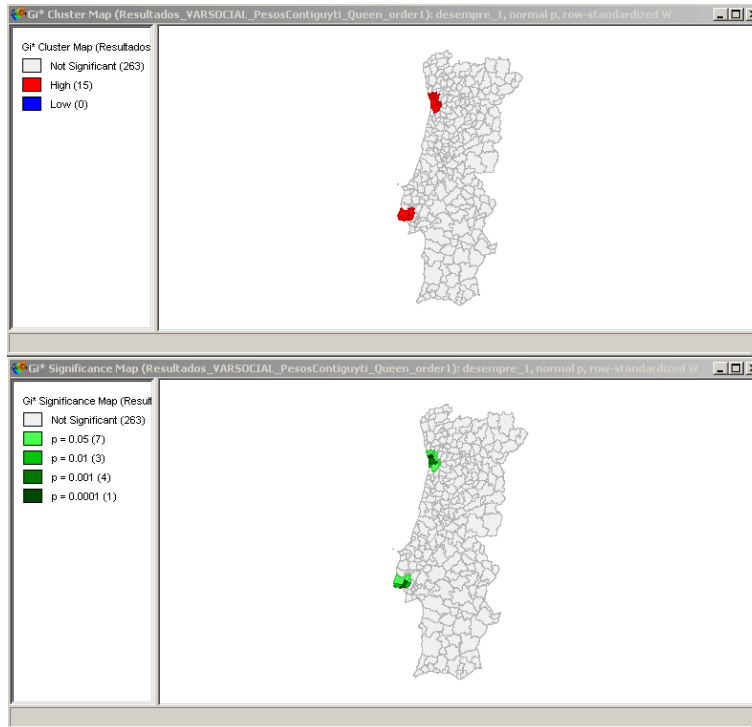


b) Resultados da função Gi com a normalização do valor p (GeoDa)

Fig. 14 - Resultados da aplicação do algoritmo de determinação da função Gi (GeoDa)



c) Resultados da função  $G_i^*$  sem normalização do valor p (GeoDa)



d) Resultados da função  $G_i$  com a normalização do valor p (GeoDa)

Fig. 15 - Resultados da aplicação do algoritmo de determinação da função  $G_i^*$  (GeoDa)

Da análise das figuras Fig. 14 e Fig. 15 verificamos que os resultados da função  $G_i$  e  $G_i^*$  são semelhantes e que existe uma tendência para a existência de valores semelhantes (aglomerados) do desemprego na área metropolitana de Lisboa e Porto, e na região do interior do país.

### 3.3. Fase 3 – Regressão Espacial

Na terceira fase do trabalho pretende-se analisar a regressão espacial. Um dos principais objetivos da análise de dados é deduzir modelos que ajudem a inferir uma determinada resposta como resultado da função de uma ou várias variáveis. Os modelos de regressão pressupõem que o fenómeno a modelar constitui uma variável dependente que pode ser inferida a partir de outras variáveis. Partindo desse modelo e dessas assunções é possível estimar o valor da variável dependente.

Quando se elabora um modelo matemático de regressão pretendem-se atingir dois objetivos: (1) obter um bom ajuste entre os valores preditos pelo modelo e os valores observados para a variável dependente; (2) encontrar quais as variáveis que mais contribuem para os resultados da variável dependente.

A análise estatística dos modelos de regressão é muitas vezes feita sobre os resíduos. Elaboram-se testes estatísticos sobre os resíduos e a sua distribuição estatística. Neste caso assume a hipótese nula de que os resíduos são independentes e não correlacionados, logo, seguem uma distribuição aleatória ou normal, com variância constante e média igual a zero.

Quando estamos na presença de fenómenos que apresentam uma dependência espacial, é provável que a hipótese padrão das observações serem não correlacionadas seja rejeitada pelo teste. Neste caso os resíduos apresentam uma dependência espacial que vai provocar desvios sistemáticos na amostra e condicionar o resultado do teste, levando-o a rejeitar a hipótese nula.

Após construir um modelo de regressão clássica deve-se testar estatisticamente os resíduos. Efetuar o seu mapeamento e investigar a sua distribuição. A distribuição dos resíduos irá apresentar indícios da sua dependência espacial ou não. Sobre os resíduos podem ser aplicados os testes de autocorrelação espacial apresentados anteriormente, nomeadamente o teste de Moran e Geary.

Na análise de modelos de regressão de amostras com referências espaciais segue-se o esquema genérico apresentado na figura Fig. 16.

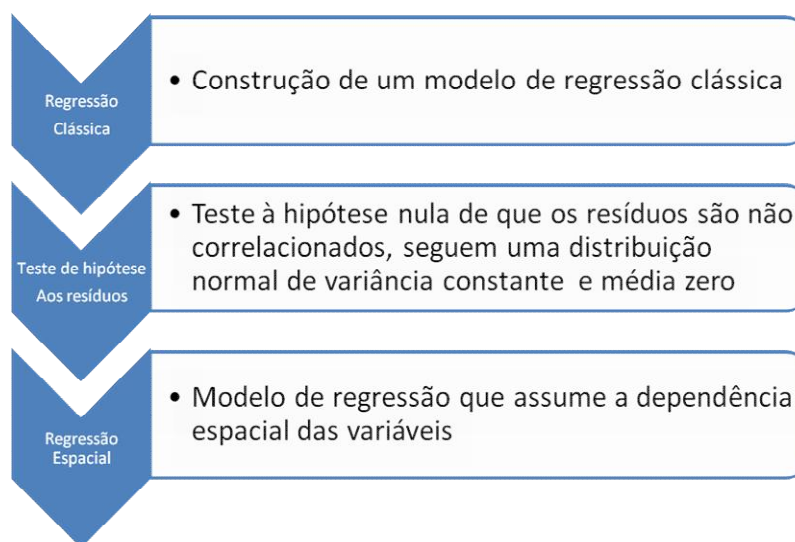


Fig. 16 – Esquema genérico da construção de modelos de regressão com dados espaciais

Após concluir que o espaço condiciona o modelo de regressão espacial, torna-se necessário ajustar o modelo de forma a contemplar os efeitos espaciais.



A inclusão dos efeitos espaciais nos modelos de regressão pode ser feita de várias formas. A mais simples, consiste na construção de modelos com efeitos espaciais globais. Nesse modelo parte-se do pressuposto que é possível modelar o efeito espacial através da inclusão de uma variável no modelo clássico. O novo parâmetro “espacial” resulta do efeito da autocorrelação espacial e pode ser inserido no modelo de duas formas. A primeira, a autocorrelação espacial é modelada pelo produto de uma variável independente pela matriz de vizinhança. Este método é designado de modelo espacial autoregressivo misto (“*Spatial AutoRegressive –SAR*” ou “*Spatial Lag Model*”). Associado ao parâmetro “espacial” temos um coeficiente designado de coeficiente espacial autorregressivo; A segunda, trata o efeito resultante da correlação espacial como ruído que é removido no processo. Esta segunda abordagem designa-se de modelo do erro espacial (“*Spatial error model*” ou “*Conditional AutoRegressive*”)<sup>12</sup>.

Quando o processo espacial é não estacionário, os coeficientes de regressão precisam de refletir a heterogeneidade espacial. A modelação da heterogeneidade espacial pode seguir duas abordagens. A primeira, modelar a tendência espacial de forma contínua, assumindo que a variação dos parâmetros é constante no espaço. A segunda, consiste em dividir o espaço em sub-regiões estacionárias, essas regiões são designadas de regimes espaciais.<sup>13</sup>

A aplicação GeoDa permite elaborar modelos de regressão linear dos seguintes tipos: clássica, *spatial lag* e *spatial error*.

De seguida iremos apresentar a abordagem de construção de um modelo de regressão espacial simples, seguindo o esquema genérico apresentado na figura 16.

Durante a recolha de dados, realizaram-se vários gráficos de forma verificar se existia algum padrão que pudesse ser utilizado para construir um modelo predictivo de regressão. Verificámos que os dados de violência doméstica, desemprego e divórcio apresentava um padrão muito similar conforme pode ser visto nos gráficos da figura 17.

---

<sup>12</sup> Para mais informação ver: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf> (acedido em abril de 2014).

<sup>13</sup> Existem vários modelos de regressão espacial global e local que podem ser adotados e estudados de forma a encontrar a melhor solução para modelar o fenómeno que pretendemos estudar. Para mais informação sobre este assunto recomenda-se: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf> (acedido em abril de 2014).

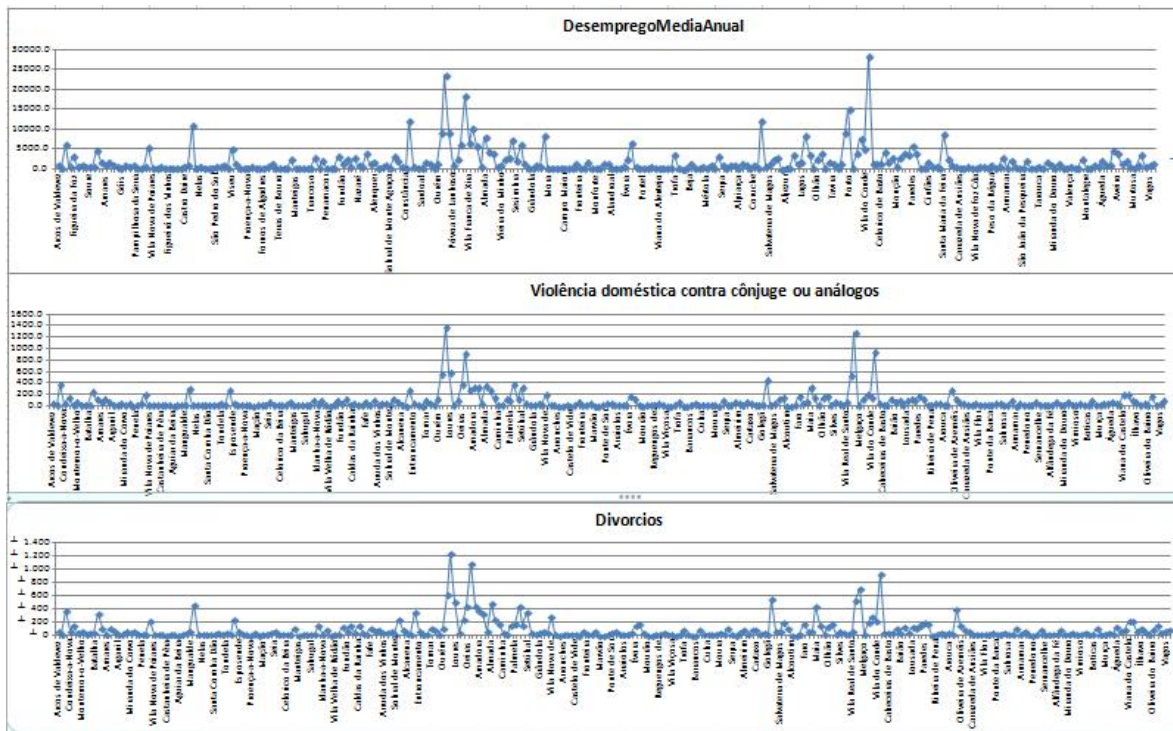


Fig. 17 – Gráficos relativos ao desemprego, violência doméstica e divórcios nos municípios de Portugal Continental (ano 2011)

Graticando no GeoDa obtivemos uma elevada relação entre os divórcios e os crimes de violência doméstica.

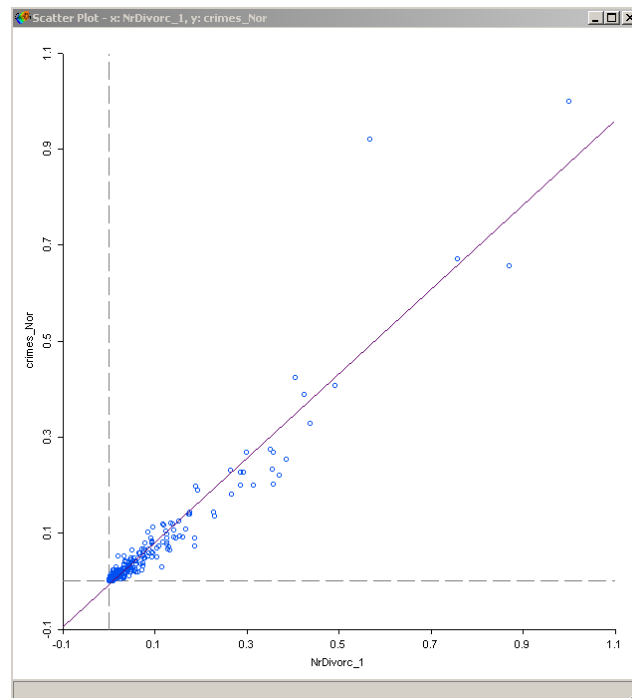


Fig. 18 – Gráfico do número de divórcios (normalizado) versus número de crimes (normalizado)

De seguida iremos utilizar os conhecimentos aprendidos para construir um modelo de regressão que possa explicar o fenómeno do divórcio em função das seguintes variáveis: desemprego, violência doméstica, salário e população sem escolaridade.

Para colocar todas as variáveis no mesmo modelo existia um problema de escala que deveria ser resolvido e existe a necessidade de manter as variáveis independentes para não condicionar os pressupostos da regressão. Tendo em conta essas questões optou-se pela normalização das variáveis de zero a um, através da razão entre o valor da variável e o valor máximo da amostra.

Utilizando o GeoDa foram criados os modelos de regressão (Clássico, Spatial Lag e Spatial Error) e obtiveram-se os seguintes parâmetros. Primeira análise: Divórcio em função do número de desempregados médios (D), número de crimes de violência doméstica (CR), Salário Médio (S) e escolaridade (E)

Tabela IV – Resultados do modelo de regressão para modelar o divórcio

Método	Modelo	R <sup>2</sup>	σ <sup>2</sup>
Clássico	-0.05189304+0.2798042E+0.09391802S+0.4120098D+0.3893482CR	0.96	0.00063
<i>Spatial Lag</i>	-0.04875592+0.06923363Y+0.2932841E+0.08049574S+0.3863475D+0.3655519CR	0.96	0.00059
<i>Spatial Error</i>	-0.03219481+0.4445413Err+0.2752679E+0.05568026S+0.4713195D+0.328188CR	0.96	0.00054

Da análise dos coeficientes verifica-se que o modelo de regressão apresenta uma boa estimativa para o fenómeno. Ou seja, estamos na presença de um modelo que ajusta bem ao fenómeno e produz resíduos reduzidos.

Os resultados apresentados indicam que os modelos de regressão linear clássica e espacial são semelhantes, essa questão é analisada através de um teste de correlação espacial aos resíduos através de índice de Moran.

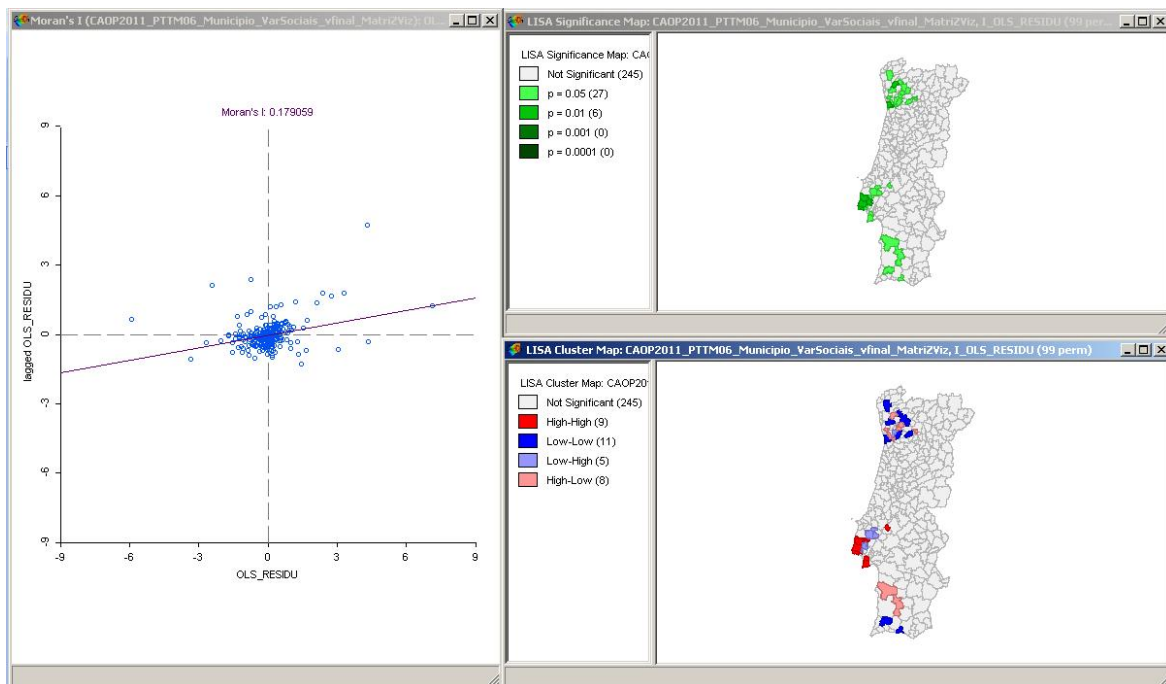


Fig. 19 – Resultado do teste de Moran aos resíduos do modelo de regressão linear clássica.

O teste estatístico sobre os resíduos resultantes do modelo de regressão clássica indica que neste caso os resíduos apresentam uma baixa autocorrelação espacial. Este teste permite concluir que não existe uma dependência espacial entre as variáveis e que bastaria a regressão clássica para modelar o fenómeno.

## 4 - Análise Espacial de Superfícies Contínuas e Redes

### 4.1 – Métodos de Interpolação:

A recolha de dados em geociências ocorre normalmente seguindo um método de amostragem pontual, partindo do pressuposto que os valores das variáveis recolhidas nesses pontos (amostras) serão explicativos do comportamento da população. Após a recolha de dados *in loco* é necessário estimar os valores da variável em novas posições a partir dos dados conhecidos. Este processo é designado de interpolação.

Seguindo a classificação de processos de interpolação sugeridos em (Camargo *et. al.*, 2004) os processos de interpolação seguem três grandes abordagens:

- Modelos determinísticos de efeitos locais: nesta classe a interpolação cada novo ponto é estimado apenas a partir da interpolação das amostras mais próximas, utilizando funções determinísticas em que não é elaborada qualquer hipótese estatística sobre a variabilidade espacial do fenómeno.
- Modelos determinísticos de efeitos globais: a suposição implícita nesta classe de interpoladores é que, para a caracterização do fenómeno em estudo, predomina a variação em larga escala, e que a variabilidade local não é relevante. Com estes interpoladores pretende-se obter a superfície de tendência.
- Modelos estatísticos de efeitos locais e globais (*Krigagem*): o valor da variável em estudo em cada novo ponto é estimado apenas a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico. Neste caso assume-se que estamos na presença de um fenómeno influenciado pelas condições locais (dependência espacial) que é modelado por uma componente que se considera estocástica.

Os métodos de interpolação seguem um processo que pressupõe as seguintes etapas genéricas:

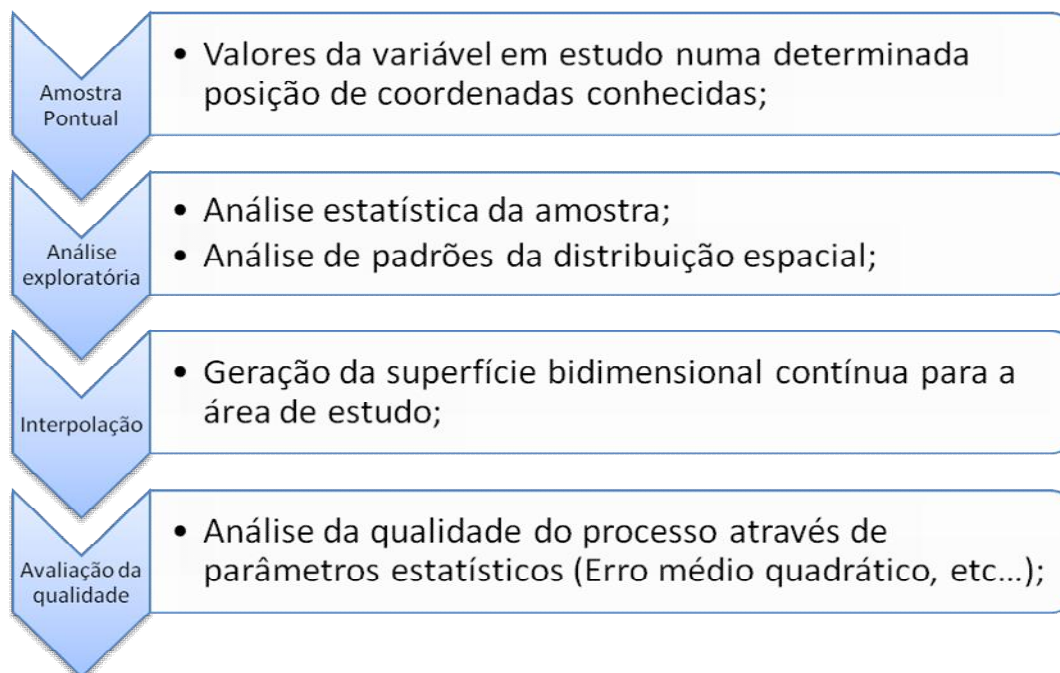


Fig. 20 – Fluxo do processo de interpolação

#### 4.1.1 – Métodos de Interpolação Determinísticos:

Os modelos de interpolação determinísticos são utilizados para gerar superfícies contínuas a partir de amostras pontuais. Neste caso considera-se que o fenómeno tem uma distribuição espacial independente das características do espaço. Baseiam-se na lei de Waldo Tobler e partem do único pressuposto de que as amostras que estão mais próximas se assemelham mais do que as que estão maior distância. Os métodos de interpolação determinísticos podem ser classificados de acordo com a sua extensão espacial em locais ou globais, e de acordo com a forma como se ajustam à amostra pontual de base. Em exatos, quando a função de interpolação estima um valor da variável igual ao valor observado; em inexatos, quando o valor estimado pela função de interpolação para a superfície difere do valor observado.

No âmbito do trabalho prático foi proposto que se aplicasse um método determinístico para aumentar a resolução de modelos digitais de elevação. Este é um exemplo de tarefa que normalmente não se realiza porque não existe uma melhoria do modelo de elevação resultante, ou seja uma descrição mais realista do fenómeno – relevo. Contudo, estas técnicas são particularmente úteis quando se pretende realizar o aumento ou diminuição da resolução geométrica de modelos de elevação para produzir GRID de igual resolução às quais se possam aplicar técnicas de álgebra matricial, por exemplo: construção de modelos algébricos de análise multi atributo.

Para a análise proposta, utilizou-se um modelo digital de elevação disponível no *blog* Via SIG (Duarte, 2010)<sup>14</sup> com uma resolução de 30 metros disponibilizado à comunidade de utilizadores de aplicações SIG.<sup>15</sup> Segundo o autor, o modelo disponibilizado em formato RASTER GeoTIFF, resulta da reprojeção dos mosaicos originais do projeto *Advanced Spaceborne Thermal Emission and Reflection Radiometer* para o sistema Hayford-Gauss Datum 73, comprimido no formato 7zip. Os dados originais dizem respeito a um modelo digital de elevação resultante de um projeto de colaboração entre a *United States National Aeronautics and Space Administration* (NASA) e *Japanese Ministry of Economy, Trade, and Industry* (METI), os dados foram obtidos por um sensor espacial designado *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER) disponibilizado de forma gratuita com uma resolução de 30 metros sobre a forma de telhas quadradas de 1' de lado (NASA, 2004)<sup>16</sup>.

Na *Internet* é possível encontrar outros modelos digitais de elevação como o modelo *Shuttle Radar Topography Mission* (SRTM)<sup>17</sup> que apresenta uma resolução de 30 metros nas zonas de interesse dos Estados Unidos da América e de 90 metros nas restantes áreas. Os dados utilizados para produzir este modelo foram recolhidos pela técnica de interferometria RADAR com o equipamento instalado no *Space Shuttle Endeavour* em fevereiro de 2000<sup>18</sup>.

---

<sup>14</sup> <http://blog.viasig.com/2010/03/mdt-30m-para-portugal/> (acedido em junho de 2014).

<sup>15</sup> <https://onedrive.live.com/?cid=cff1b4093141519b&id=CFF1B4093141519B%21154> (acedido em junho de 2014).

<sup>16</sup> <http://asterweb.jpl.nasa.gov/gdem.asp> (acedido em junho de 2014).

<sup>17</sup> <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1> (acedido em junho de 2014).

<sup>18</sup> <http://www2.jpl.nasa.gov/srtm/> (acedido em junho de 2014) ver também <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1> (acedido em fevereiro de 2014).

E a nível europeu, o modelo digital de elevação da área europeia resultante do projeto GMES RDA<sup>19</sup> com uma resolução de 30 metros.

Para medir a qualidade do modelo final foi determinado o erro médio quadrático das diferenças entre as cotas do modelo de elevação e as cotas ortométricas dos marcos geodésicos da rede nacional geodésica (DGT, 2013)<sup>20</sup>. À partida, são expectáveis grandes diferenças entre os modelos globais e as cotas dos marcos geodésicos, sendo importante referir que o modelo digital de elevação vai ser utilizado numa aplicação em que se admite uma reduzida exatidão dos dados de elevação – determinação de bacias de visão a partir dos postos de vigia do distrito de Leiria.

Para interpolar o modelo digital de elevação utilizou-se a função de interpolação determinística da média *Inverse Distance Weighed* (IDW)<sup>21</sup>. O IDW é um interpolador determinístico local, cuja vizinhança é definida em termos de raio de busca, ou número de vizinhos a considerar. O processo de interpolação consistiu em criar uma grelha regular de pontos a partir do RASTER do DEM com uma resolução de 30 metros, à grelha de pontos foi aplicado a interpolação IDW obtendo-se o RASTER com 15 metros de resolução. De seguida, foi determinado o valor da superfície para cada um dos pontos correspondentes aos marcos geodésicos, e por fim foram determinados os parâmetros estatísticos que permitem medir a exatidão do modelo interpolado.

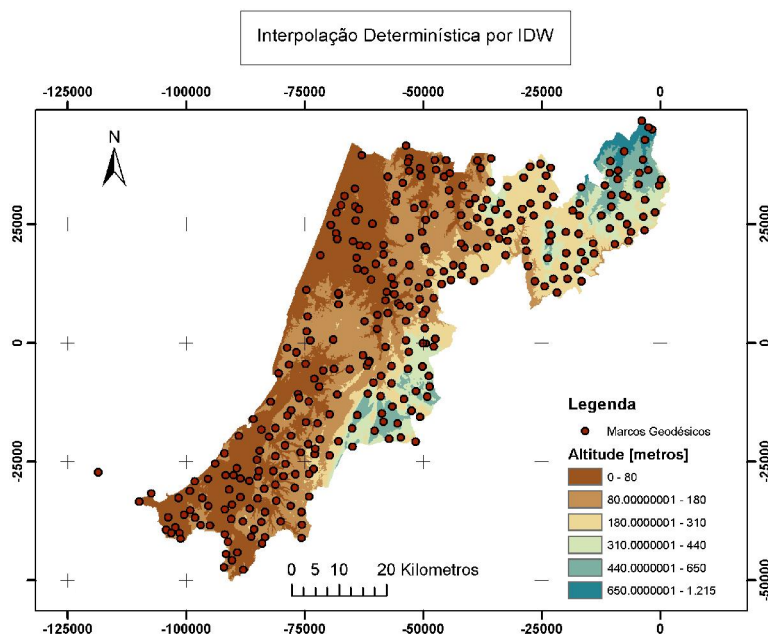


Fig. 21 – Modelo digital de elevação com uma resolução de 15 metros gerado por uma interpolação IDW

<sup>19</sup> <http://www.eea.europa.eu/data-and-maps/data/eu-dem> (acedido em junho de 2014),

<sup>20</sup>

[http://www.dgterritorio.pt/cartografia\\_e\\_geodesia/geodesia/redes\\_geodesicas/rede\\_geodesica\\_nacional/](http://www.dgterritorio.pt/cartografia_e_geodesia/geodesia/redes_geodesicas/rede_geodesica_nacional/) (acedido em junho de 2014).

<sup>21</sup> Para informação sobre o algoritmo IDW implementado no ArcGIS ver: [http://resources.arcgis.com/en/help/main/10.1/index.html#/How\\_IDW\\_works/009z00000075000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/How_IDW_works/009z00000075000000/) (acedido em junho de 2014).

Tabela V – Resultados estatísticos das diferenças entre as cotas dos marcos geodésicos e o modelo digital de elevação interpolado por IWD com uma resolução de 15 metros.

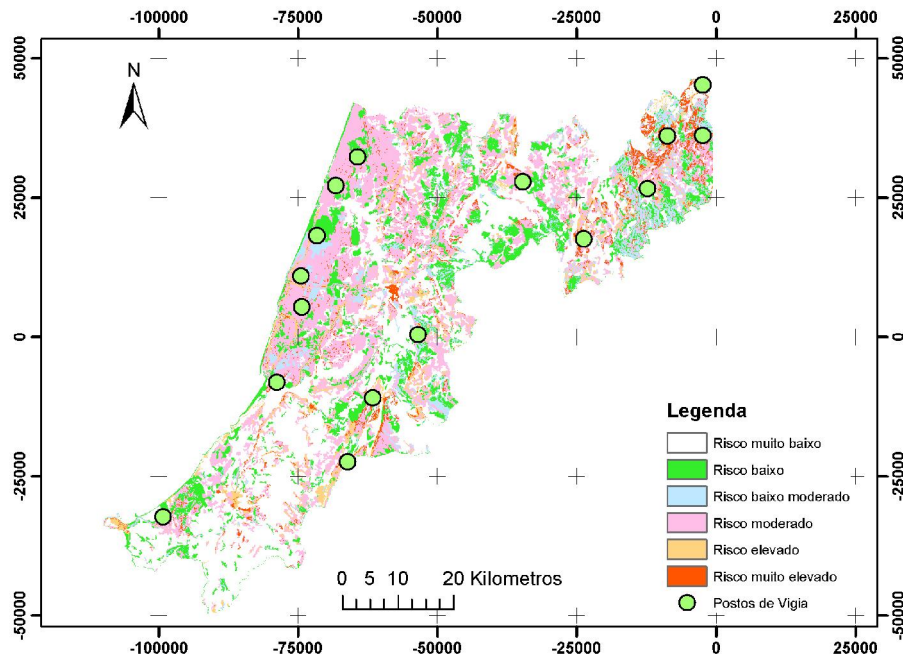
Diferenças entre os marcos geodésicos e o Modelo Digital de Elevação ASTER IDW 15 metros de resolução	
Número de valores comparados	351
Máximo	66.0 m
Mínimo	-5.0 m
Média	22.0 m
Desvio Padrão	9.7 m
EMQ	24.0 m

Tal como esperado, verificamos que o erro médio quadrático é muito elevado (24 metros), o que demonstra a reduzida exatidão global do modelo.

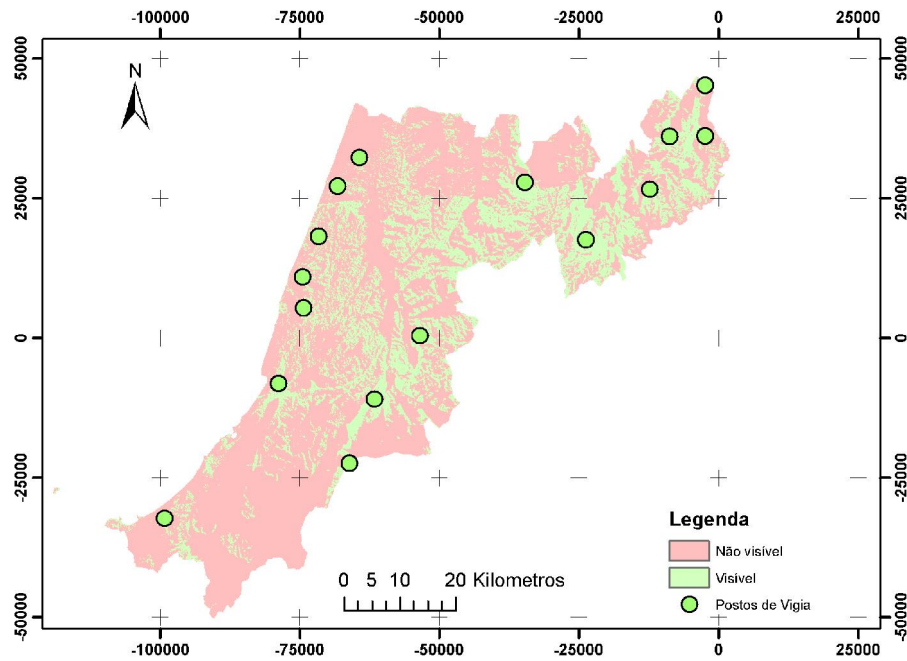
Nos sistemas de informação geográfica existem algoritmos específicos que permitem gerar os seguintes produtos a partir dos modelos digitais de elevação: superfícies de declive, exposição, escoamento, e bacias de visão. No âmbito deste trabalho vai ser gerada um RASTER com as bacias de visão a partir dos postos de vigia que existem no distrito de Leiria.

As posições dos postos de vigia utilizados na análise foram recolhidas em dezembro de 2013 a partir do sítio: [http://scrif.igeo.pt/ASP/postos\\_f.asp](http://scrif.igeo.pt/ASP/postos_f.asp) (IGEO - Grupo Crise).

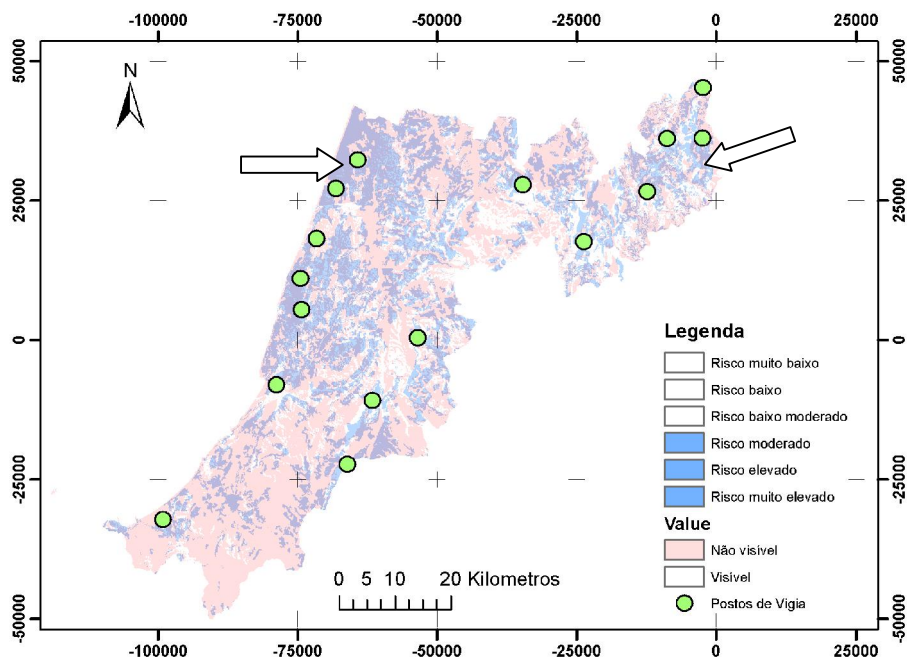
A nova *layer* de informação gerada com o algoritmo de bacias de visão irá ser comparada com o mapa de vulnerabilidade calculado no âmbito da disciplina CSIG (2013). O mapa de vulnerabilidade fig. 22 a) diz respeito ao terceiro trimestre, período mais propício à ocorrência de incêndios. Esta análise tem como objetivo verificar se a atual distribuição de postos de vigia garante uma completa cobertura e vigilância sobre todas as áreas de maior vulnerabilidade à ocorrência de incêndios.



(a)



(b)



(c)

Fig. 22 – (a) Áreas de vulnerabilidade de incêndio no terceiro trimestre, (b) Bacias de Visão, (c) Cruzamento das áreas de risco moderado a muito elevado com as bacias de visão

Na figura anterior mostra-se a sobreposição do mapa de vulnerabilidade de incêndio no Distrito de Leiria do terceiro trimestre com o mapa de bacias de visão. Verifica-se que existem áreas de risco elevado de ocorrência de incêndio que se situam em zonas que não são visíveis dos postos de vigia existentes na área (setas). Todavia, verificamos que a atual rede de postos de vigia garante a cobertura visual da maioria das zonas classificadas como vulneráveis.



#### 4.1.2 – Métodos de Interpolação Probabilísticos:

Nos métodos de interpolação probabilísticos (geoestatística) pretende-se efetuar a modelação dos fenómenos considerando a existência de processos de autocorrelação espacial inerente aos dados (Matos, 2008). Estes métodos são particularmente úteis para modelar fenómenos que ocorrem mediante a existência de características espaciais muito específicas que geram padrões que devem ser estudados e interpretados.

Os processos de interpolação que preveem a existência de fenómenos estocásticos tornam o processo de modelação mais complicado e demorado, contudo, apresentam as seguintes vantagens:

1. Permitem criar mapas mais realistas;
2. A análise exploratória permite identificar *outliers* nos dados, elaborar hipóteses e entender melhor os fenómenos;
3. A modelação da dispersão do fenómeno permite considerar a influência das características do espaço;

Antes da aplicação de ferramentas de geoestatística (p. ex.: *Kriging*) deve ser realizada a análise exploratória dos dados onde se pretende verificar a sua eventual aleatoriedade (Soares, 2006), isto não significa que o fenómeno foi gerado de forma aleatória mas sim que o seu padrão apresenta uma determinada organização espacial. Nos processos de interpolação probabilísticos devem ser realizadas etapas genéricas que se encontram resumidas no diagrama seguinte:

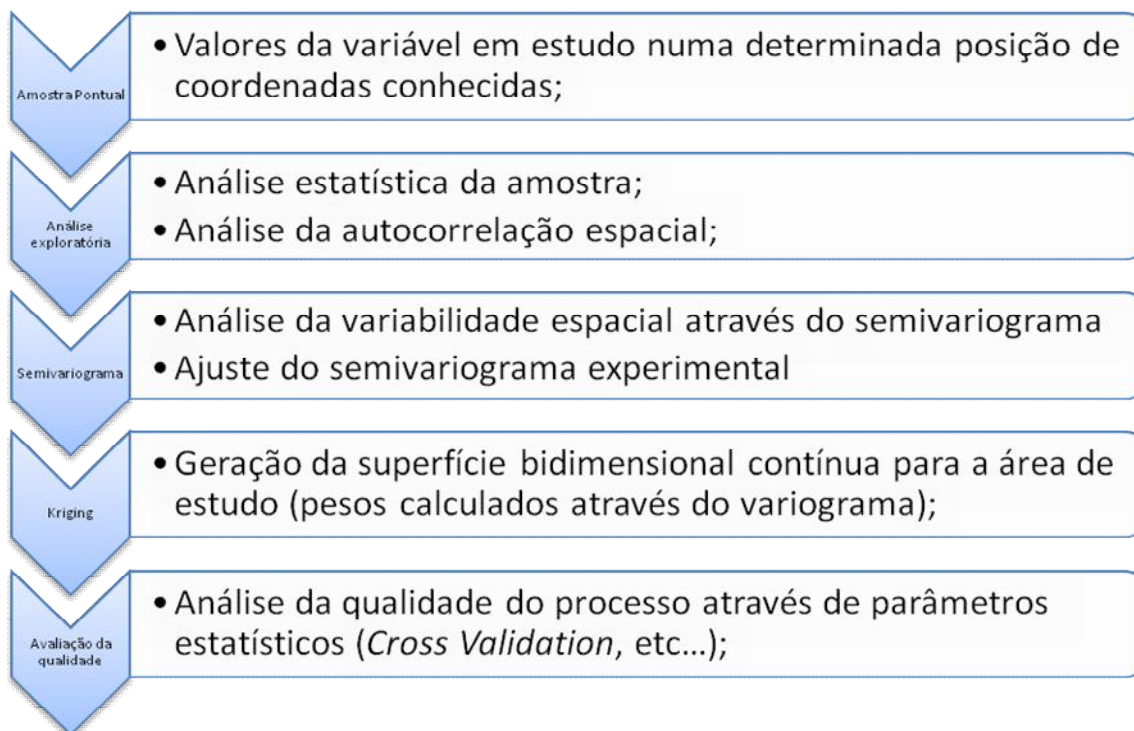


Fig. 23 . Esquema genérico da aplicação de métodos de interpolação probabilísticos<sup>22</sup>

No caso destes interpoladores, os coeficientes do modelo matemático são determinados através da teoria probabilística. Por esse motivo, ao contrário dos modelos determinísticos, a aplicação destas ferramentas tem de ser antecedida de uma análise estatística dos dados. O método que iremos utilizar para a interpolação probabilística será o *kriging*. Entre os métodos genericamente designados de *kriging* vamos abordar o *kriging* genérico (*ordinary kriging*) em que se assume que o fenómeno tem uma componente estacionária global e uma componente estocástica que resulta do efeito do espaço (Hengl, 2009). Esta interpolação é basicamente uma média ponderada sofisticada em que os pesos são determinados com base na função semivariograma, esta função permite determinar a distribuição dos pesos na vizinhança de um ponto de forma que o cálculo algébrico entre em linha de conta com a autocorrelação espacial. No âmbito deste relatório não será apresentado o formulário matemático que suporta o método. A teoria poderá ser encontrada com maior ou menor profundidade na bibliografia dedicada à geoestatística (Soares, 2006; Hengl, 2009).

Os dados que serão analisados nesta etapa estão referidos na tabela seguinte, sendo que a descrição presente neste relatório não será aprofundada sobre todos os elementos. Procura-se exemplificar a análise exploratória e interpolação sobre os dados de gravimetria, porém, restantes conjuntos de dados servirão para cruzamento de informação e análise. O trabalho nesta etapa continua a análise iniciada na etapa 1 em que se verificou a existência de valores elevados na concentração de nitratos em várias áreas, neste caso iremos estudar a bacia hidrográfica do Rio Vouga.

Tabela VI – Fontes de dados para a interpolação determinística

Dados	Fonte de dados	Data de recolha	Observações
Dados da concentração de nitratos nas águas subterrâneas do continente.	Sistema Nacional de Informação de Recursos Hídricos da Agência Portuguesa do Ambiente: <a href="http://snirh.apambiente.pt">http://snirh.apambiente.pt</a>	Março de 2014	Amostra pontual referente a 100 estações de análise de águas subterrâneas limitadas pela BH do Rio Vouga.
Dados precipitação média diária	Sistema Nacional de Informação de Recursos Hídricos da Agência Portuguesa do Ambiente: <a href="http://snirh.apambiente.pt">http://snirh.apambiente.pt</a>	Junho de 2014	Amostra pontual referente a 20 estações meteorológicas na zona da BH do Rio Vouga.
Modelo digital de elevação ASTER	<a href="http://blog.viasig.com/2010/03/mdt-30m-para-portugal/">http://blog.viasig.com/2010/03/mdt-30m-para-portugal/</a>	Junho de 2014	DEM da zona da BH do Rio Vouga com uma resolução de 30 metros
Dados de sedimentos (geologia)		Mai de 2014	Grelha regular com os valores da velocidade dos sedimentos (relacionada com a densidade dos elementos geológicos)
Componente perturbadora da gravidade	<a href="http://ddfe.curtin.edu.au/gravitymodels/GGMplus/">http://ddfe.curtin.edu.au/gravitymodels/GGMplus/</a>	Junho de 2014	Componente de perturbadora da gravidade relativa ao modelo global do Có-geoide GGMPlus
Limites da Bacia	Atlas do Ambiente	Junho de 2014	

Hidrográfica do Rio Vouga	<a href="http://sniamb.apambiente.pt/webatlas/I_17.html">http://sniamb.apambiente.pt/webatlas/I_17.html</a>		
Rede Hidrográfica	Atlas do Ambiente <a href="http://sniamb.apambiente.pt/webatlas/I_17.html">http://sniamb.apambiente.pt/webatlas/I_17.html</a>	Junho de 2014	

A construção da superfície contínua pelo método de interpolação por *kriging*, será implementada sobre os valores do potencial perturbador gravitacional do modelo de co-geoide GGMPPlus.

#### 4.1.2.1 – Análise exploratória dos dados:

A análise exploratória dos dados descrita de seguida será realizada com as ferramentas existentes no módulo *Geostatistical Analysis* do ArcGIS10.2 sobre a amostra de dados do co-geoide GGMPPlus, dados de concentração de nitratos nos aquíferos e dados de precipitação média diária na bacia hidrográfica do rio Vouga.

Os dados da componente perturbadora da gravidade do GGMPPlus encontram-se numa grelha regular de aproximadamente 350 metros (longitude) por 450 metros (latitude). A amostra de dados é constituída por 23 534 pontos cujos elementos estatísticos se encontram na figura seguinte.

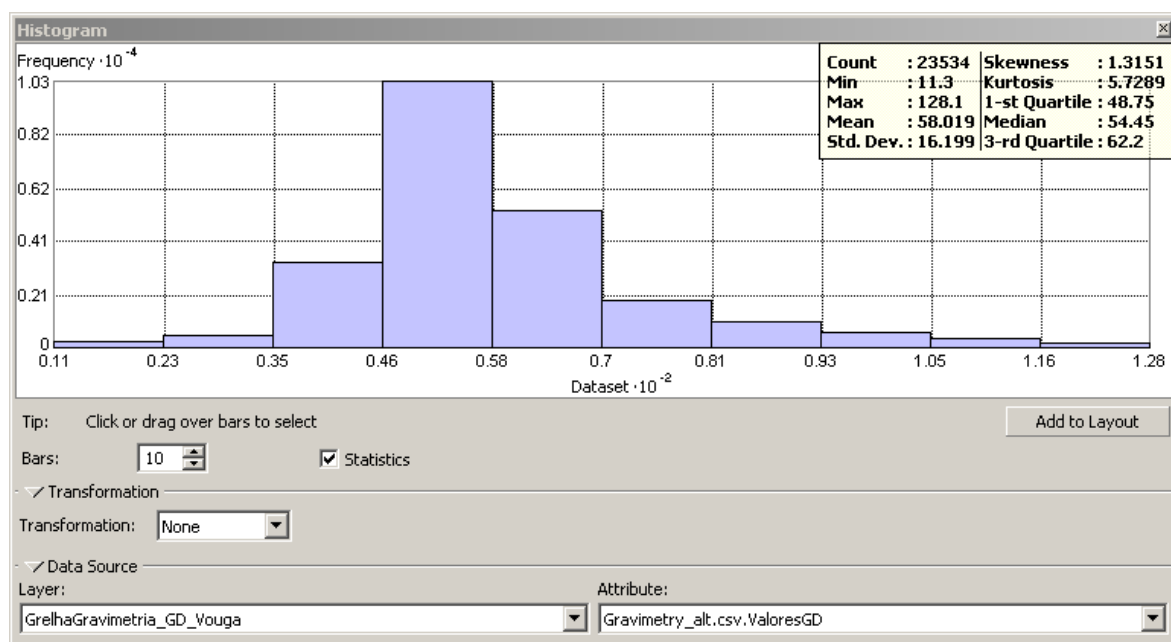


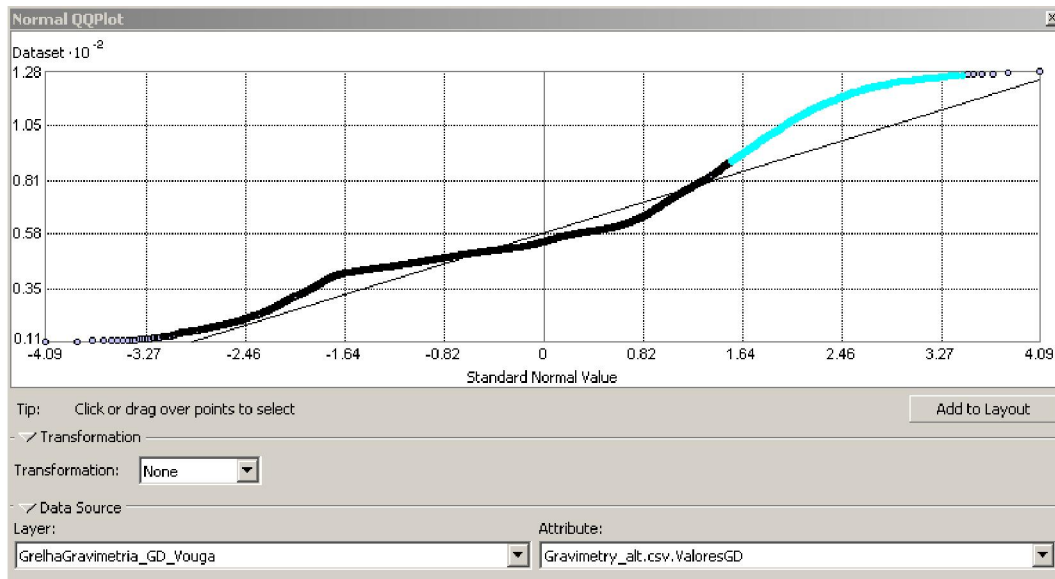
Fig. 24 – Elementos estatísticos da amostra pontual regular da componente perturbadora da gravidade do co-geoide GGMPPlus (unidades em 0.1 mGal)

O *Normal QQPlot*<sup>23</sup> permite verificar se os dados seguem uma distribuição normal de média igual a 0 e de desvio padrão igual a 1 é o. Do gráfico seguinte podemos inferir que se os valores da amostra se aproximam da semi-recta a 45° estamos na presença de uma amostra com distribuição espacial normal ou aleatória, no caso da amostra se afastar da semi-recta de referência temos uma distribuição não aleatória.

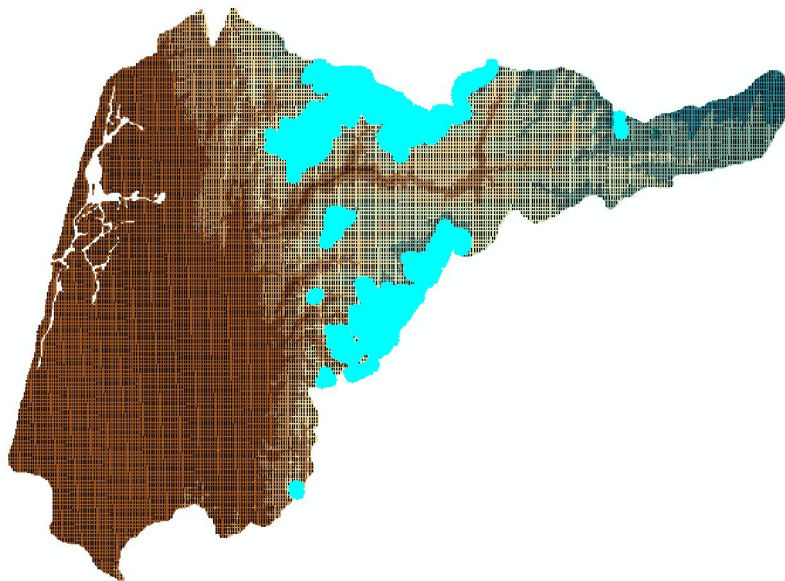
<sup>23</sup> Sobre o *Normal QQ plot* ver também:

[http://resources.arcgis.com/en/help/main/10.1/index.html#/Normal\\_QQ\\_plot\\_and\\_general\\_QQ\\_plot/0031000000q000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/Normal_QQ_plot_and_general_QQ_plot/0031000000q000000/) (acedido em junho de 2014)

Na figura 25.a) foram selecionados alguns pontos que apresentam características estatísticas de distribuição não aleatória, a sobreposição da amostra ao modelo digital de elevação indica-nos que os selecionados correspondem às zonas de maior altitude no modelo.



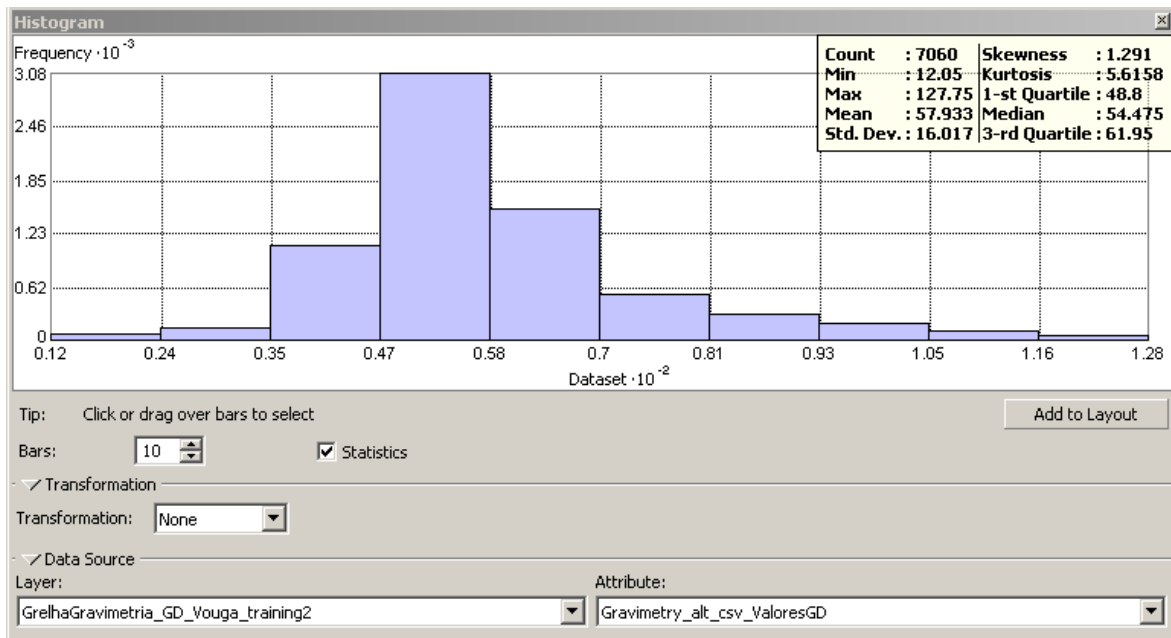
(a)



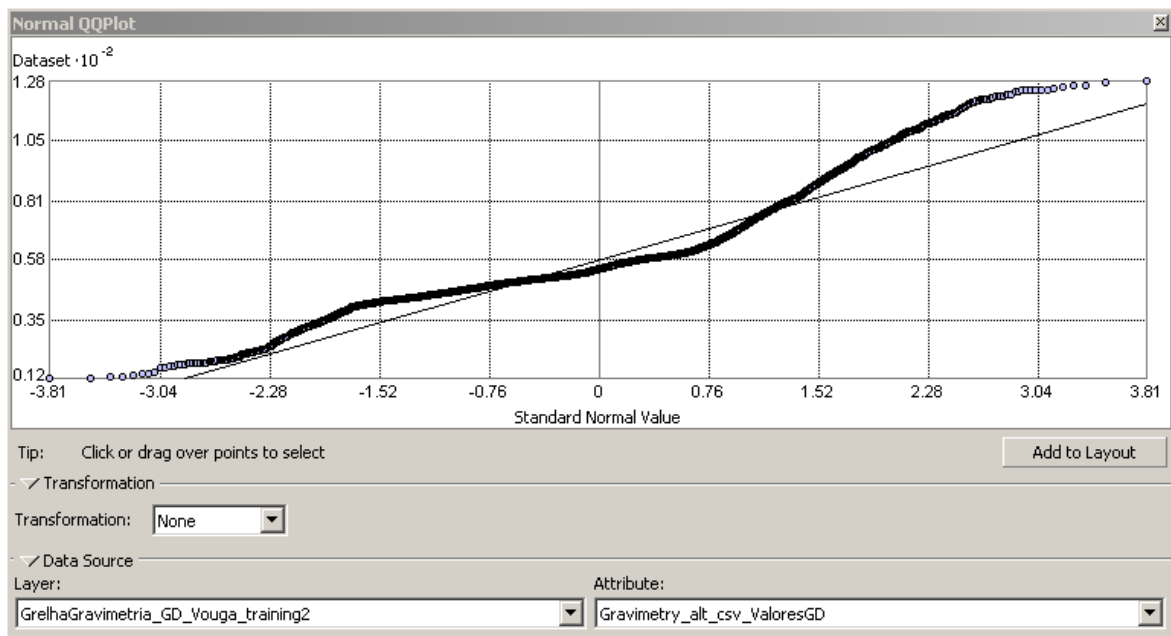
(b)

Fig. 25 – (a) Resultados do Normal QQPLOT aplicado à amostra da componente de perturbação da gravidade do modelo GGMPPlus. (b) Seleccção dos pontos que se afastam da distribuição normal sobrepostos ao modelo digital de elevação da bacia hidrográfica do Vouga.

A dimensão da amostra torna o processo computacional de cálculo e análise muito exigente. Por essa razão, considerou-se a extração de uma amostra de treino que compreende 30 % da amostra original (7 060 pontos). Depois de efetuar este passo é necessário analisar a amostra de treino em comparação com os dados originais, pelos gráficos presentes na figura seguinte conclui-se que estatisticamente a amostra de treino caracteriza a amostra original. Com base nestes elementos, podemos continuar a análise exploratória sobre a amostra de treino inferindo a partir dela as características da amostra global.



(a)

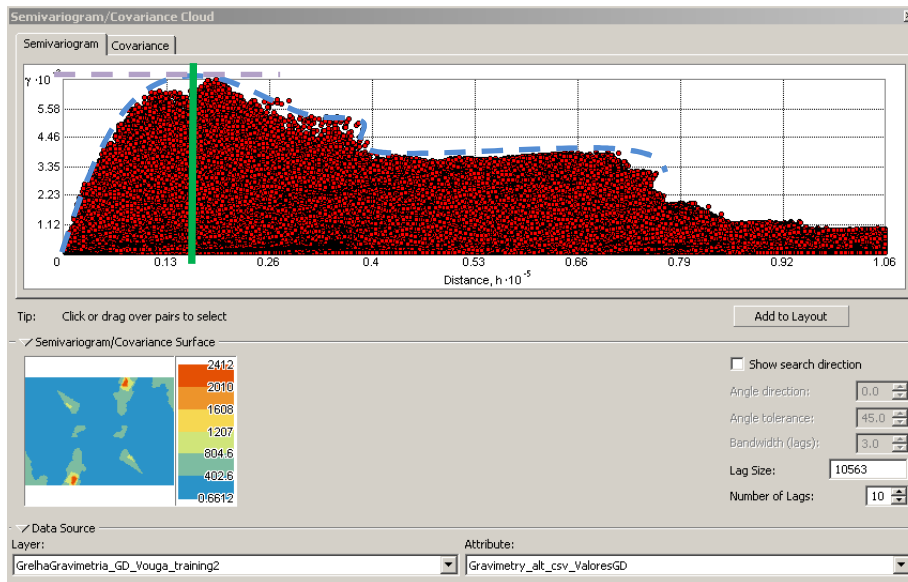


(b)

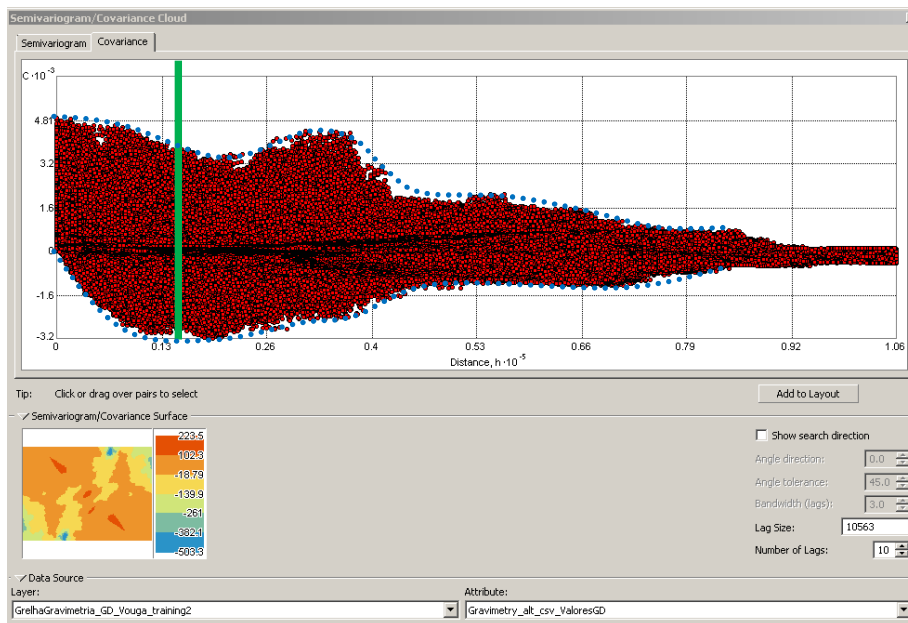
Fig. 26 – Parâmetros estatísticos da amostra de treino.

No processo de interpolação por *kriging* existe o mesmo problema de base que em todos os modelos de interpolação: qual deve ser o peso a atribuir aos vizinhos por forma a termos uma boa

modelação da realidade? No *kriging* esta questão é resolvida com uma função auxiliar designada de semivariograma. Esta função permite determinar a variância das diferenças da variável numa vizinhança. A novidade introduzida pelo método é oferecer um método objetivo de determinação dos pesos de forma que estes reflitam a auto correlação espacial da estrutura, relacionada diretamente com o parâmetro estatístico da variância.



(a)

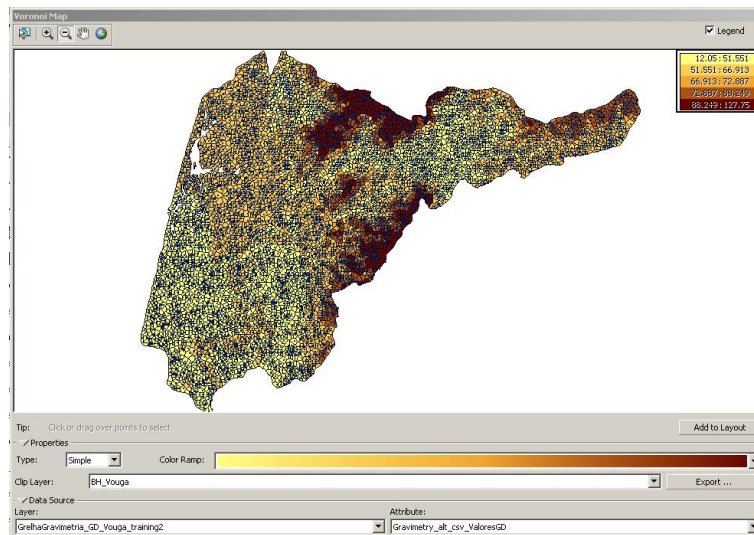


(b)

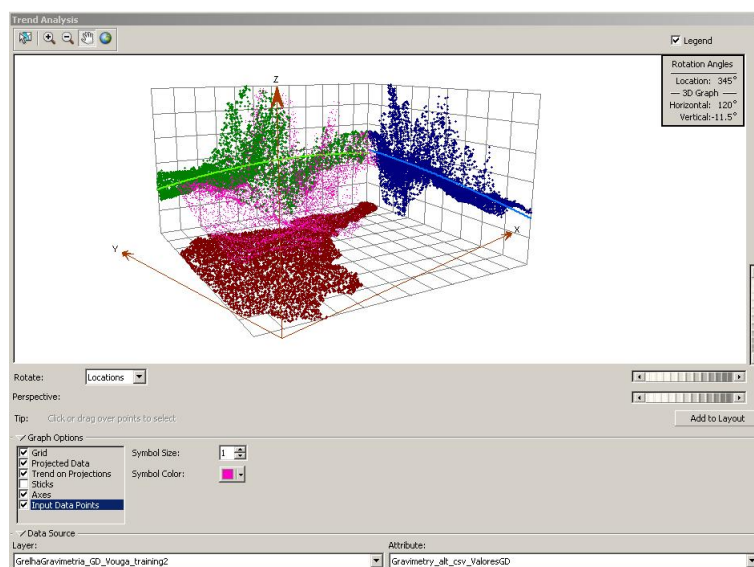
Fig. 27 – Função (a)semivariograma e (b)covariância

As figuras anteriores apresentam o resultado gráfico da função semivariograma e covariância. Nos gráficos anteriores verificamos que a função variograma apresenta um crescimento progressivo até um ponto em que começa a decrescer (linha tracejada azul na figura 29. a)). Este comportamento pode ser interpretado da seguinte forma, os valores da variável são semelhantes (autocorrelação espacial) até à distância em que começa a decrescer a semelhança (esta distância é designada de alcance e traduz a distância a partir da qual a diferença da variável nos pares de vizinhos é próxima

da variância global do fenômeno (tendência) – linha a verde). Neste caso o alcance será aproximadamente  $0.15 \times 10^5$  metros. Porém, estamos a analisar o comportamento global, como veremos com mais pormenor à frente este valor apresenta valor diferentes variando o valor da *lag*. O valor máximo onde ocorre a estacionaridade ou inflexão da função variograma é designado de patamar (linha de cor rosa).



(a)



(b)

Fig. 28 – Análise exploratória (a) Rede de Voronoi; (b) Análise de tendências

Na figura 28 são apresentados os resultados de duas ferramentas de análise exploratória. (a) Os polígonos de *Thiessen* ou rede de *Voronoi* que resulta de um processo de interpolação exacto, em que cada polígono assume o valor da variável no ponto conhecido, o estabelecimento da rede de polígonos resulta de uma triangulação que segue o critério de *Delaunay*. É de salientar que com este método verificamos a existência de máximos da variável nas zonas onde anteriormente identificamos maiores afastamentos do gráfico *Normal QQPlot*. (b) O estudo da distribuição tridimensional dos valores da variável e a comparação com a linhas de tendência em cada coordenada. Este gráfico é útil para verificar se existem padrões nos dados que merecem um estudo mais pormenorizado ou indiciem fenómenos localizados.

O ArcGIS tem um auxiliar (*Geostatistical Wizard*) que auxilia o utilizador a seguir os passos de criação do semivariograma experimental e o processo de análise da qualidade do processo. Com o auxílio do *wizard* iremos realizar a interpolação por *kriging* (*ordinary kriging*).

Para realizar a interpolação por *kriging* (*ordinary kriging*) o utilizador necessita de verificar se estão reunidas as condições seguintes através da análise exploratória dos dados (Hengl, 2009, p. 16):

1. A tendência é constante ( $\mu$ =constante);
2. O variograma é constante em toda a área onde se pretende realizar a interpolação;
3. A variável a interpolar segue aproximadamente uma distribuição normal;

O primeiro passo passa por determinar o semivariograma utilizando as diferenças de variâncias como já indicado atrás. Quando colocamos os valores da função semivariograma num gráfico bidimensional verificamos que existe uma dispersão dos dados (com dados reais nunca encontramos a simplicidade do modelo matemático). Por esse motivo, torna-se necessário ajustar os dados a uma curva de valores únicos. O processo de modelação neste caso consiste em encontrar uma função empírica com formulação matemática conhecida (*Circular, Spherical, Tetraspherical, Pentaspherical, Exponential, Gaussian, Rational Quadratic, Hole Effect, K-Bessel, J-Bessel, Stable*) que se ajuste aos valores médios do semivariograma resultante (cruzes azuis da figura 30). Depois de escolhida a equação pretendida, a aplicação determina os coeficientes por um processo de ajustamento baseado no critério dos mínimos quadrados. Neste caso foram experimentadas várias opções tendo-se optado por uma função do tipo K-Bessel<sup>24</sup> ( $0 * \text{Nugget} + 319.59 * \text{K-Bessel}(28391, 0.51051)$ ).

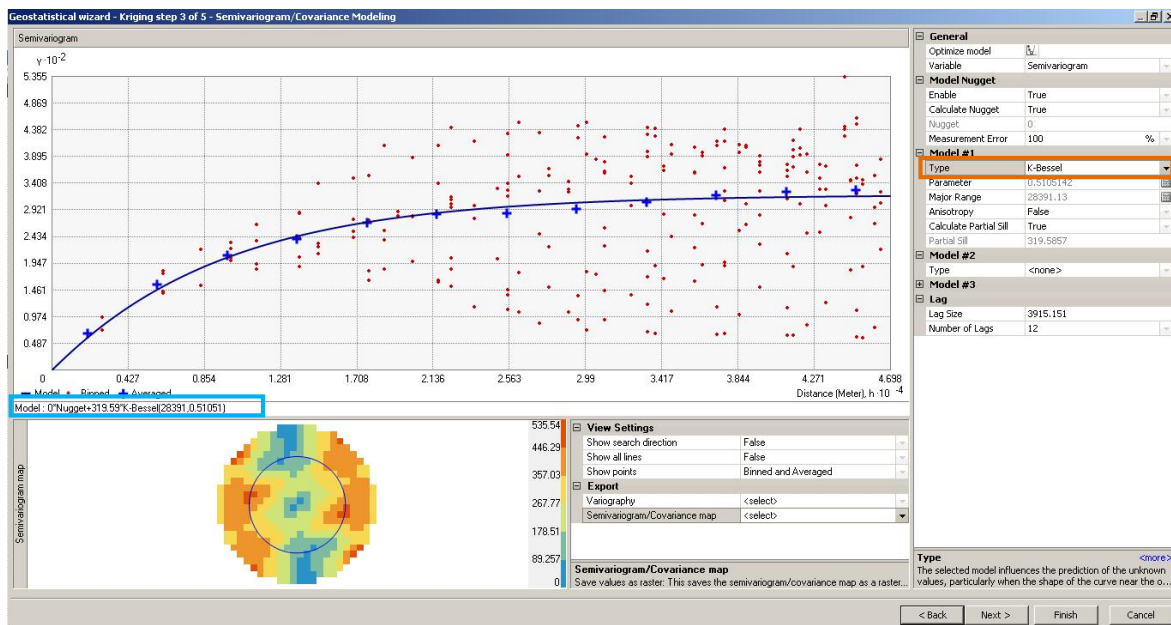


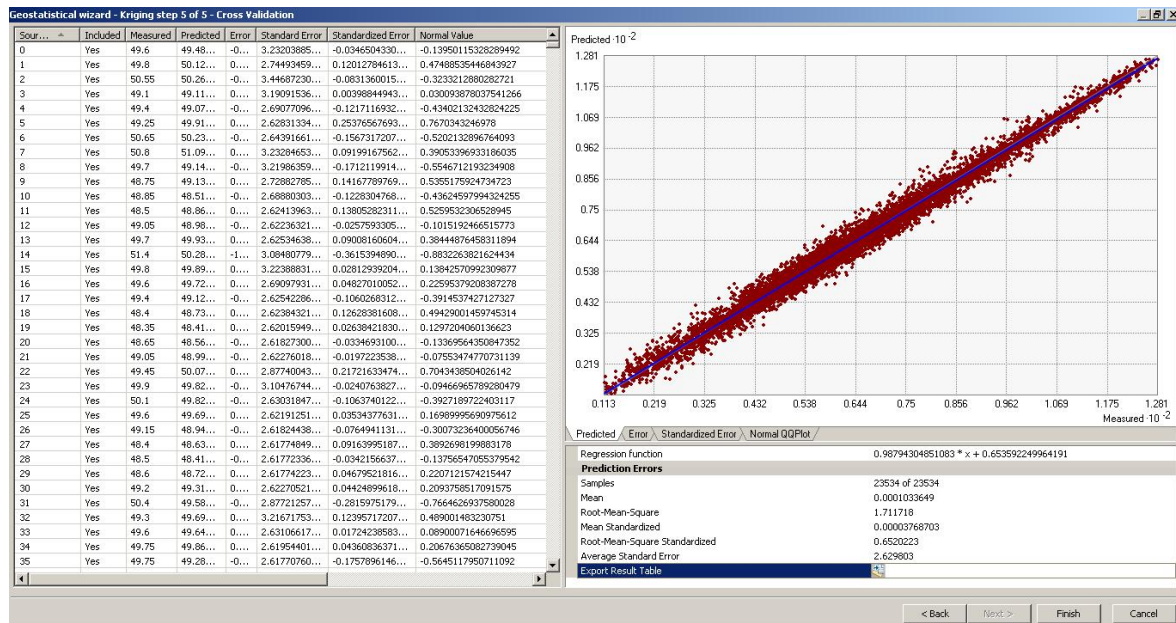
Fig. 29 – Modelação do semivariograma

<sup>24</sup> Sobre *empirical Bayesian Kriging* ver também:

<http://resources.arcgis.com/en/help/main/10.2/index.html#//0031000000q9000000>



Na figura seguinte são apresentados os resultados do processo de controlo de qualidade efetuado por *crossvalidation*<sup>25</sup>. Verificamos que temos uma dispersão de resíduos (de aproximadamente<sup>26</sup> 0.6 (unidades 0.1 mGal). Pensamos que isto estará relacionado com a dimensão espacial da amostra. Porém, no parâmetro global da qualidade *root mean square* obtivemos 1.71 (unidades 0.1 mGal). Podemos também concluir que a resolução espacial da grelha utilizada de aproximadamente 350 em 350 metros é um fator que facilita a qualidade da interpolação, visto que a gravidade é um fenómeno que apresenta uma tendência espacial de variação de reduzido gradiente.

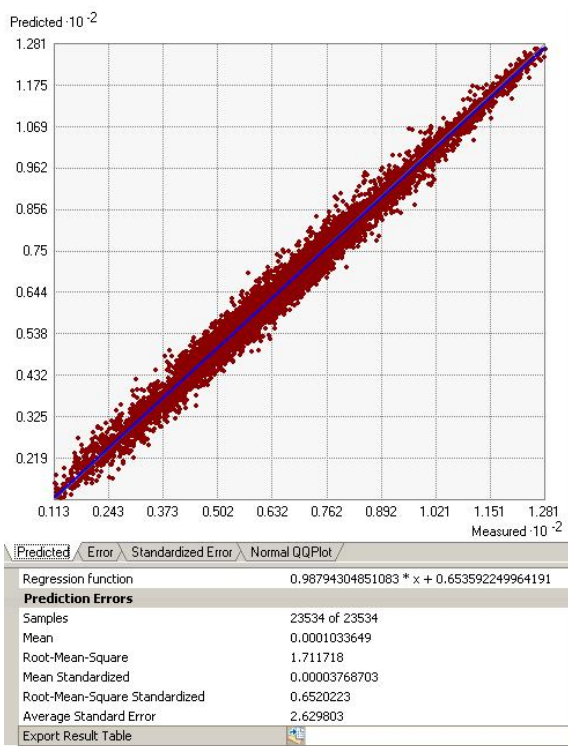


(a)

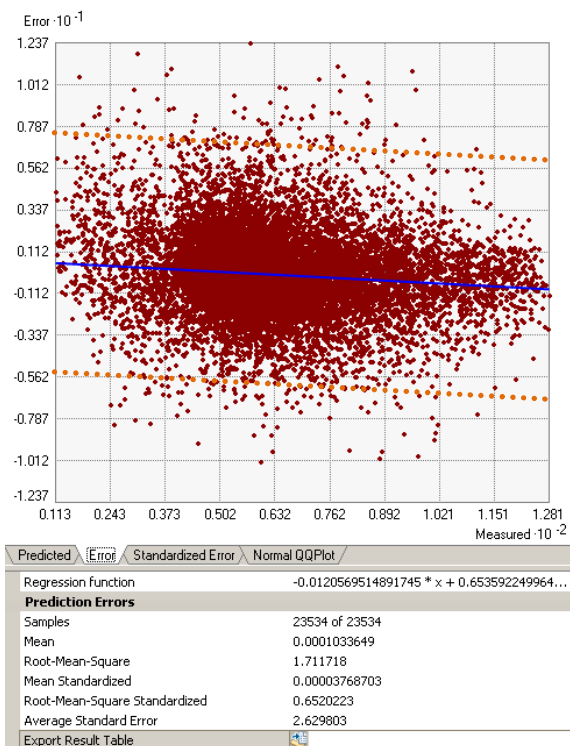
<sup>25</sup> Sobre crossvalidation no ArcGIS ver também:

[http://resources.arcgis.com/en/help/main/10.2/index.html#/Performing\\_cross\\_validation\\_and\\_validation/003100000059000000/](http://resources.arcgis.com/en/help/main/10.2/index.html#/Performing_cross_validation_and_validation/003100000059000000/)

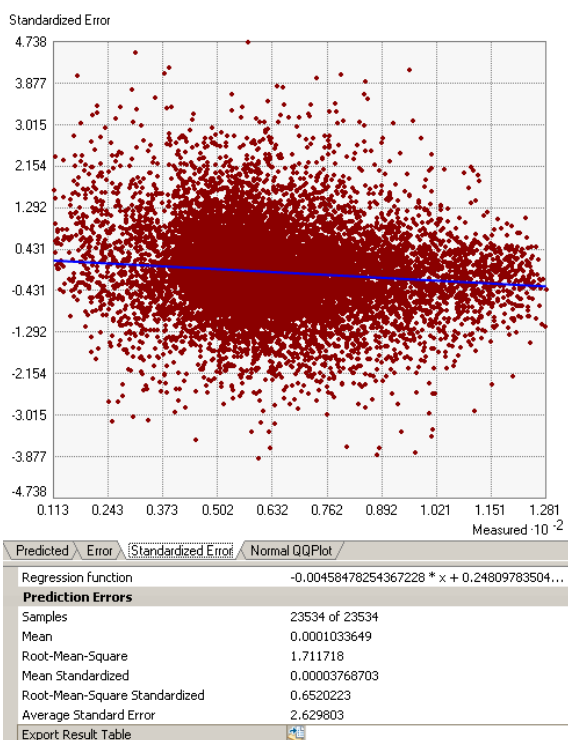
<sup>26</sup> Valor estimado visualmente através do gráfico dos erros.



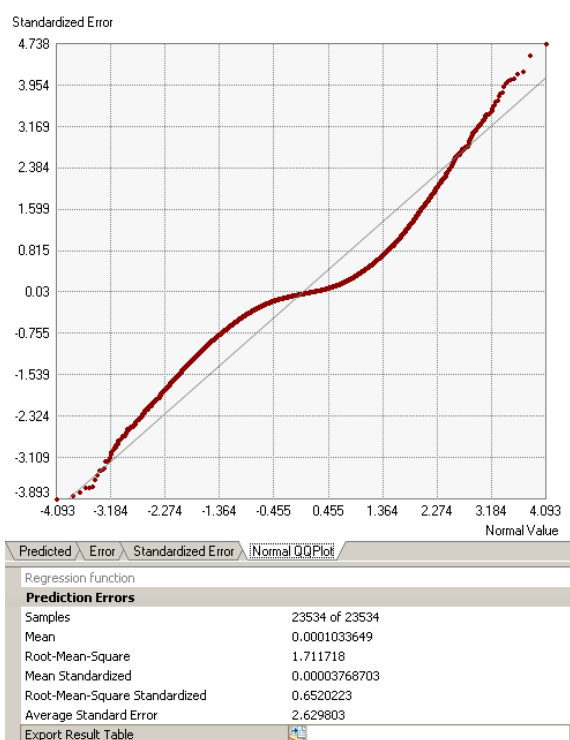
(b)



(c)



(d)



(e)

Fig. 30 – Parâmetros estatísticos da qualidade da interpolação (Crossvalidation)

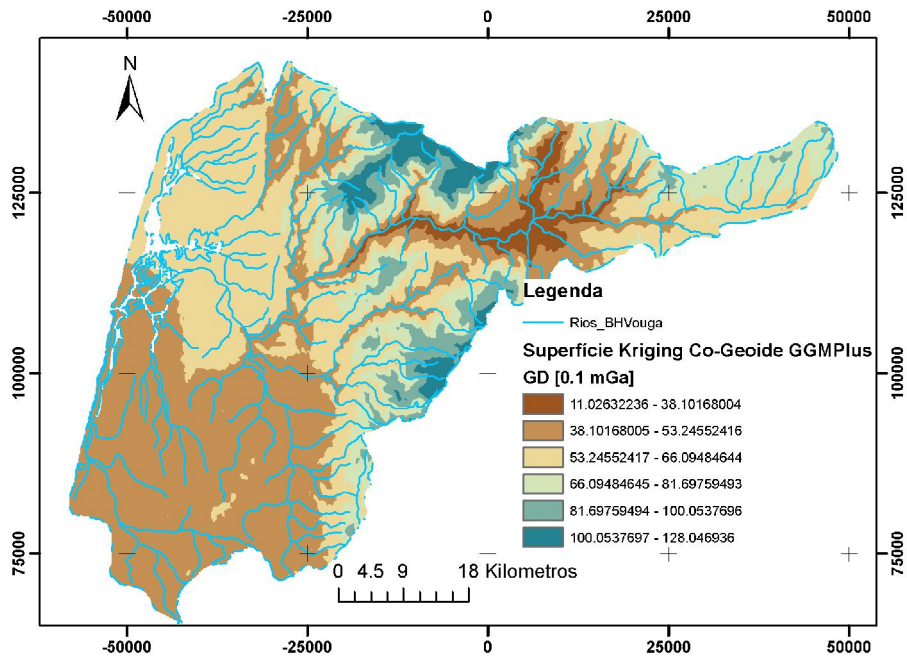
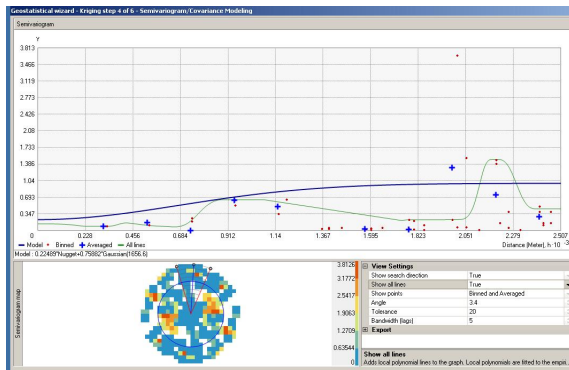
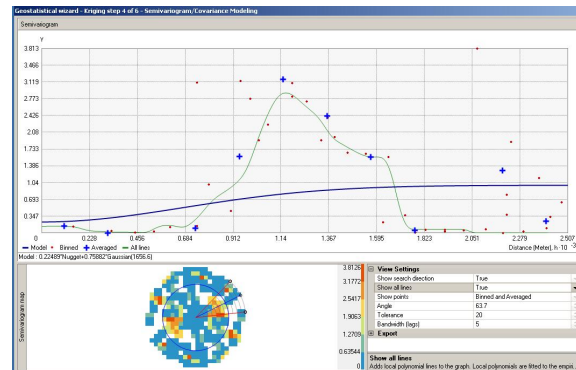


Fig. 31 – Resultado da interpolação de *Kriging* (variograma modelado com uma função do tipo *K-Bassel*) aplicado aos dados de gravidade do co-geoide GGMPPlus na bacia hidrográfica do rio Vouga

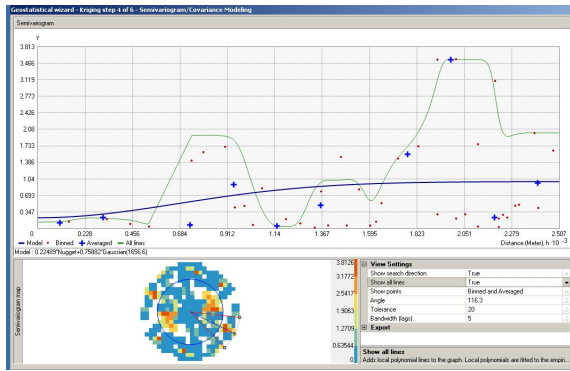
A metodologia de interpolação por *kriging* foi aplicada aos dados da concentração média de nitratos na bacia hidrográfica do rio Vouga tendo-se obtido os seguintes resultados.



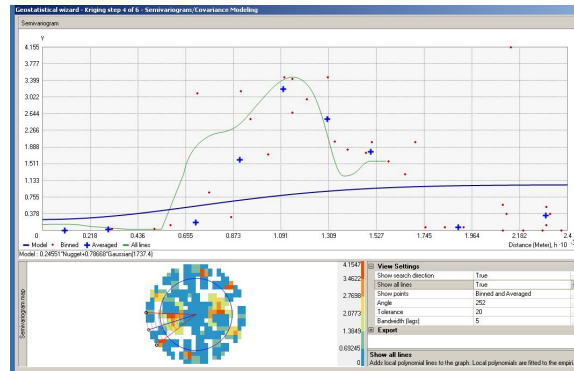
(a)



(b)



(c)

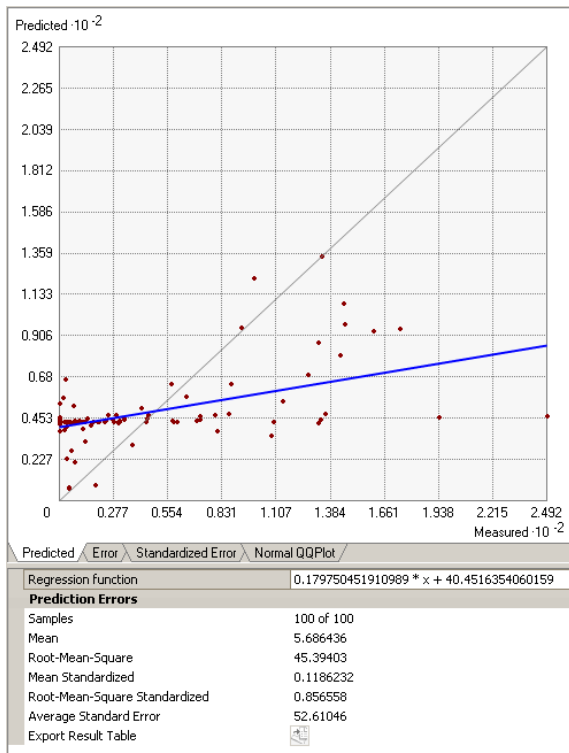


(d)

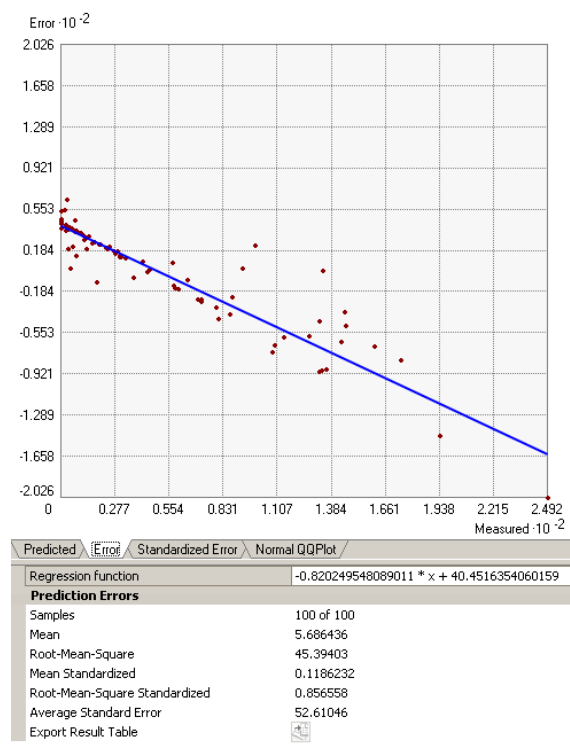
Fig. 32 – Análise exploratória através do semivariograma da amostra de dados da concentração de nitratos em águas subterrâneas.

Neste caso verificamos que estamos na presença de um conjunto de amostras que indicam uma anisotropia zonal do fenómeno. Nas figuras anteriores temos representado o semivariograma em diversas direções.

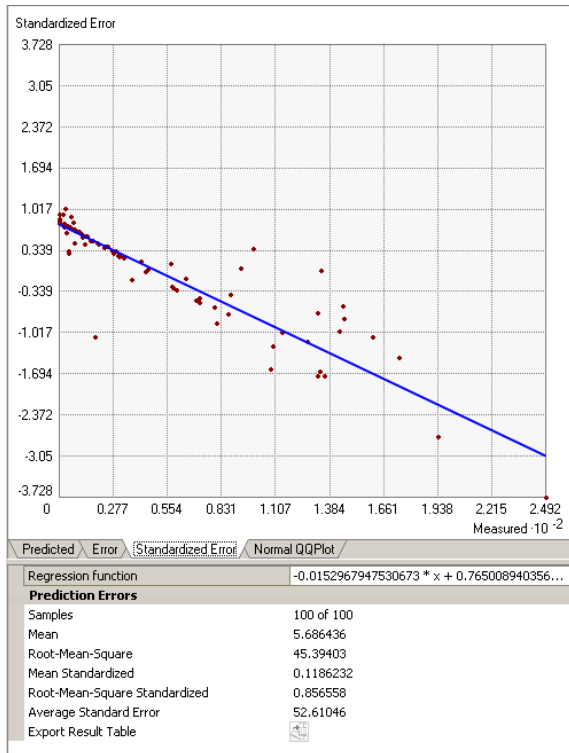
Na imagem seguinte, são apresentados os resultados do método de interpolação por *kriging* com a aplicação de um só modelo (simulando a isotropia do fenómeno). Nestas condições estimamos que a superfície resultante da interpolação será de má qualidade. Esta estimativa é comprovada pela análise dos resultados da *crossvalidation*, encontramos um grande desfasamento entre a curva de previsão (cinza na figura 33 (a)) e a curva de ajustamento aos dados (azul na figura 33 (a)), isto também pode ser observado no erro médio quadrado resultante que é de 45.4 mg/l, um valor muito elevado. Este é um exemplo em que para melhorar o processo temos de considerar a anisotropia do processo e modelar o fenómeno de forma diferencial.



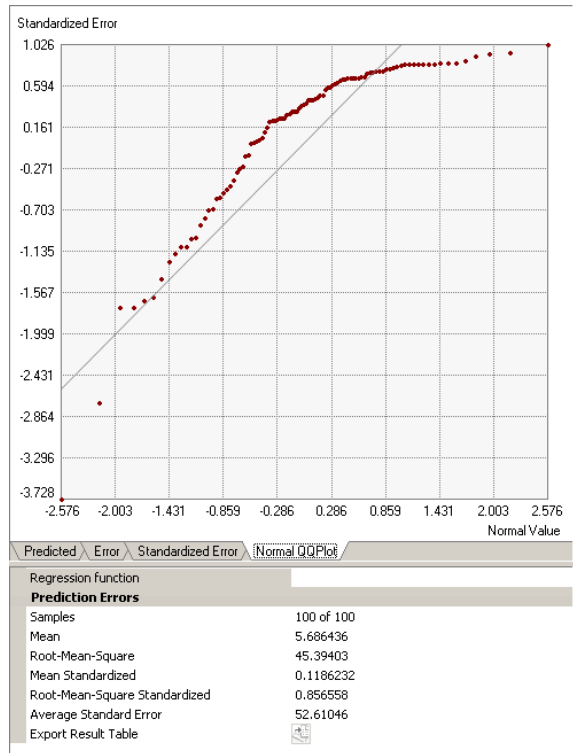
(a)



(b)



(c)



(d)

Fig. 33 – Resultados do crossvalidation para os dados de concentração de nitratos nas águas subterrâneas

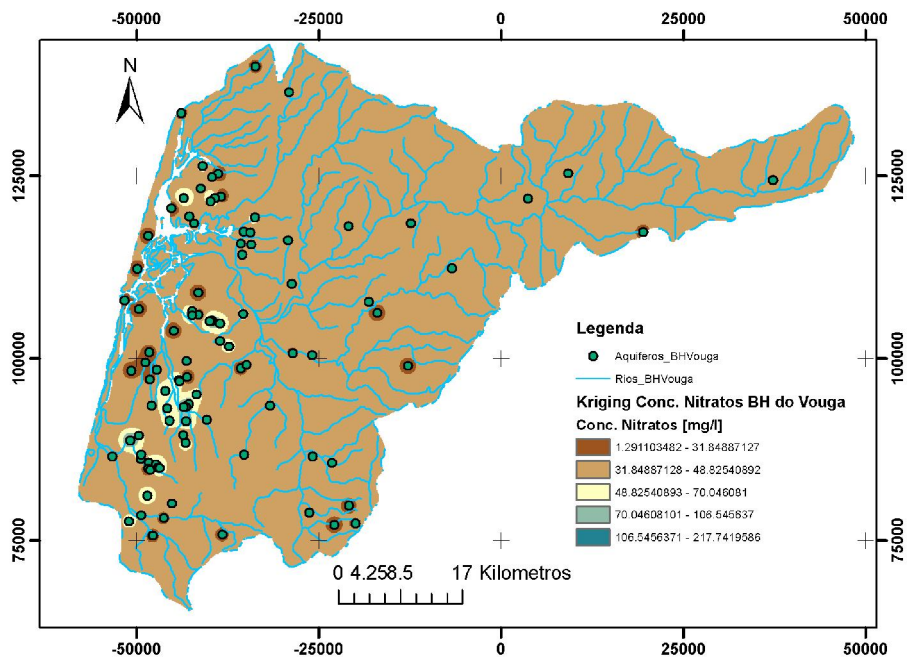
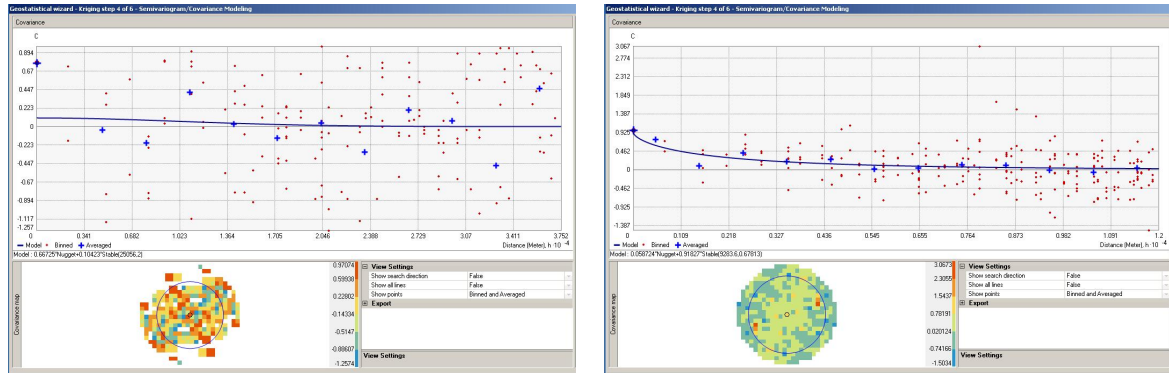


Fig. 34 – Resultado da interpolação por kriging considerando o fenómeno com isotrópico

Para terminar este bloco, iremos apresentar um terceiro conjunto de dados, os dados de precipitação em vinte estações meteorológicas distribuídas na bacia hidrográfica do rio Vouga.

Da análise do semivariograma verificamos que estamos na presença de um fenómeno que apresenta uma reduzida auto correlação espacial, com a covariância a tender rapidamente para zero, é um exemplo onde a utilização de um processo de interpolação estocástico não traz melhoria. Por esse motivo optou-se por realizar uma interpolação determinística por IDW.



(a)

(b)

Fig. 35 – Análise exploratória do semivariograma para interpolação dos dados meteorológicos da observação da precipitação média diária em 20 estações da bacia hidrográfica do rio Vouga

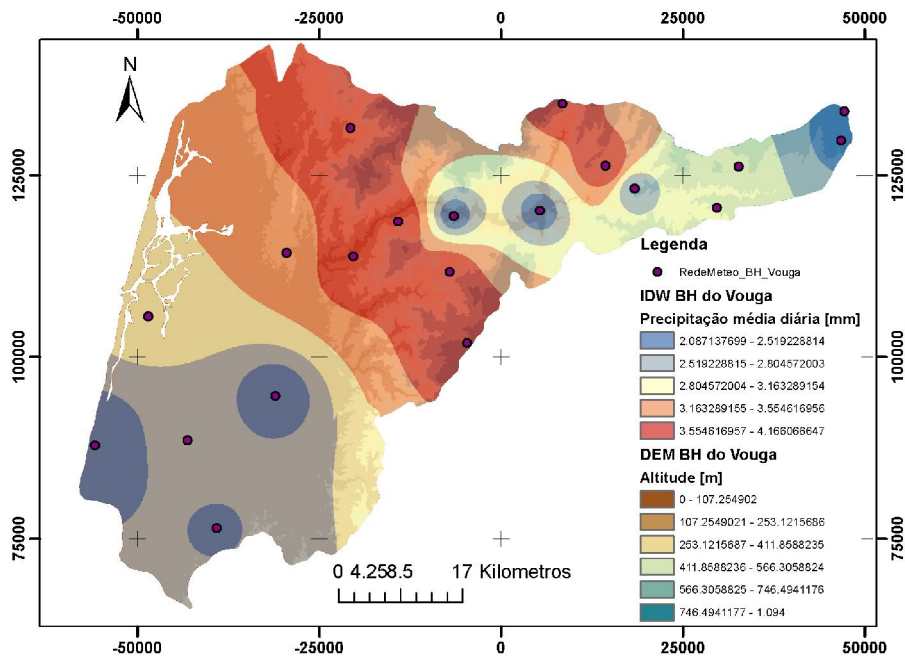


Fig. 36 – Superfície resultante da interpolação determinística por IDW na precipitação média diária na bacia hidrográfica do rio Vouga.

Analisando a superfície resultante sobreposta ao modelo digital de elevação verificamos que os valores mais elevados da precipitação ocorrem na vertente oeste da serra de Freita onde os ar húmido oceânico será forçado a subir o declive acentuado da serra decrescendo a sua temperatura a potenciando a precipitação. Este fator de natureza orográfica deverá ser preponderante para justificar o pico de precipitação que aí se verifica.

## 4.2 – Análise de Redes:

Neste capítulo do relatório iremos abordar os algoritmos de análise de redes. A rede «é uma estrutura [organizada e classificada topologicamente] que representa a interação ou movimento entre determinadas localizações» (Catita, 2014). Uma rede é constituída por nós, localizações pontuais que marcam o início, transmissão ou fim do fluxo; por arcos, são os elementos condutores do fluxo e garantem a ligação entre nós.

As redes são elementos muito importantes em plataformas SIG que permitem responder aos problemas seguintes:

- Qual o melhor caminho (mais próximo / menor custo) entre o ponto A e B?
- Com base na rede onde se devem situar os serviços?
- Onde está localizado o serviço que serve um determinado local?
- Qual a acessibilidade de um local a partir de outro?
- Etc...

A análise de redes em SIG baseia-se na teoria de grafos e nos algoritmos que foram sendo criados nesse campo da matemática. Mantendo o procedimento que tem sido adotado nos capítulos anteriores não iremos aprofundar a formulação matemática dos algoritmos, focando a descrição do trabalho nas aplicações práticas. Existe muita bibliografia dedicada a estes assuntos, no entanto, recomenda-se (Matos, 2008, pp. 104 - 129)

Os dados utilizados a esta fase do trabalho foram recolhidos das várias fontes que se encontram na tabela seguinte:

Tabela VII – Fontes dos dados utilizados na análise de redes

Dados	Fonte de dados	Data de recolha	Observações
Rede viária do concelho de Lisboa	<i>OpenStreetMap</i> : <a href="http://www.openstreetmap.org/">http://www.openstreetmap.org/</a> <a href="http://download.geofabrik.de/">http://download.geofabrik.de/</a>	Junho de 2014	Rede viária foi recolhida em formato OSM e foi realizado o corte da <i>layer</i> pelos limites do distrito de Leiria.
Posições das corporações de bombeiros	Portal dos bombeiros portugueses <a href="http://www.bombeiros.pt/corpos-de-bombeiros">http://www.bombeiros.pt/corpos-de-bombeiros</a> GPSVisualizer <a href="http://www.gpsvisualizer.com/geocoder/">http://www.gpsvisualizer.com/geocoder/</a>	Dezembro de 2013	A partir das moradas foi necessário utilizar uma ferramenta em linha de <i>geocoding</i> que associasse

			as coordenadas geográficas à morada.
Posições dos hospitais do concelho de Lisboa	As localizações dos hospitais no município de Lisboa foram recolhidos a partir do <i>Google Maps</i> <a href="https://www.google.pt/maps/preview">https://www.google.pt/maps/preview</a>	Junho de 2014	
Posições dos centros de freguesia com maior número de idosos por unidade de área	Informação recolhida no sítio do Instituto Nacional de Estatística <a href="http://censos.ine.pt/xportal/xmain?xpid=CENSOS&amp;xpgid=censos_quadros">http://censos.ine.pt/xportal/xmain?xpid=CENSOS&amp;xpgid=censos_quadros</a>	junho de 2014	Dados referentes aos Censos de 2011
Organização administrativa de Portugal CAOP2011	Carta Administrativa Oficial de Portugal - Versão 2011 <a href="http://tinyurl.com/lr49zcn">http://tinyurl.com/lr49zcn</a>	Junho de 2014	

O enquadramento que serve de base a esta fase do trabalho é a existência de freguesias no município de Lisboa que apresentam uma elevada densidade de população idosa, suscetível à futura necessidade de apoio dos serviços de emergência médica. Com esse objetivo, pretende-se identificar quais as freguesias de Lisboa que apresentam maior número de idosos por unidade de área e verificar quais as corporações de bombeiros mais próximas, assim como, quais as áreas que são servidos pelos diversos hospitais que estão espalhados pelo município de Lisboa. A necessidade deste tipo de estudo surge na identificação prévia de planos de apoio à população idosa.

Para fazer a análise indicada, iremos utilizar os algoritmos disponíveis na ferramenta *network analyst* do ArcGIS 10.2 para fazer uma análise conjunta dos dados da rede viária, localização dos bombeiros e hospitais, e centros das freguesias com maior densidade de população idosa (população com 65 ou mais anos). Pretendem-se realizar duas tarefas de análise, respondendo às seguintes questões:

1. Quais as corporações de bombeiros que se encontram mais próximas dos centros de freguesia que apresentam uma densidade de população idosa superior a 20 idosos por hectare?
2. Qual a área servida pelos diversos hospitais do município de Lisboa?

Os dados da rede viária foram recolhidos a partir da comunidade *open street map* em formato .osm que tem a rede viária europeia. Os dados em formato OSM podem ser abertos de duas formas utilizando o QGis ou diretamente no ArcGIS através do *Editor for OSM*<sup>27</sup>. Após a importação dos dados verificou-se que estes apresentavam inconsistências que impossibilitavam a análise e aplicação dos algoritmos de rede. O objetivo deste trabalho não é o tratamento da informação de base, por esse motivo, abandonou-se a ideia original e utilizou-se a rede viária que foi disponibilizada no âmbito da disciplina de Deteção Remota, proveniente do Instituto Geográfico do Exército em Datum 73.

Foram realizadas as operações necessárias a ter todas as fontes de informação em PTTM06. As freguesias da CAOP 2011 foram juntadas às tabelas de dados instituto nacional de estatística

<sup>27</sup> Ver [http://wiki.openstreetmap.org/wiki/ArcGIS\\_Editor\\_for\\_OSM](http://wiki.openstreetmap.org/wiki/ArcGIS_Editor_for_OSM) (acedido em junho de 2014)



relativos à quantidade de população idosa de forma a calcular a variável densidade populacional de população idosa nas freguesias do concelho de Lisboa<sup>28</sup>.

Depois de ter uma *shapefile* com a rede viária para do município de Lisboa, foi necessário conduzir operações de construção de um *network dataset* utilizando o *ArcCatalog* e permite obter uma rede organizada através da *layer* vetorial da rede viária e iniciar as operações de análise, aplicando algoritmos.

A resposta à nossa primeira questão é conseguida com a aplicação do algoritmo *Closest Facility* onde se pretende encontrar as corporações de bombeiros que estão mais perto dos centros das freguesias com maior densidade populacional. Obtiveram-se os resultados apresentados de seguida.

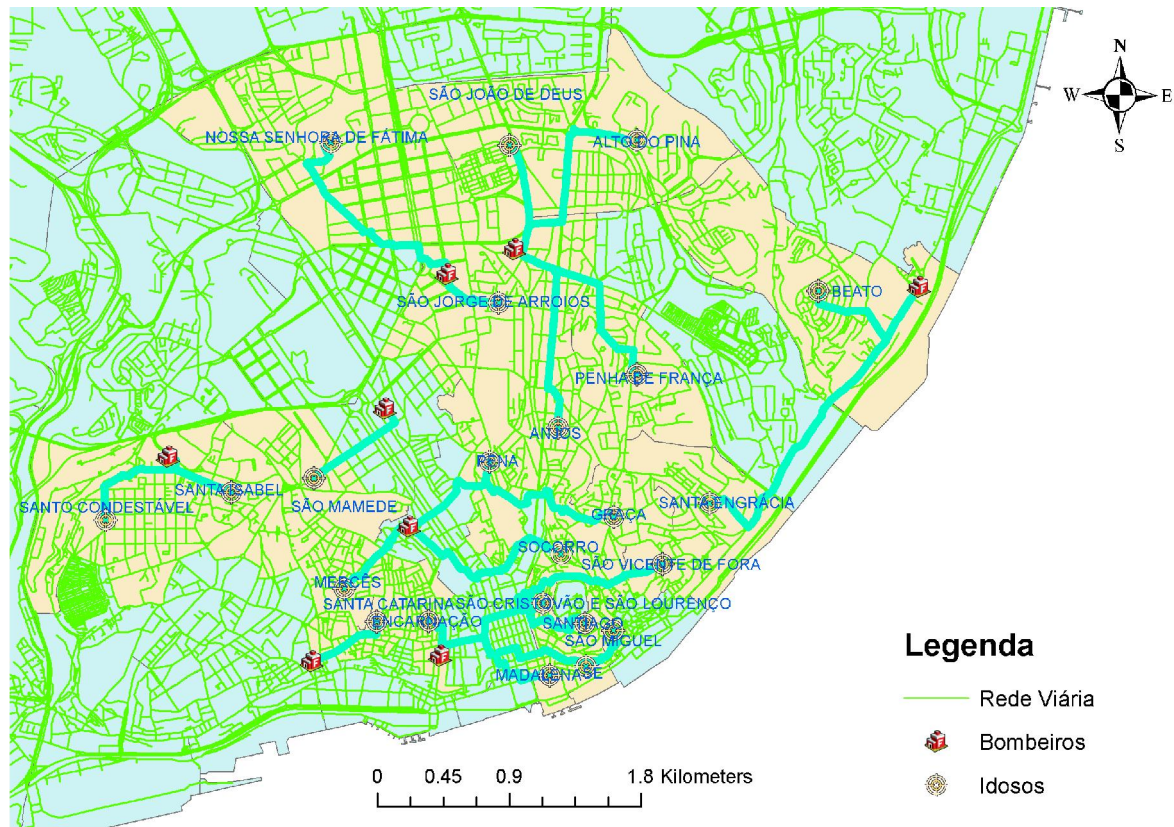


Fig. 37 – Resultado da aplicação do algoritmo *Closest Facility* (a amarelo estão representadas as freguesias do município de Lisboa com mais de 20 idosos por unidade de área [hectare])

Na fase seguinte do trabalho pretende-se responder à pergunta dois, recorrendo ao algoritmo de análise de redes: *Service Area*. Com este algoritmo pretendemos saber criar as áreas de influência dos hospitais, ou seja, vão ser gerados polígonos que se encontram respetivamente a 1 000, 2 500 e 5 000 metros dos hospitais. Com este algoritmo obteve-se o resultado que se encontra na figura seguinte:

<sup>28</sup> Para mais informação sobre a conversão de formatos recomenda-se: <http://wiki.openstreetmap.org/wiki/Shapefiles> (acedido em junho de 2014).

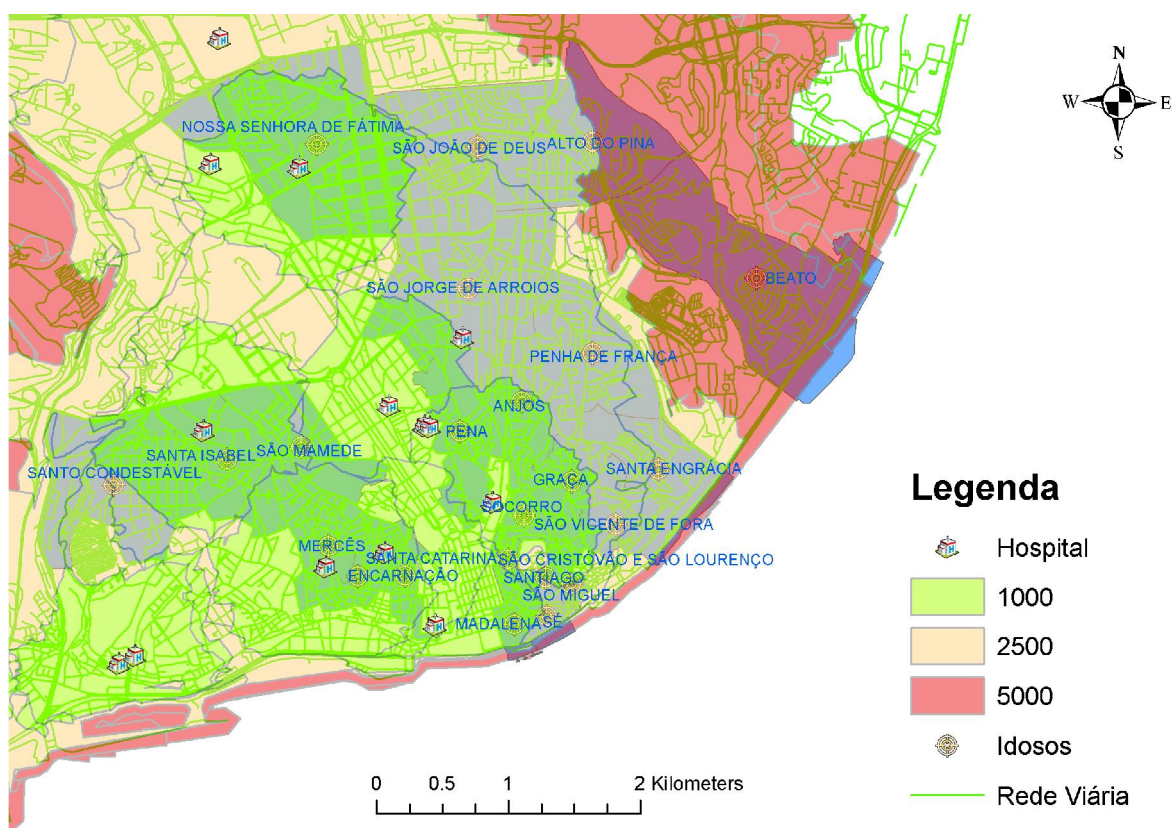


Fig. 38 - Resultado da aplicação do algoritmo *Service Area* (a azul estão representadas as freguesias do município de Lisboa com mais de 20 idosos por unidade de área [hectare])

Da análise da figura anterior concluímos que existem duas freguesias com elevada densidade de população idosa que são parcialmente cobertas pela área de influência a 2,5 km dos hospitais (referimo-nos à freguesia do Beato e Alto do Pina). Fica demonstrado que uma análise simples permite como a apresentada permite ajudar os decisores num melhor planeamento e ordenamento dos serviços.

## 5 – Conclusões

A análise de dados é realizada em vários domínios do conhecimento com o objetivo de construir modelos preditivos dos fenómenos. No nosso caso realizámos a análise de dados espaciais, de padrão pontual e agregados por áreas. De seguida, efetuaram-se operações de interpolação para construir superfícies contínuas, por fim desenvolvemos uma análise simples da localização de serviços com base em algoritmos de análise de redes.

No decurso do trabalho foi possível concluir que existem variáveis que apresentam uma dependência espacial e que a sua ocorrência só pode ser corretamente modelada com a inserção do fator espaço e sua variação. Ao longo do trabalho verificou-se que é importante desenvolver processos preliminares de análise espacial que permitam caracterizar estatisticamente as amostras.

Na primeira fase do trabalho, foi realizada a análise de padrões pontuais, no caso, concentrações máximas de nitratos nas reservas subterrâneas de água. Os métodos aplicados descreveram os dados como sendo de aglomerados. Coincidentes com as zonas onde encontramos um elevado esforço agrícola.

Na segunda fase, foram utilizadas técnicas de análise de dados agregados de padrão social ao nível do município. Os métodos aplicados indicaram a presença de aglomerados, de zonas em que as

variáveis apresentavam as mesmas características que a sua vizinhança. E fenómenos de auto correlação espacial.

Na terceira fase, aplicaram-se técnicas que visavam a criação de modelos simples de regressão clássica e espacial. No fenómeno analisado de crimes de violência doméstica, foi possível verificar que os resíduos não apresentavam uma correlação espacial. Esses resultados foram submetidos ao teste de *Moran* que comprovou que na modelação do fenómeno não é necessário inserir o fator espaço no modelo de regressão.

Na quarta fase, realizaram-se operações de interpolação utilizando métodos deterministas e probabilísticos para gerar superfícies contínuas que permitam obter o valor esperado da variável em estudo em toda a área. Neste bloco de tarefas concluímos que os processos de interpolação devem ser sempre precedidos de uma análise exploratória que permite inferir a natureza do fenómeno e seleccionar o método de interpolação que melhor serve o objetivo. E que a interpolação deverá ser sempre seguida de um processo de análise de qualidade dos resultados que permite verificar a eficiência do processo utilizado através da exatidão da superfície resultante.

Na quinta e última fase, realizou-se a análise da localização de serviços de emergência médica no município de Lisboa com base em algoritmos de redes. Analisou-se se a actual localização dos hospitais e corporações de bombeiros correspondem às necessidades das freguesias com população mais envelhecida. Neste caso, concluiu-se que a atual rede de hospitais e bombeiros garante as necessidades gerais da população idosa, sendo que identificou duas freguesias, o Beato e o Alto do Pina que se situam a mais de 2,5 km do hospital mais próximo.

Como conclusão geral, ao longo do trabalho verificamos que os modelos da realidade são aproximações do mundo real, pretendem-se que sejam o mais precisos possível. Contudo, nenhum consegue modelar a totalidade da complexidade da realidade. Nos modelos matemáticos e estatísticos convencionais é normalmente desprezada a influência do espaço quando ela existe como demonstrado. As aplicações SIG têm vindo a integrar ferramentas de geostatística e análise espacial que permitem inferir e modelar a influência do espaço nos fenómenos. Por esse motivo, são poderosas ferramentas de modelação com provas dadas para a análise, previsão, planeamento e decisão.

## Referências Bibliográficas

- Agência Portuguesa do Ambiente. (1995). "Sistema Nacional de Informação de Recursos Hídricos." No endereço: <http://snirh.apambiente.pt/index.php?idMain=> (Acedido março de 2014).
- Arizon State University. "GeoDa center for geospatial analysis and computation "; No endereço: <http://geodacenter.asu.edu/> (Acedido abril de 2014).
- Catita, Cristina (2014). Apontamentos das Aulas de Análise Espacial de Informação Geográfica. Lisboa, Faculdade de Ciências da Universidade de Lisboa.
- DGT. (2013). "Rede Geodésica Nacional." No endereço: [http://www.dgterritorio.pt/cartografia\\_e\\_geodesia/geodesia/redes\\_geodesicas/rede\\_geodesica\\_nacional/](http://www.dgterritorio.pt/cartografia_e_geodesia/geodesia/redes_geodesicas/rede_geodesica_nacional/) (Acedido junho de 2014).
- Druck, Suzana, M. S. Carvalho, *et al.* (2004). Análise Espacial de Dados Geográficos. Brasília, EMBRAPA.
- Duarte. (2010). "Blog Via SIG." No endereço: <http://blog.viasig.com/2010/03/mdt-30m-para-portugal/> (Acedido junho de 2014).
- ESRI. (2013). "ArcGIS Help Library." No endereço: <http://resources.arcgis.com/en/help/main/10.1/index.html#/> (Acedido abril de 2014).

- Fischer, Manfred M. and Jinfeng Wang (2011). Spatial Data Analysis, Models, Methods and Techniques. London, Springer.
- Hengl, Tomislav (2009). A practical Guide to Geostatistical Mapping. Amsterdam.
- Matos, João (2008). Fundamentos de Informação Geográfica, LIDEL.
- Ministério da Agricultura do Desenvolvimento Rural e das Pescas (1997). Código de Boas Práticas Agrícolas para a protecção da água contra a poluição com nitratos de origem agrícola.
- NASA. (2004). "ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer." No endereço: <http://asterweb.jpl.nasa.gov/gdem.asp> (Acedido junho de 2014).
- Parlamento Europeu (2000). Quadro de acção comunitária no domínio da política da água. **2000/60/CE**.
- Parlamento Europeu (2006). Directiva relativa à protecção das águas subterrâneas contra a poluição e a deterioração. **2006/118/CE**.
- QGIS ORG QGIS - A Free and Open Source Geographic Information System.
- Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Nova Iorque, Chapman & Hall/CRC.
- Soares, Amílcar (2006). geoestatística para as ciências da terra do ambiente. Lisboa, IST Press.