

# Modelação Espacial da Concentração de Sulfatos na Bacia Hidrográfica do Tejo

Nogueira, W., Alves, G., Negreiros, J.  
Universidade de Cabo Verde / UniCV  
Campus de Palmarejo  
Cidade da Praia, Santiago, Cabo Verde

Negreiros, J.

Instituto Superior de Estatística e Gestão de Informação – Universidade Nova de Lisboa / ISEGI-UNL  
Campus de Campolide  
1070-312 Lisboa, Portugal

**Resumo:** Segundo Isaaks and Srivastava (1989), a geoestatística permite descrever a continuidade espacial e adapta as técnicas de regressão clássica para tirar vantagem dessa continuidade. Neste artigo, foi realizada uma delimitação espacial da bacia hidrográfica do Tejo e, através das amostras relativo à medição da concentração de sulfatos nas águas subterrâneas, foi estimado os seus valores para a bacia em causa. Assim, realizou-se a Krigagem ordinária (OK) e a de indicatriz (IK). Verificou-se que ambos os estimadores apresentam características e objectivos diferentes e que a sua aplicabilidade é enorme (agricultura, ambiente, meteorologia, hidrologia, mineração) quando comparada com a estatística tradicional.

**Palavras-chave:** Geoestatística, Geostatistical Analyst<sup>®</sup>, Interpolação Espacial, Kriging, Sulfatos, Bacia Hidrográfica do Tejo.

## 1. ENQUADRAMENTO

Por definição, a geoestatística é a aplicação do formalismo de funções aleatórias para o reconhecimento e estimação de fenómenos naturais. Segundo Deutch and Journel (1998), a geoestatística teve as suas primeiras aplicações em mineração e hidrologia. No início da década de 60, o Professor Matheron e Krige, trabalhando com dados de concentração do ouro na África de Sul, concluíram que somente a informação dada pela variância seria insuficiente para explicar o fenómeno em estudo. Para tal, seria necessário levar em consideração a distância entre as observações e a sua dependência espacial. Confirmado por Ribeiro (2006), as ferramentas da geoestatística permitem a análise da dependência espacial, como também a estimativa de dados para locais não amostrados através de um estimador sem tendenciosidade.

Este trabalho tem como objectivo medir o teor de sulfatos na rede hidrográfica da bacia do rio Tejo, recorrendo a técnicas de análise geoestatísticas. O ião sulfato é um dos iões mais abundantes nas águas naturais e, normalmente, tem origem em formações geológicas (dissolução de gessos, oxidação

de perites, erosão de granito e basalto), inquinação por dejectos, utilização de fertilizantes como o sulfato amónio ou infiltração da água do mar. O problema é que o elevado teor de sulfatos na água podem causar sérios problemas tais como mau cheiro e corrosão em tubagens de águas residuais devido à existência de bactérias que infectam as suas paredes laterais.

Segundo a Figura 1, foi executado o levantamento de dados a partir da base de dados do Atlas do Ambiente, onde se diligenciou o download dos ficheiros shapefiles (.shp) relativos ao nível de nitratos da rede hidrográfica de Portugal (<http://www.igeo.pt/atlas/index1.html>). Dado que o shapefile residia em formato poligonal, foi necessário transformar os mesmos polígonos em centróides de modo a poder realizar a análise exploratória de dados (secção 2 e 3), variografia (secção 4), estimação (secção 5) e conclusão (secção 6). Note-se que a realização deste processo foi apenas possível através da utilização do módulo de software Geostatistic Analyst® da ESRI®. É também crucial referir que se assume na leitura deste documento que o leitor já apresenta os conhecimentos básicos de Sistemas de Informação Geográficos e estatística descriptiva.



Figura 1. Esquema estrutural do estudo em causa.

## 2. ÁREA DE ESTUDO

Sendo a quinta maior bacia hidrográfica da Europa Comunitária, a bacia do Tejo cobre uma superfície total de 80629 Km<sup>2</sup>, dos quais 24800Km<sup>2</sup> (29,8%) reside em Portugal. Esta apresenta um largo corredor no centro-oeste da Península Ibérica com uma orientação de Este/NorteEste para Oeste/SulOeste e 700 Km de comprimento (230 Km em Portugal e 43 Km de fronteira). A nascente reside nos Montes Universais e Serranias de Cuenca (perto de Albarracín, a cerca de 1600 metros de altitude). Pela sua disposição e por apresentar regularmente alguma elevação, quer em Portugal quer em Espanha, a sua faixa montanhosa a norte contribui com afluentes caudalosos, beneficiando de maiores precipitações. Globalmente, o relevo da bacia é acentuado, por vezes vigoroso, atingindo na Cordilheira Central (Estrela, Gredos, Guadarrama, Ayllón, Serra de Gredos) uma altitude superior a 2,6 Km. As zonas planas ou pouco acidentadas são raras com excepção na região Portuguesa, a jusante da

confluência com o rio Zêzere, findando nas Lezírias e Terrços do Tejo onde decresce suavemente até à foz (Instituto da Água, 2000).

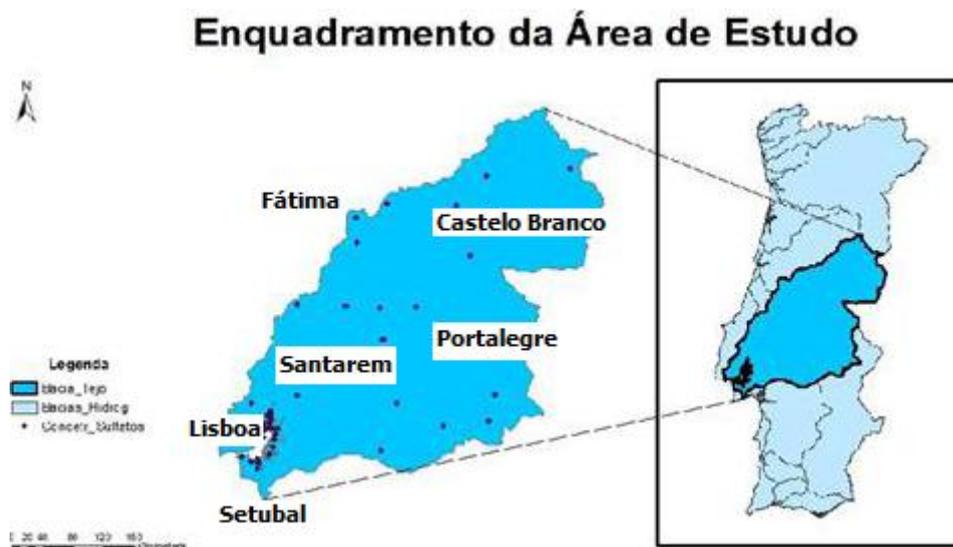


Figura 2. Localização das estações de sulfatos.

### 3. ANÁLISE EXPLORATÓRIA DE DADOS

A AED é um conjunto essencial de funções que permite organizar, resumir e interpretar dados num conjunto de procedimentos estatísticos para compreensão dos dados de input. Tem por objectivo caracterizar as propriedades dos dados, sintetizar as características da sua distribuição empírica (comportamento médio, simetria, dispersão...), apoiar hipóteses que suportam outros modelos, representar relações/correlações entre variáveis e comparar distribuições. Neste caso particular, computou-se os parâmetros principais da análise descritiva, histograma, QQPlot e a análise de tendência (Trend Analysis) pelo Geostatistic Analyst<sup>®</sup> (<http://www.esri.com/software/arcgis/extensions/geostatistical/index.html>) da ESRI<sup>®</sup>.

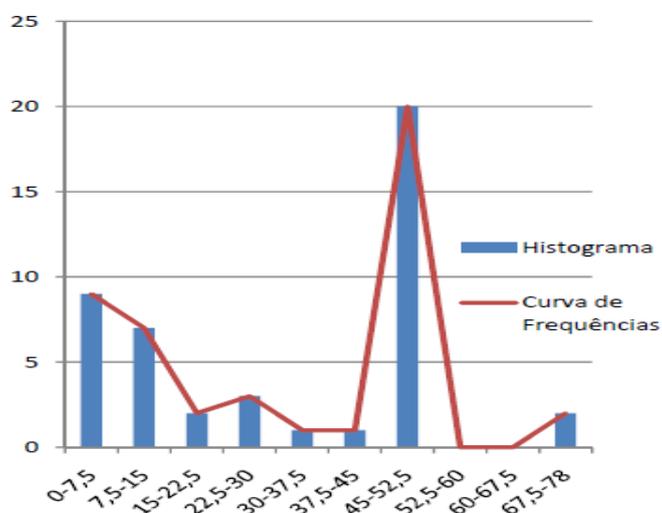


Figura 3. Histograma de 10 classes das 45 amostras de concentração de sulfato.

O coeficiente de Skewness (0,07962) é um valor ligeiramente positivo, o que indica que a distribuição está desviada ligeiramente para a direita. Tecnicamente, considera-se a distribuição Platicúrtica. O valor médio das concentrações de sulfatos na bacia hidrográfica do Tejo é 31,667 mg/l, o desvio padrão é 20,642 mg/l enquanto a mediana é 40 mg/l (50% das amostras recolhidas tem uma concentração inferior a 40 mg/l). O primeiro Quartil é 10 mg/l (25% das amostras têm uma concentração de sulfatos inferior a 10 mg/l) e o 3º Quartil é 40 mg/l (75% das amostras têm uma concentração de sulfatos inferior a 40 mg/l). Como seria de esperar, o intervalo inter-quartil é 30 mg/l. O valor máximo e mínimo são 77,5 mg/l e 2,5 mg/l, respectivamente (a amplitude de variação é, obviamente, 75 mg/l).

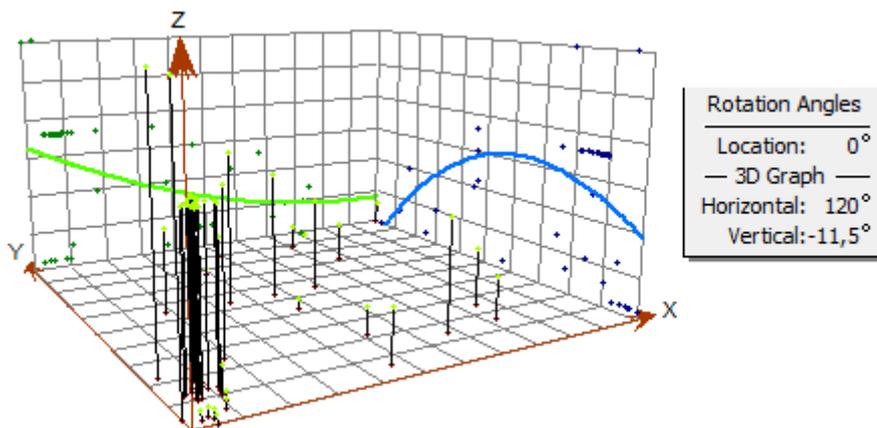


Figura 4. Trend analysis do Geostatistic Analyst® da ESRI®.

Na função Trend Analysis, o eixo vertical representa a localização e o valor de cada amostra no espaço e projectado em planos perpendiculares (Este-Oeste e Norte-Sul). No caso de a linha ser achatada, esta indica a não existência de qualquer tendência. Neste caso particular, a linha de tendência no plano Este-Oeste inicia-se com um valor alto e vai decrescendo, indicando que as amostras obtidas têm um valor alto a Este mas diminuindo em direcção a Oeste. O plano Norte-Sul, com uma linha de tendência curva, regista um aumento no centro e uma diminuição do factor de poluição nos extremos.

#### 4. VARIOGRAFIA

A etapa da variografia apresenta como objectivo avaliar os principais padrões espaciais do fenómeno físico e desenvolver um modelo matemático de correlação espacial que caracteriza a região de estudo. Assim, o variograma permite descrever quantitativamente a variação no espaço de um fenómeno regionalizado (Goovaerts, 1997). A sua natureza estrutural é definida a partir da comparação de valores tomados simultaneamente em dois pontos, segundo uma determinada direcção. Matematicamente, o variograma  $\lambda(h)$  é definido como sendo a esperança do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância  $h$  (ver Equação 1), onde  $N(h)$  é número de pares para uma distância  $h$  e  $z(u)$  representa o valor da amostra num determinado local ou *site*. Sendo esta função uma medida de variância, o seu comportamento rege-se por baixos valores

para distâncias de h curtas, mas que vai aumentando à medida que os pontos se distanciam entre si. Ao utilizar-se esta fórmula nas principais direcções de continuidade do fenómeno, é possível calcular as variações médias dos corpos ao longo de cada uma das direcções. Assim, quando a variabilidade é sensivelmente igual em todas as direcções, o modelo é designado isotrópico e depende somente do comprimento do vector h. Caso contrário, o modelo é considerado anisotrópico (Soares, 2000). Os principais modelos matemáticos de variograma são o esférico, exponencial, gaussiano e potêncial.

$$\lambda(h) = \frac{1}{N(h)} \sum_{N(h)} [z(u) - z(u+h)]^2 \quad (1)$$

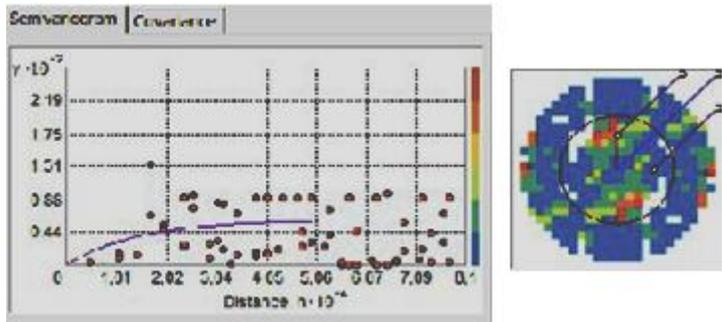


Figura 5. O variograma exponencial resultante apresenta uma distância angular de 45°. O efeito de pepita (nugget-effect) é nulo com um range máximo de 4373,2 metros.

### 5. ESTIMAÇÃO

A Krigagem ordinária ou Ordinary Kriging (OK) é uma técnica de regressão linear generalizadas para minimizar uma variância de estimação a partir de um modelo de covariância definido à priori (Deutsch and Journel, 1998). O processo de OK estima o valor de um atributo a partir de um conjunto de amostras vizinhas  $z(u_a)$ ,  $a=1, \dots, n$ . Genericamente, OK é semelhante à interpolação por média móvel ponderada. Porém, os pesos dados a cada observação da Krigagem são determinados a partir do variograma experimental. A equação 2 apresenta a formulação para o estimador de OK, onde  $w_j$  são os pesos a serem atribuídos para a observação  $i$ ,  $\Psi$  representa o valor dos multiplicadores de Lagrange ou a folga necessária para forçar o peso total ao valor um, aquando do sistema de derivadas parciais,  $\gamma(x_i, x_j)$  é o valor do variograma da amostra entre  $i$  e  $j$  enquanto  $\gamma(x_i, x_0)$  é igual ao valor do variograma entre a amostra  $i$  e a estimativa  $x_0$ .

$$\sum_{j=1}^n w_j \gamma(x_i, x_j) - \Psi = \gamma(x_i, x_0), \quad i = 1 \dots n \quad \text{e} \quad \sum_{j=1}^n w_j = 1 \quad (2)$$

Assim, o objectivo é encontrar o estimador de média ponderada que fornece um estimador imparcial e com o menor erro de estimação de variância usando o multiplicador de Lagrange e a soma unitária dos pesos como restrição. Partindo de uma média constante desconhecida e uma variável intrínseca (a ideia subjacente é trabalhar com incrementos mais do que com a própria variável), o regime geral de configurar as equações de Matheron é revelada na Figura 6. Uma vez calculado os pesos  $W_i$ , basta

multiplicar o valor de cada peso pelo respectivo valor da amostra onde a sua soma corresponde à estimativa desejada.

<b>A</b>	=	$Cov(x_1, x_1)$	...	$Cov(x_1, x_n)$	1	<b>B</b>	=	$Cov(x_1, x_0)$	<b>W</b>	=	$w_1$
		...	...	...	...			...			...
		$Cov(x_n, x_1)$	...	$Cov(x_n, x_n)$	1			$Cov(x_n, x_0)$			$w_n$
		1	...	1	0			1			$\psi$

Figura 6. O sistema genérico de Krigagem ( $A \times W = B$ ) para o cálculo dos pesos  $W_i$  (com  $i=1 \dots n$ ) através da covariância espacial existente entre as  $n$  amostras existentes,  $Cov(x_1, x_n)$ , com o ponto a estimar  $x_0$ .

A processo de validação-cruzada permite comparar os valores reais das amostras e valores estimados do OK de modo a quantificar a precisão (ou não) do modelo espacial de estimação. Sumariamente, a validação cruzada inicia-se por retirar do conjunto total das 45 amostras existentes uma amostra e, seguidamente, é feita a estimação pelo OK nesse ponto mas apenas baseado nas restantes 44 amostras. Feita a comparação dessa estimação com o valor real, obtem-se o erro amostral. Este processo é realizado continuamente para todas as amostras existentes. Teóricamente, um modelo é considerado bom se a média estimada for próxima de zero, o  $R^2$  igualar 1 enquanto o root-mean-square/average standart error devem apresentar valores baixos (ver Figura 7). O root-mean-square standardized deve ser o mais próximo de um.

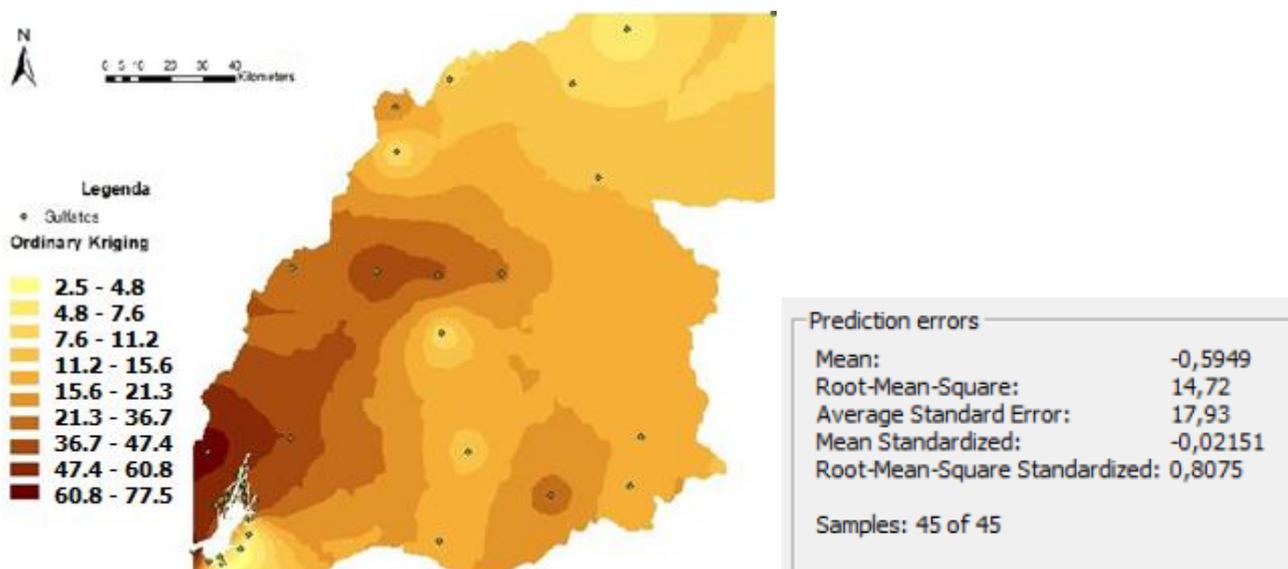


Figura 7. Estimação final do teor de sulfato pela Krigagem Ordinária na área de estudo (esquerda) e erros de estimação pela validação-cruzada (direita).

O Indicator Kriging (IK) ou Krigagem de Indicatriz é um estimador não paramétrico que usa a posição e os valores transformados dos dados para produzir uma distribuição de probabilidades, isto é, a incerteza do valor da variável pode ser caracterizada pela sua função de distribuição de probabilidade calculada em qualquer local desconhecido, condicionada à informação disponível. Segundo Journal

(1987), a aplicação desta metodologia não paramétrica não está restrita a qualquer tipo de distribuição estatística da variável, como é o caso da distribuição normal na abordagem paramétrica, justificando-se plenamente em problemas de poluição, onde o normal é a ocorrências de valores anormais.

De um ponto de vista prático, os dados são codificados como maiores ou menores a valores limite (thresholds) especificados. Assim, é criada uma nova variável de indicatriz que toma unicamente os valores 0 ou 1, caso o valor da amostra seja menor e maior, respectivamente, que o limite pré-especificado. Confirmado por Ribeiro, Santos e Silva (2011), numa codificação binária, os resultados da interpolação são valores necessariamente entre 0 e 1 e representam as probabilidades da variável ser menor ou maior ao valor limite seleccionado.

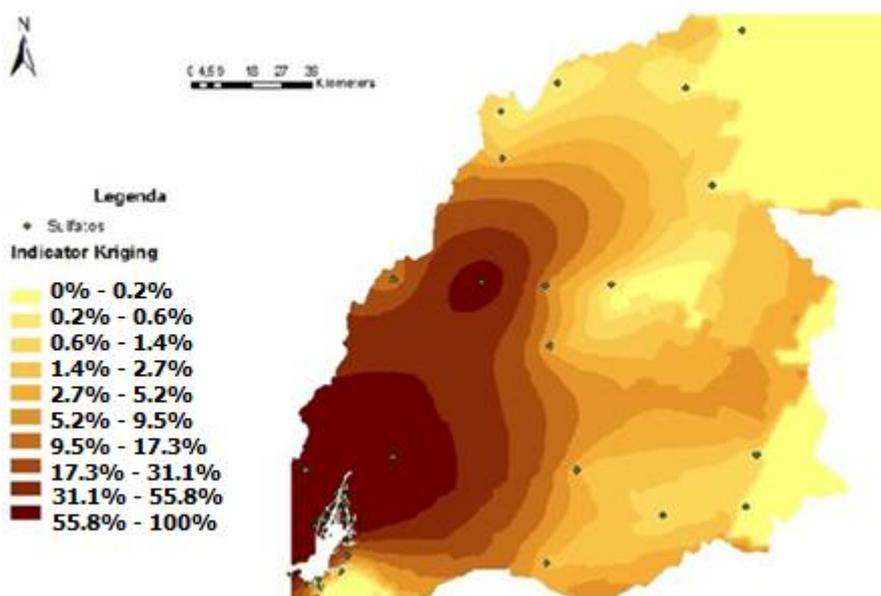


Figura 8. Mapa de probabilidades aquando da concentração de cloretos exceda o valor crítico de 33 mg/l (threshold considerado perigoso para o ser humano).

## 6. OBSERVAÇÕES FINAIS

Burrough e McDonnel (1994) identificam algumas questões que se devem ter em conta quando se pretende obter mapas interpolados, nomeadamente o número de pontos da amostra, a distribuição dos mesmos, o tamanho da área de estudo, topologia da vizinhança das amostras e o método matemático a ser usado. Neste caso particular, denota-se que o esquema espacial de amostragem foi equilibrado para o estudo apresentado. Segundo Griffith (1996), o tamanho da amostra deve ser igual ou superior a 30, situação perto de ser verificada neste caso. Obviamente, o tempo e os custos associados ditam o tamanho de qualquer amostragem.

Ordinary Kriging e Indicator Kriging são métodos de Krigagem que usam a dependência espacial entre amostras vizinhas, expressa no semivariograma, para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima (estimadores considerados óptimos). Contudo, a

Krigagem ordinária dá-nos mapas de interpolação ou estimação enquanto a Krigagem de indicatriz dá-nos mapas de incerteza (Silva and Monteiro, 2010). Como seria de esperar, a aplicação destes métodos geoestatísticos para um conhecimento mais profundo dos valores da concentração dos sulfatos apresenta-se altamente eficaz. As zonas correspondentes a uma maior concentração de sulfatos correspondem à região da Grande Lisboa, possivelmente com origem em actividades agrícolas (utilização de adubos, por exemplo), podendo assim esta informação disputar como um factor crítico, por exemplo, numa decisão da localização de uma futura instalação de equipamento de limpeza de águas residuais.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

Burrough, P., McDonnel, R. (1998). Principles of geographical information systems (spatial informations systems and geostatistics). Oxford University Press. 2<sup>nd</sup> Edition, 90-6164-200-0, 832 pp.

Deutsch, C., Journel, A. (1998). GSLIB. Geostatistical software library and user's guide. Oxford, New York: Oxford University Press, ISBN 978-019-510-0150, 369 pp.

ESRI (2001). ArcGIS 9 - ArcGIS Geostatistical Analyst Tutorial, USA, 39 pp. ([http://webhelp.esri.com/arcgisdesktop/9.3/pdf/Geostatistical\\_Analyst\\_Tutorial.pdf](http://webhelp.esri.com/arcgisdesktop/9.3/pdf/Geostatistical_Analyst_Tutorial.pdf)).

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press, ISBN 0-19-511538-4, 483 pp.

Griffith. D. (1996). Some guidelines for specifying the geographical weights matrix contained in spatial statistical models. In Practical Handbook of Spatial Statistics, CRC Press, 0-8493-0132-7, Boca Raton, Florida, 336 pp.

Instituto da Água (2000). Plano de Bacia Hidrográfica do Rio Tejo, Ministério do Ambiente e do Ordenamento do Território, 42 pp. ([http://www.inag.pt/inag2004/port/a\\_intervencao/planeamento/pbh/pbh03\\_tejo/4/09000md\\_4f.pdf](http://www.inag.pt/inag2004/port/a_intervencao/planeamento/pbh/pbh03_tejo/4/09000md_4f.pdf)).

Isaaks, E. H. and R. M. Srivastava. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, USA, ISBN 3-7908-1418-0, 442 pp.

Ribeiro, A. (2006). Diagnóstico de uma área compactada por actividade mineraria, na floresta amazônica, empregando metodos geoestatisticos à variavel resistência mecânica à penetração do solo. Acta Amazônica (<http://www.scielo.br/pdf/aa/v36n1/v36n1a10.pdf>).

Ribeiro, L., Santos, A., Silva, J. (2011). Caracterização Geoestatística da Contaminação da Contaminação Salina por Dissolução na Zona da Maragota (<http://www.aprh.pt/congressoagua98/files/com/069.pdf>).

Silva, J., Monteiro, P. (2010). Modelação Espacial da Precipitação na Ilha de Santiago de Cabo Verde. Praia.