

ANÁLISE ESPACIAL DA DISTRIBUIÇÃO DA ESPÉCIE *SALIX SALVIIFOLIA* EM PORTUGAL CONTINENTAL

Sandra Borges (nº41263)

(patriciaborges2000@yahoo.com)

RESUMO

No presente relatório visa apresentar os resultados da exploração de diferentes técnicas de análise espacial com base nos conhecimentos adquiridos no âmbito da cadeira de Análise Espacial de Informação Geográfica. Para o efeito será estudada a espécie arbórea ripícola, o *Salix salviifolia* (salgueiro branco), onde serão aplicadas diferentes técnicas de análise espacial em três domínios diferentes: a distribuição espacial da espécie no território de Portugal Continental (padrão pontual), a análise da autocorrelação e regressão espacial e distribuição espacial do *Salix salviifolia* com base nas variáveis ambientais e climáticas que melhor poderão explicar as suas exigências ecológicas, e por fim, definir áreas de repositório de material vegetal nas galerias ripícolas tendo em conta os pontos de ocorrência da espécie que foram inventariadas no âmbito da Directiva Quadro da Água (INAG), a rede hidrográfica para todo o território de Portugal Continental e a principais bacias hidrográficas. O objectivo final é a sinalização geográfica destes locais de repositório de material vegetal, para posterior utilização em obras de requalificação ambiental. Com o desenvolvimento deste relatório pretende-se apresentar as diferentes etapas da metodologia desenvolvida no âmbito da análise espacial para este caso de estudo e em simultâneo analisar os resultados obtidos de acordo com os métodos utilizados.

1. Introdução

Neste trabalho, pretende-se compreender a distribuição da espécie *Salix salviifolia* no território de Portugal Continental, tendo como base os inventários realizados no âmbito da implementação da Directiva Quadro da Água (DQA) entre 2004 e 2006, e as exigências ecológicas da própria espécie.

Neste sentido, este processo de análise espacial divide-se por etapas que respeitam uma sequência lógica de procedimentos para este caso de estudo.

Em primeiro lugar foi feita uma análise da distribuição espacial do padrão pontual dos pontos de ocorrência da espécie *Salix salviifolia*, envolvendo diversas técnicas de visualização, técnicas de análise exploratória (1ª e 2ª ordem) e testes de hipótese de aleatoriedade espacial.

Depois foi feita uma análise, de modo a compreender a autocorrelação da ocorrência da espécie segundo o inventário da DQA entre 2004 e 2006 no território de Portugal Continental, envolvendo diversos procedimentos como técnicas de visualização de dados de área, indicadores de associação espacial (globais e locais) e aplicação de testes de significância estatística aos mesmos.

Ainda através do estudo da regressão espacial, tentou-se definir um modelo de regressão explicativo para a ocorrência da espécie no território de Portugal Continental.

E por último, fez-se uma análise de modo a compreender a distribuição espacial da espécie *Salix salviifolia* em função das suas exigências ecológicas e respectivas variáveis climáticas e ambientais que justificam a sua distribuição na rede hidrográfica do território português.

Na sequência das várias análises realizadas a este caso de estudo, pretendeu-se obter como resultado prático final, as áreas com melhor aptidão para servirem de repositório de material vegetal a ser utilizado em futuras obras de requalificação ambiental.

No desenvolvimento deste relatório procurou-se na metodologia aplicada, analisar e explicar os resultados de uma forma sucinta.

2. Análise da Distribuição Espacial da Espécie *Salix salviifolia* em Portugal Continental

O objectivo da análise de padrões de pontos espaciais é verificar existência uma tendência nos eventos observados exibindo um padrão sistemático, como regularidade, agrupamentos ou aleatoriedade. Por outras palavras, o objectivo da análise dos processos pontuais é estudar a distribuição espacial desta espécie no território de Portugal Continental com base no inventário DQA realizado entre 2004 e 2006 e testar hipóteses sobre os padrões observados.

2.1 Tratamento de dados

Os dados foram disponibilizados na página do Instituto da Água, I.P. (INAG) e referem-se aos dados recolhidos nas campanhas de amostragem que decorreram entre 2004 e 2006 no âmbito da implementação da Directiva Quadro da Água em Portugal Continental.

A informação encontra-se compilada num ficheiro WinZip® que contém 5 ficheiros Microsoft Office Excel® correspondentes às características gerais das estações de amostragem e aos dados recolhidos para cada elemento de qualidade biológica (**macrófitos e matriz com dados ambientais dos pontos de amostragem da rede hidrográfica nacional, 404 pontos**).

Nota: Estes dados poderão ser utilizados livremente, sendo, contudo, necessário referir a origem dos dados.

As coordenadas (X, Y) dos pontos de amostragem DQA vêm expressas no sistema de referência *Hayford Gauss* Militar - Datum Lisboa (EPSG: 20790).

Uma vez que é muito importante saber quais as variáveis que influenciam a ocorrência desta espécie no território de Portugal Continental e tendo como base teórica a definição de nicho ecológico também se utilizaram ficheiros vectoriais com variáveis ambientais e climáticas da fonte Atlas do Ambiente (*Hayford Gauss* Militar) e ficheiros raster da fonte Monteiro-Henriques (2010) com variáveis climáticas (ETRS89).

Para a construção de uma matriz com toda a informação necessária para a realização das análises espaciais que se pretende efectuar neste trabalho, importou-se a matriz de base dos inventários de campo para 404 pontos e no *ArcGis Versão 9.3* [ESRI Inc., 1998-2009], obteve-se toda a informação climática e ambiental para cada ponto de amostragem. Esta extração de valores foi feita através da função *Extract Values to Points* da ferramenta *spatial analyst*, em ambiente *ArcGis 9.3*, de seguida efectuou-se um *join* desta tabela com uma *shape* das principais bacias hidrográficas no território de Portugal Continental e por fim, procedeu-se à sua gravação como layer. A partir desta layer procedeu-se às diversas análises espaciais da espécie *Salix salviifolia* para todo o território de Portugal Continental.

2.2 Visualização do padrão pontual

Neste ponto fez-se a identificação e localização da frequência relativa da espécie *Salix salviifolia* no território de Portugal continental, informação resultante dos inventários florísticos em 404 pontos de amostragem da rede hidrográfica no âmbito da implementação da Directiva Quadro da Água.

Nesta primeira etapa o objectivo foi perceber a disposição espacial da espécie em conformidade com as suas exigências ecológicas e perceber se esta distribuição apresentava um padrão sistemático e quais os factores/variáveis ambientais e climáticas que poderiam determinar esse padrão espacial.

Para melhor entender a distribuição espacial dos dados, mediu-se a direcção da distribuição espacial, o centro médio dos incêndios e a elipse e círculo padrão que delimitam o espaço no qual estão concentrados 68% das observações.

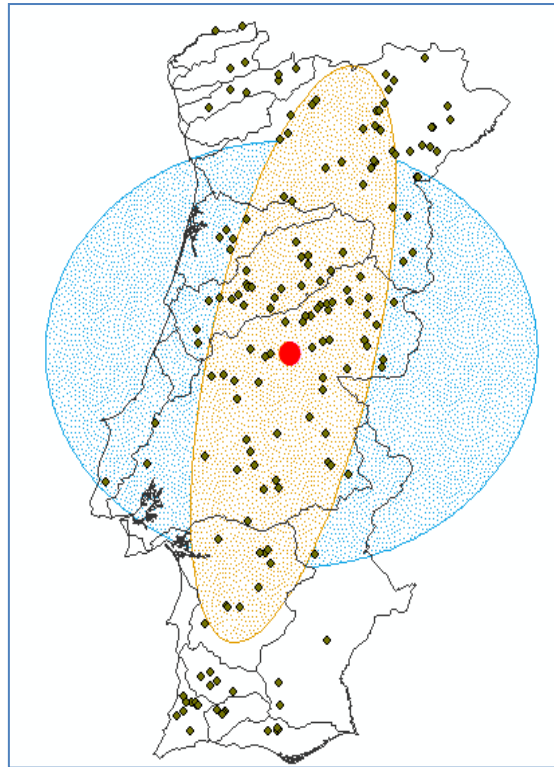


Figura 1 – Centro médio, elipse e círculo padrão.

Com este mapa, foi possível ter uma percepção preliminar da forma como a espécie *Salix salviifolia* se encontra distribuída no território de Portugal Continental, através da sua frequência, densidade, centro geométrico e orientação espacial.

2.3 Análise Exploratória de Dados

Nesta etapa efectuou-se uma análise mais rigorosa dos eventuais padrões espaciais, de forma a verificar associações espaciais entre as variáveis que caracterizam os fenómenos estudados. Os métodos desenvolvidos permitiram explorar as propriedades de 1ª e 2ª ordem do padrão pontual.

2.3.1 Propriedades de 1ª Ordem

Para explorar as **propriedades de 1ª ordem** e efectuar o reconhecimento e medição do padrão de uma distribuição de pontos utilizou-se o método da contagem por quadrantes e o estimador de densidade Kernel.

O **método da contagem por quadrantes** consiste numa grelha estruturada na área de estudo procedendo-se posteriormente à simples contagem de eventos em cada um dos quadrantes. Esta técnica é mais simples de aplicar embora tenha mais limitações pois está dependente da dimensão da célula escolhida. Para o estudo da distribuição da espécie *Salix salviifolia*, consideraram-se duas grelhas distintas, uma grelha 20x20 (malha pequena) e uma grelha 10x10 (malha grande), de modo a comparar-se resultados.

Neste procedimento é calculada a média de eventos por quadrante, a variância associada e ainda a estatística *Variance To Mean Ratio* (VTMR), que é a medida da divisão da variância pela média, sendo assim uma medida normalizada em torno de 1.

Se o valor de **VTMR for inferior a 1**, o padrão é regularmente disperso ou seja, os eventos estão distribuídos quase regularmente no espaço não contendo um padrão de aglomerados. No caso de **VTMR ser superior a 1**, o padrão é de aglomerados ou clusters ou seja, existência de conjuntos de dados agrupados e áreas extensas sem pontos. Para o caso da **VTMR ser igual a 1**,

o padrão apresenta uma distribuição aleatória, não tendendo nem para aglomerado nem para a dispersão.

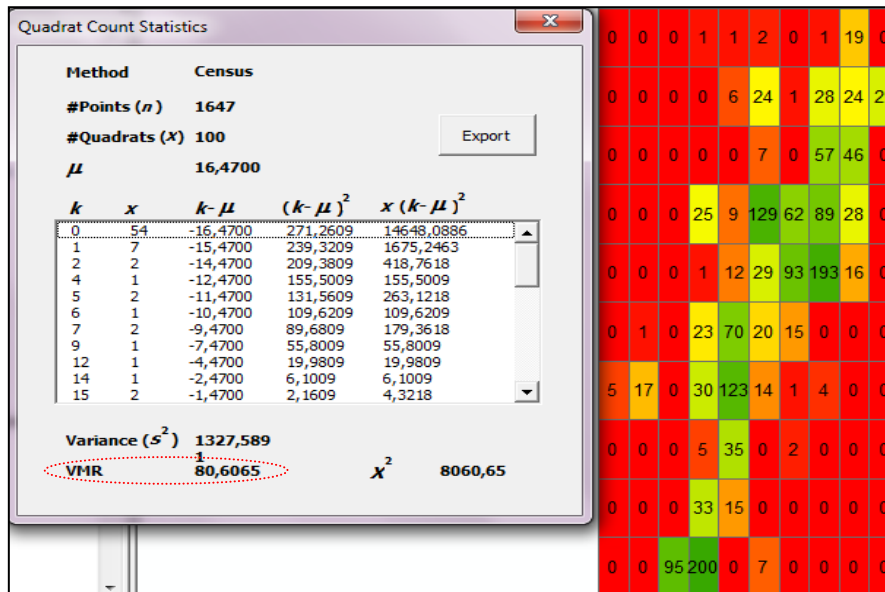


Figura 2 – Estatística por método dos quadrantes e Grelha 10x10

Neste caso, o **VTRM é superior a 1**, a grelha testada, revela que apresenta um padrão de aglomerados ou clusters. Da análise da figura 2 (grelha 10x10), pode-se concluir que esses clusters estão mais concentrados nas bacias hidrográficas do Douro, Tejo e várias Algarve

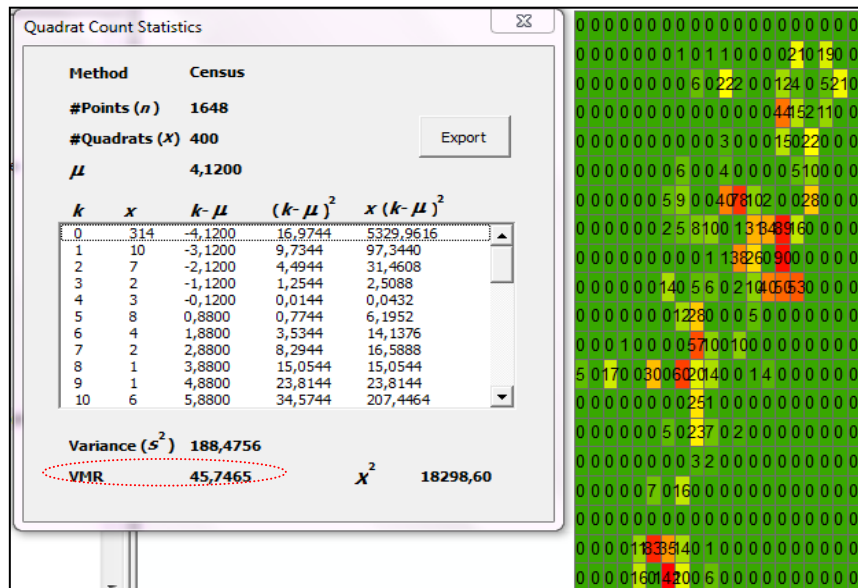


Figura 3 – Estatística por método dos quadrantes e Grelha 20x20

Para este caso, o **VTRM é superior a 1**, sendo que esta grelha testada também apresenta um padrão de aglomerados ou clusters. Da análise da figura 3 (grelha 20x20), com um aumento do numero de quadrantes, também se conclui que os cluster se concentram nas bacias hidrográficas do Douro, Tejo e várias Algarve.

O **estimador de densidade Kernel**, consiste em converter o conjunto dos eventos pontuais numa superfície contínua que represente a densidade. Esta função realiza uma contagem de

todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse. Neste caso, a visualização do padrão de pontos permite a detecção de *hot-spots*, ou seja, regiões de influência, verificando se o processo é ou não estacionário a partir das variações locais de intensidade. No *Kernel* a densidade é determinada em todos os locais da território de estudo e não somente nos locais onde existem eventos.

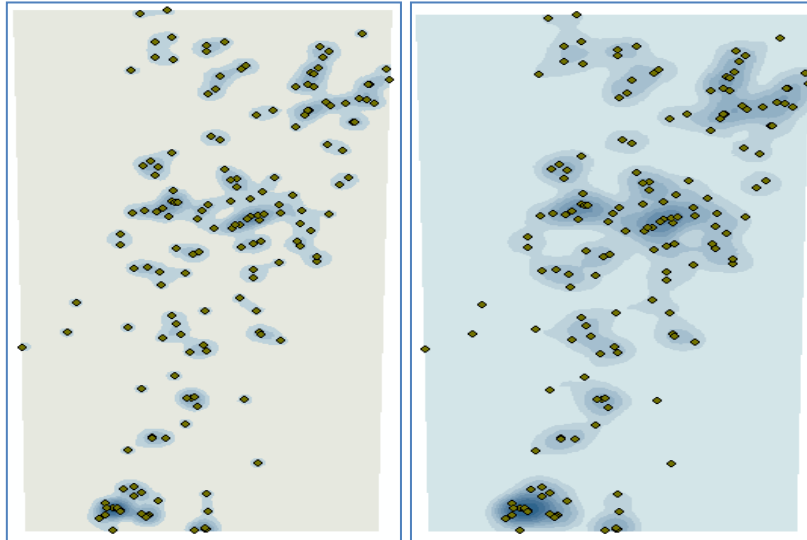


Figura 4 –Mapas de densidade: Kernel 10000m e Kernel 20000m

Os resultados da densidade de *Kernel* estão em conformidade com o método anterior, ou seja, verifica-se que a zona de *hot-spots* se encontra nas bacias hidrográficas do Douro, Tejo e Várias Algarve.

2.3.2 Propriedades de 2ª Ordem

Para explorar as **propriedades de 2ª ordem** utilizaram-se os métodos de distância ao vizinho mais próximo e a função K, que se tratam de métodos baseados na medição de distâncias entre pontos. Estes métodos assumem estacionaridade sobre pequenas escalas na região de trabalho, significando que a média e a variância do processo são independentes da localização e constantes na região.

O **método do vizinho mais próximo** é o método mais directo para descrever as propriedades de 2ª ordem de um processo pontual, estimando a função de distribuição cumulativa a partir de distâncias entre os eventos da área em análise. Utilizando a ferramenta *Average Nearest Neighbor* obteve-se o seguinte resultado:

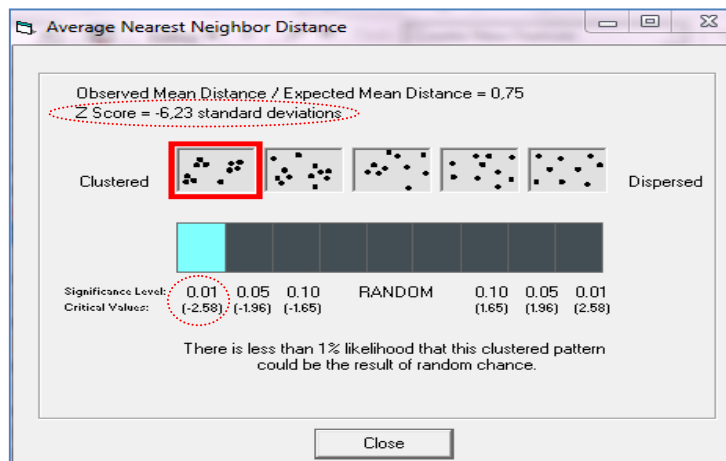


Figura 5 – Vizinho mais próximo

O resultado desta função de distribuição indica que existem aglomerados/agrupamentos na distribuição da espécie *Salix salviifolia*. Embora este método do vizinho mais próximo forneça uma boa indicação da distribuição espacial, ele considera apenas escalas pequenas.

A função K é um método para estimar o padrão de eventos na área de estudo e baseia-se na distância entre todos os pontos e não apenas na distância mínima entre dois eventos. Esta função permite **avaliar o grau de dispersão dos eventos pontuais**, através da dependência espacial de uma variável relativamente a todas as variáveis da área de estudo.

Considerando diversas distâncias (0.5 e 1 grau decimal), incrementos de 10 e zero permutações, obtiveram-se os seguintes gráficos:

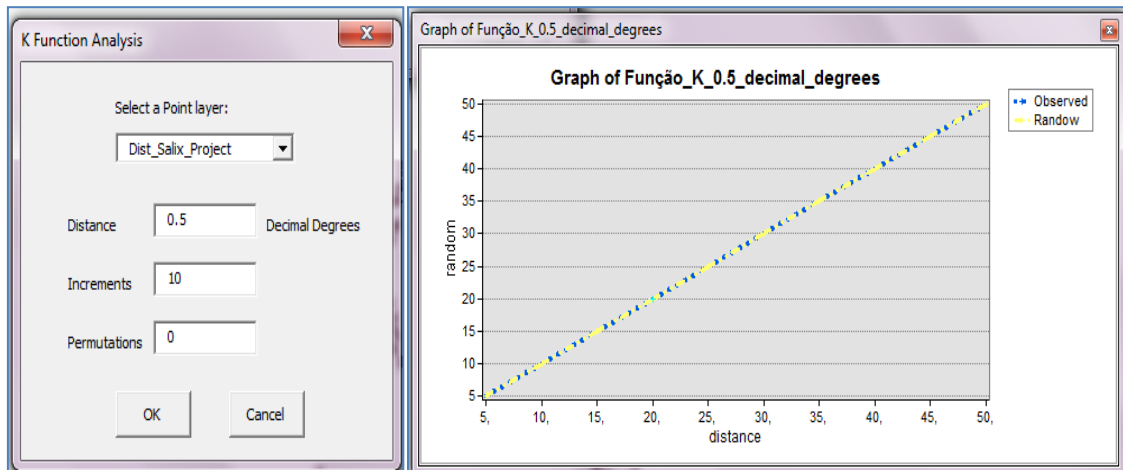


Figura 6 – Aplicação da Função K (0.5 graus decimais) e gráfico resultante

Da análise do gráfico (Figura 6), verifica-se que os valores observados são coincidentes com os valores esperados e que aumentam gradualmente com o aumento de distância. Isto leva a concluir que o padrão pontual correspondente à distribuição de *Salix salviifolia* é um padrão essencialmente de aglomerados.

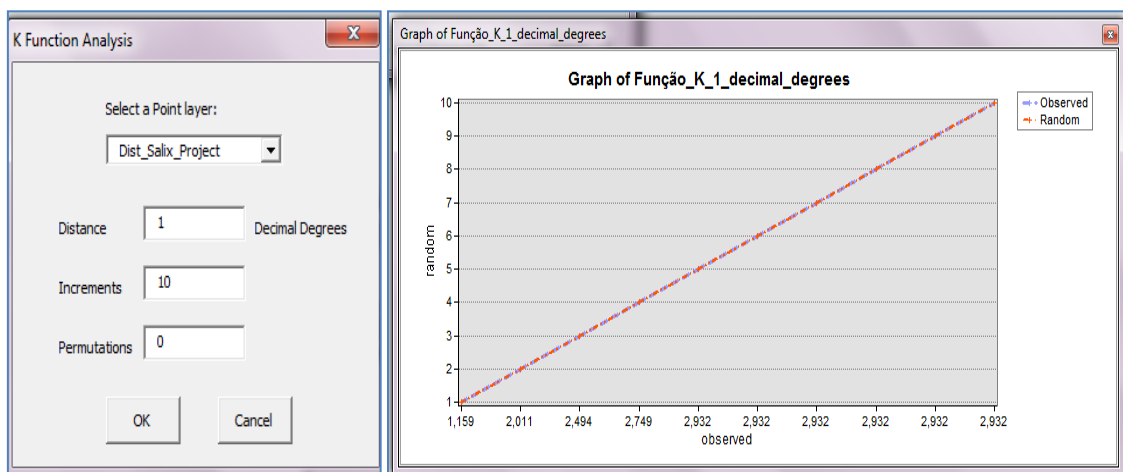


Figura 7 – Aplicação da Função K (1 grau decimal) e gráfico resultante

Da análise do gráfico (Figura 7), verifica-se que os valores observados são coincidentes com os valores esperados e que aumentam gradualmente com o aumento de distância. Isto leva a concluir que o padrão pontual correspondente à distribuição de *Salix salviifolia* é também um padrão essencialmente de aglomerados.

De seguida procedeu-se ao **cálculo da Função K para um subconjunto** dos pontos da área de estudo e a determinação de um intervalo de confiança para os resultados.

No passo anterior, calculou-se a função K para a totalidade dos layers de pontos, mas não se calculou um intervalo de confiança, para o efeito criou-se uma nova layer a partir de um subconjunto de dados, onde se procede à análise da função K neste subconjunto e recorrendo a simulações para gerar um intervalo de confiança.

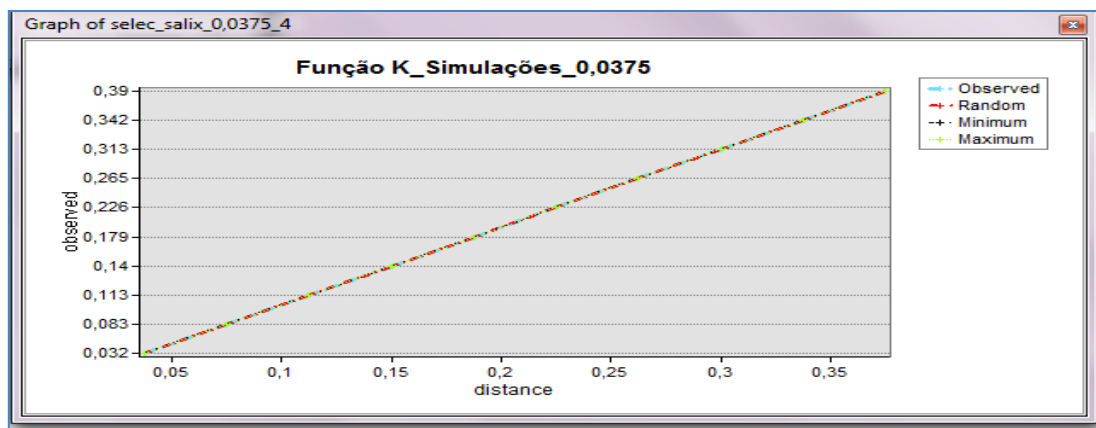
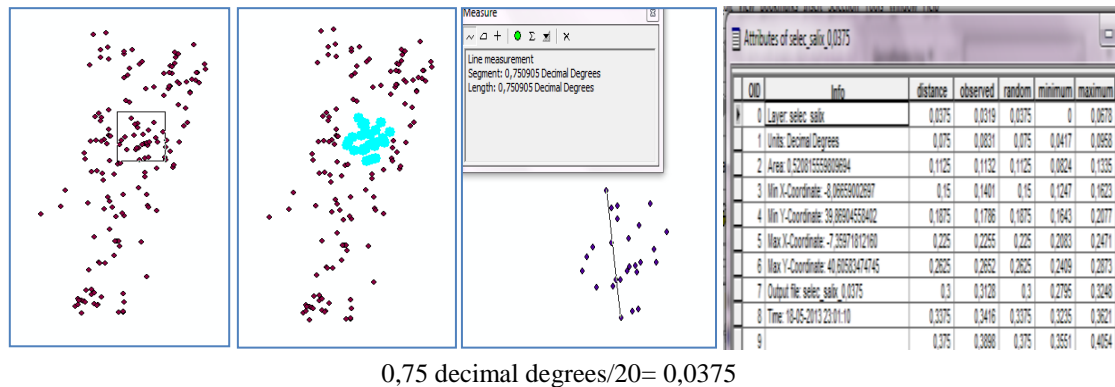


Figura 8 – Aplicação da **Função K** com simulações

Relativamente à **função K** com envelopes de simulação, da análise da **figura 8**, conclui-se que os valores observados coincidem com os envelopes, então o padrão observado é de aglomerados.

2.4 Testes para hipótese de aleatoriedade espacial

Para **testar a hipótese de CSR (Complete Spatial Randomness) do conjunto de dados em análise**, recorreu-se à comparação com distribuições teóricas de referência e ao **teste de significância estatístico** aplicado através do método do vizinho mais próximo assim como à simulação do processo aleatório recorrendo à função K para obter o **gráfico com envelopes de simulação**.

Relativamente ao **método do vizinho mais próximo**, este testa a hipótese **nula de CSR**, ou seja, de que o padrão é aleatório (H0) ou disperso/aglomerado (H1). Assim, fazendo-se uma análise dos resultados obtidos (figura 5), verifica-se que para $\alpha = 0,01$, ou seja, 99% de confiança, o valor $Z_{critico} = -2,58$ e o valor $Z_{teste} = -6,23$. Considerando que estamos a usar uma distribuição normal $N(0,1)$, como $Z_{teste} < Z_{critico}$, rejeita-se H0 e aceita-se H1, logo não se aceita a hipótese nula e neste sentido as variáveis apresentam um padrão disperso e/ou aglomerado

Relativamente à **função K com envelopes** de simulação, da análise da **figura 8**, conclui-se que dado os valores observados se encontrarem dentro dos envelopes, o padrão observado é

aglomerado. Pelo facto dos valores observados serem coincidentes com o máximo e mínimo do envelope, conclui-se que se trata de um padrão de aglomerado.

3. Análise da Autocorrelação Espacial e Regressão Espacial

O objectivo na **análise de áreas** é **descrever e visualizar as distribuições espaciais**, identificar **situações atípicas** (*ouliers*), descobrir padrões de associação espacial, *clusters* e sugerir regimes espaciais ou formas de heterogeneidade espacial. Esta análise será efectuada face aos dados amostrados da distribuição da espécie *Salix salviifolia*, num período de tempo entre 2004 e 2006 em todo o território de Portugal Continental.

3.1 Tratamento de dados

Em termos de dados adicionais para efeitos de regressão (variáveis independentes), utilizaram-se as variáveis ambientais e climáticas que já tinham sido extraídas para cada ponto de amostragem e juntou-se a *shape* das principais bacias hidrográficas do território de Portugal Continental, onde consta a informação da área individual de cada bacia.

No *ArcGis Versão 9.3* [ESRI Inc., 1998-2009] fez-se um *join* da *layer* bacias hidrográficas com a *layer* dos 404 pontos amostrados da DQA e respectivas variáveis ambientais e climáticas previamente extraídas através da função *Extract Values to Points* da ferramenta *spatial analyst*. Tendo já toda a informação dos pontos de amostragem com correspondência a cada umas das principais bacias hidrográficas do território de Portugal Continental, de seguida exportou-se toda a informação para formato *shapfile*. Esse formato foi posteriormente usado no *software OpenGeoDa Versão 1.4.1* [GeoDa Center e filiados (R-Geo)].

3.2 Visualização dos dados espaciais

Nesta fase, e organizando-se as variáveis de estudo em unidades de área (distribuição da espécie *Salix salviifolia* por bacia hidrográfica), e usando o *software* referido obtiveram-se 2 mapas: por **intervalos iguais** (5vclasses) e por *percentis*. Com estes mapas, foi possível ter uma melhor percepção da forma como a espécie se encontra distribuída por bacia hidrográfica principal. Numa primeira análise, podemos verificar que a maior concentração da espécie *Salix salviifolia* nas bacias hidrográficas principais do Centro e Sul do território de Portugal Continental.

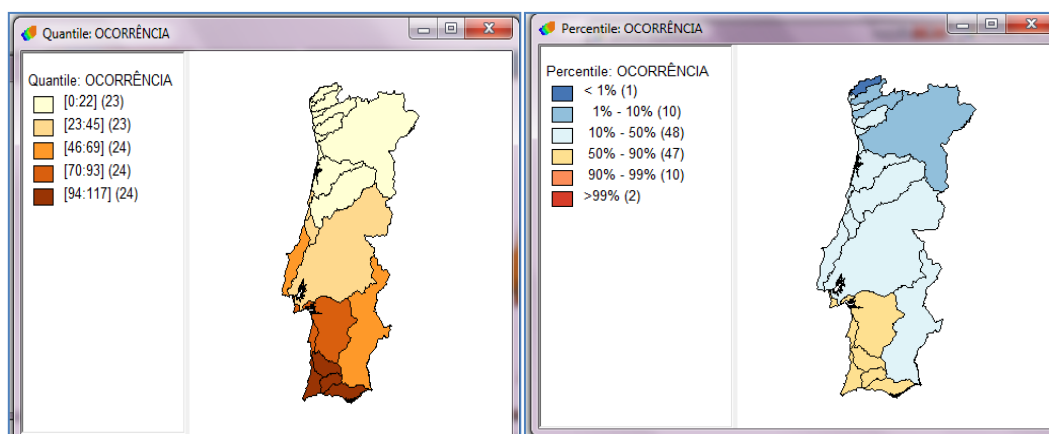


Figura 9 – Mapa por Intervalos Iguais e Percentagem

3.3 Análise Exploratória de Dados

Nesta etapa, pretende-se **medir a estrutura da correlação espacial e estimar a magnitude da autocorrelação espacial** entre bacias hidrográficas. Para tal, foi necessário determinar a **matriz de vizinhança**, que foi gerada através do *OpenGeoDa Versão 1.4.1* [GeoDa Center e filiados (R-Geo)], usando o método Distance Weight (*Distance metric* > *Euclidean Distance* > *Threshold Distance = 57701,575865*), depois de usar este método para determinar o ficheiro *gal.*, determinou-se o histograma de conectividade, apresentado na figura 10:

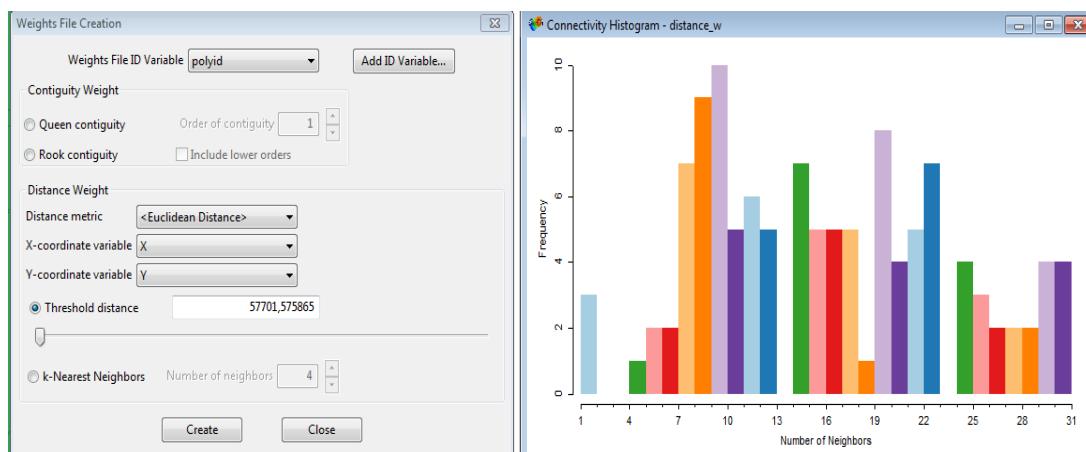


Figura 10 – Matriz de vizinhança

3.3.1 Índices Globais de Autocorrelação Espacial

Para as **estatísticas de autocorrelação espacial global**, consideraram-se todas as presenças da espécie *Salix salviifolia* em cada bacia hidrográfica, tendo por objectivo caracterizar a dependência espacial mostrando como os valores estão correlacionados no espaço e fornecendo uma medida geral da associação espacial.

O **Índice Global de Moran** mede a autocorrelação espacial, comparando o valor de uma variável com os valores de outras variáveis em quaisquer localizações. Este índice mede a autocorrelação espacial baseado simultaneamente nos valores das variáveis e nas suas localizações. O Índice de Moran considera a covariância entre pares e sendo um indicador global, determina o nível de interdependência espacial entre todos os polígonos (bacias hidrográficas) em estudo.

Valores próximos de **+1 indicam aglomerados** e a autocorrelação espacial é **positiva** o que indica que valores altos tendem a estar na vizinhança de valores altos e valores baixos tendem a estar na vizinhança de valores baixos. Se o coeficiente de correlação é **nulo**, significa **aleatoriedade** e se a autocorrelação espacial for **negativa**, o valor do índice indica que **valores altos** estão rodeados de **valores baixos** e **valores baixos** estão rodeados de **valores altos** (Ver figura 11).

O **gráfico de Moran** compara os valores normalizados do atributo numa área com a média dos valores normalizados dos seus vizinhos. Como se verifica a orientação do gráfico (**figura 11**) é no primeiro e segundo quadrante (Q1 e Q2) logo a autocorrelação é positiva, ou seja, localizações com vizinhos de valores semelhantes (clusters). Se a orientação fosse no terceiro e quarto quadrante (Q3 e Q4) então a autocorrelação seria negativa isto é localizações com vizinhos com valores distintos e padrão dispersivo.

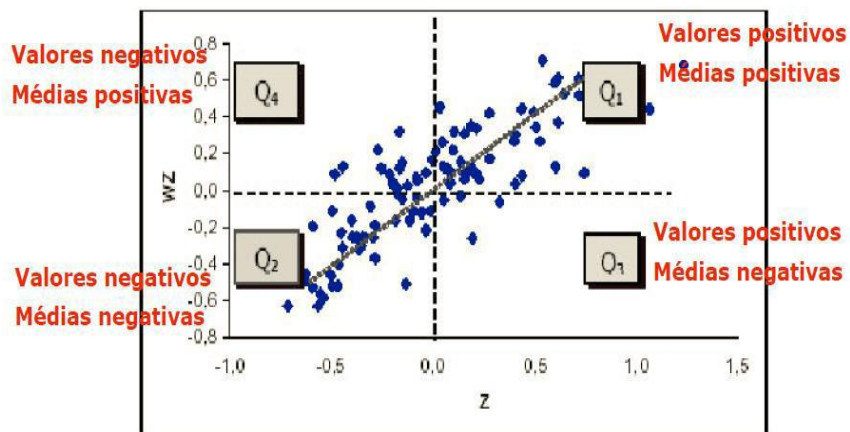


Figura 11 – Gráfico explicativo de Moran

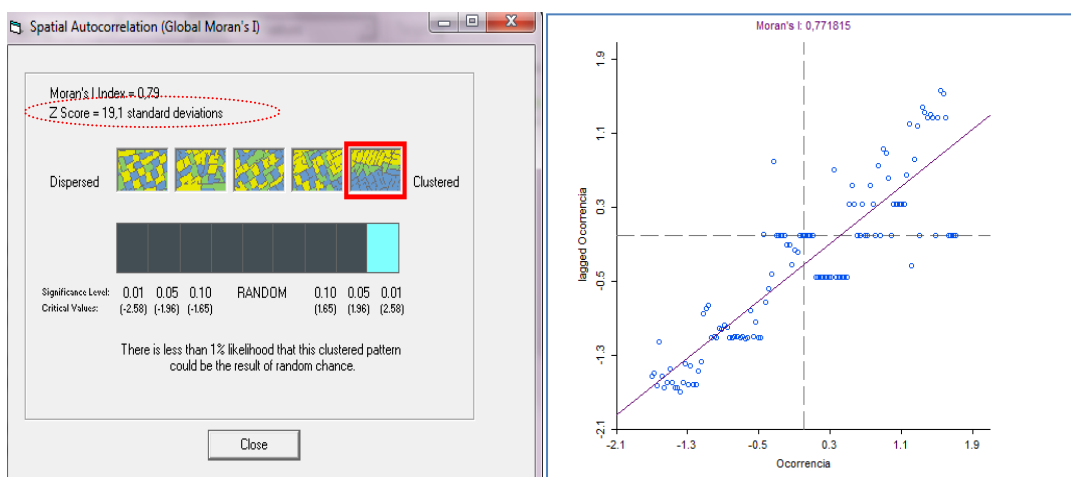


Figura 12 – Índice Global de Moran ArcGis e OpenGeoDa

Da análise dos resultados obtidos por dois softwares diferentes (*ArcGis* e *OpenGeoDa*), verifica-se que o índice no *ArcGis* é 0,78 e no *OpenGeoDa* é 0,771, verificando-se consistência nos resultados. Tratando-se como já foi referido, de uma **autocorrelação espacial positiva com aglomerados**.

3.3.2 Índices Locais de Autocorrelação Espacial

Para as estatísticas de autocorrelação espacial local, também são considerados todas as presenças da espécie *Salix salviifolia* em cada bacia hidrográfica, e avaliam a covariância entre um determinado polígono e uma certa vizinhança definida em função de uma distância. O objectivo é produzir um valor específico para cada área, permitindo a identificação de aglomerados (*clusters*), de valores extremos (*outliers*) e a existência de vários regimes espaciais, que não são identificados pelos indicadores globais.

O **Índice Local de Moran** consiste num produto do resíduo no polígono de referência com a média local dos resíduos dos seus vizinhos adjacentes, em que os valores altos e positivos apontam para a presença de **cluster** e valores baixos indicam um regime espacial de **desigualdade ou zonas de transição**.

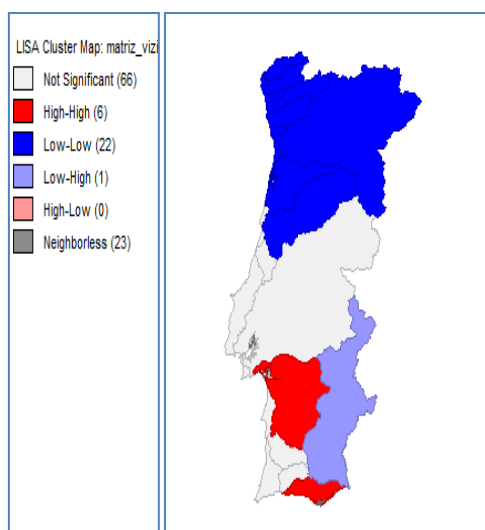


Figura 13 – LISA cluster Map

O *LISA cluster map* indica quais os quadrantes do mapa de dispersão de Moran a que pertence cada bacia hidrográfica. Da análise do mapa, verifica-se que bacias hidrográficas a **vermelho e azul são os que contribuem significativamente para a autocorrelação positiva (Q1 e Q2)** pois são as bacias cujos vizinhos têm valores similares (altos próximos de altos e baixo próximos de baixos, respectivamente de acordo com as cores representadas no mapa da figura 13).

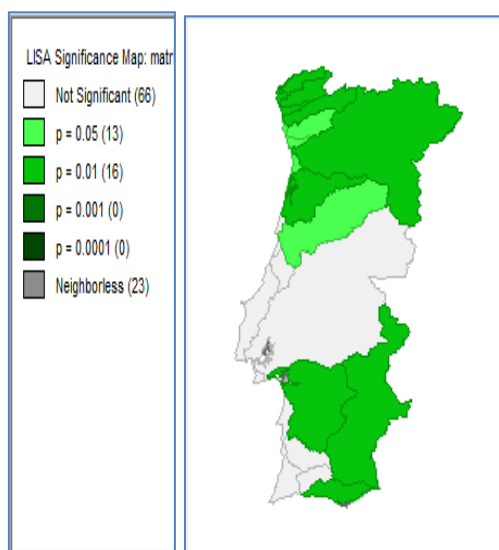


Figura 14 – LISA significance Map

O *LISA significance map* mostra o nível de significância estatística com que cada bacia hidrográfica pode ser classificada, no sentido de contribuir para o **valor global de autocorrelação**. O método usado para as simulações é o **método de Monte Carlo**. Da análise do mapa, verifica-se que as bacias hidrográficas assinaladas com os tons de verde mais claro $p=0.05$ e $p=0.01$ respectivamente são as que **contribuem mais significativamente para a autocorrelação**, diminuindo essa contribuição para as cores de verde mais escuro e não tendo qualquer significado estatístico as bacias hidrográficas a cinzento claro.

3.3.3 Significância Estatística

Para verificar a **validade estatística do Índice Global de Moran** estimou-se a sua significância utilizando o teste da **pseudo-significância** [CÂMARA et al., 2004].

No **teste da pseudo-significância**, também conhecido por técnica de permutação é necessário escolher o número de permutações, em que cada uma delas produz um novo arranjo espacial dos valores do atributo, onde os valores estão redistribuídos entre as áreas.

Neste processo apenas um dos arranjos corresponde à situação observada e o índice medido nos dados originais compara-se com a distribuição simulada. Se o índice corresponder a um extremo da distribuição simulada, indica que é um valor com significância estatística, e deve-se rejeitar a hipótese nula isto é rejeitar a aleatoriedade, tal como se verifica neste caso de estudo tanto para 99 permutações assim como para 999 (Figura 15).

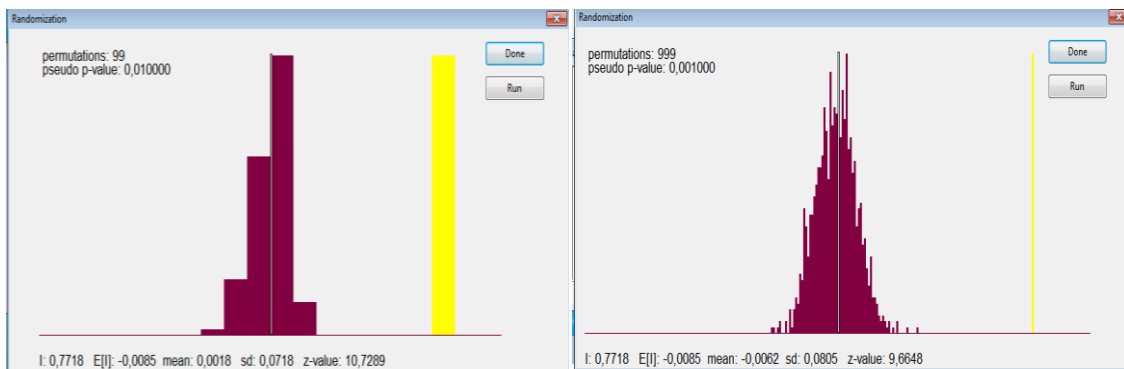


Figura 15 – Teste pseudo-significância

Para verificar a **validade estatística dos indicadores locais**, utilizou-se **valor de Z** presente na **estatística Getis e Ord**, gerada através do *ArcGis Versão 9.3* [ESRI Inc., 1998-2009]. É um indicador baseado na soma de valores vizinhos definidos a partir de uma matriz de vizinhança, valores altos apontam para altos índices de ocorrência do atributo e valores baixos indicam um agrupamento de valores baixos.

A **estatística Getis and Ord G_i^*** é um indicador LISA, ou seja, é um indicador que permite detectar **hot-spots locais**. A estatística G supõe normalidade nas suas distribuições para construir intervalos de confiança para a hipótese nula de total ausência de autocorrelação local.

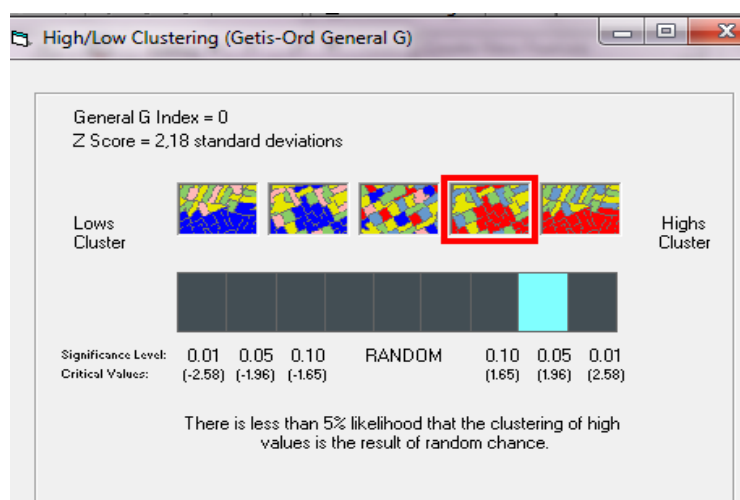


Figura 16 – Estatística G determinada no ArcGis

Assim, fazendo-se uma análise dos resultados obtidos na figura 16, verifica-se que para $\alpha = 0.05$, ou seja, **95% de confiança**, o **valor Z crítico = 1,96** e o **valor Z teste = 2,18**. Considerando que estamos a usar uma distribuição normal $N(0,1)$, como **Z teste > Z crítico** rejeita-se H_0 e

aceita-se H1, logo não se aceita a hipótese nula e neste sentido as variáveis apresentam um **padrão de aglomerados**.

3.4 Regressão Espacial

Nesta etapa, pretende-se o uso de técnicas de regressão espacial de modo a explicar eventuais associações espaciais. Enquanto a **correlação** mede o **grau de relacionamento entre duas variáveis ou mais**, a **regressão** fornece uma **equação que descreve o relacionamento matemático entre elas**.

Para o efeito, foram adicionados dados para variáveis (independentes) espaciais que poderão contribuir para a ocorrência da variável dependente (ocorrência da espécie *Salix salviifolia* nas bacias hidrográficas principais de Portugal Continental) e são esses dados: Temperatura Média do Mês mais Quente do Ano (TMMQA), Temperatura Média do Mês mais Frio do Ano (TMMFA), Temperatura Média anual (TMA), Precipitação Média Anual (PMA), Frequência Relativa da Espécie *Salix salviifolia* (FRESALIX), Coeficiente de Variação da Precipitação (CVP).

3.4.1 Regressão Clássica (OLS)

Um modelo de regressão é uma ferramenta estatística que utiliza o **relacionamento** existente **entre duas ou mais variáveis** de maneira que uma delas possa ser descrita, ou o seu valor estimado, em relação às outras variáveis.

O primeiro passo da regressão espacial é analisar os resíduos do modelo. Para verificar a existência de correlação espacial dever-se-á mapear os resíduos e medir a autocorrelação espacial dos resíduos da regressão. **No caso de existir dependência espacial** os resíduos não são independentes e correlacionam-se. Através do **diagrama de dispersão** é possível **concluir** sobre a **existência de relação entre as variáveis**, se os pontos se situam sobre uma **linha recta** **existe uma relação estatística do tipo linear que é tanto mais forte** quanto mais **próximos** estiverem os **pontos da recta**.

Os modelos de regressão espacial permitem estabelecer as relações entre as variáveis, levando em conta os efeitos espaciais. Para avaliar uma regressão é utilizado o **valor do coeficiente de determinação-R²**, este valor é uma medida da proporção da variabilidade numa variável que é explicada pela variabilidade de outra. Este **coeficiente varia de 0 a 1**, sendo **0 nenhuma correlação** entre as variações das variáveis em questão e **1 a correlação perfeita**.

REGRESSION				
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set	Join_BH+SALIX			
Dependent Variable	OCORRENCIA	Number of Observations	:	118
Mean dependent var	59.5	Number of Variables	:	7
S.D. dependent var	34.0624	Degrees of Freedom	:	111
R-squared	0.861501	F-statistic	:	115.075
Adjusted R-squared	0.854014	Prob(F-statistic)	:	2.7117e-045
Sum squared residual	18961.9	Log likelihood	:	-467.125
Sigma-square	170.828	Akaike info criterion	:	948.251
S.E. of regression	13.0701	Schwarz criterion	:	967.646
Sigma-square ML	160.694			
S.E of regression ML	12.6765			
Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	114.4771	45.96283	2.490645	0.0142317
SALISALV	10.43249	0.4136457	25.22083	0.0000000
TMIN	-0.1108696	1.164677	-0.09519341	0.9243202
TMAX	-1.685365	1.27042	-1.32662	0.1873575
CVP	-76.057	94.51504	-0.8047079	0.4227078
TMA	-1.084324	1.456691	-0.7443747	0.4582232
PMA	-0.01050954	0.005425223	-1.937163	0.0552665

Figura 17 – Relatório de Regressão Clássica 1

MODELO:

$$Y = 114.4771 + 10.43249 X_1 + (-0.1108696) X_2 + (-1.685365) X_3 + (-76.057) X_4 + (-1.084324) X_5 + (-0.01050654) X_6 + \epsilon$$

Em que:

- X1- Frequência relativa da Espécie *Salix salviifolia*;
- X2 – Temperatura Média do Mês mais Quente do Ano;
- X3 – Temperatura Média do Mês mais Frio do Ano;
- X4- Coeficiente de Variação da Precipitação.
- X5 – Temperatura Média anual;
- X6 – Precipitação Média Anual;

Tentou-se utilizar as variáveis do Regime hidrológico permanente e temporário, mas estas provocavam um erro quando se correu o modelo, porque os seus valores são apenas 1 para o regime permanente e 0 para o regime temporário, neste sentido estavam a condicionar os resultados, neste sentido optou-se por excluir estas duas variáveis e só fazer alguns testes em outras análises dos capítulos seguintes.

- X7- Regime Hidrológico Permanente;
- X8- Regime Hidrológico Temporário;

Assim, fazendo-se uma análise dos resultados obtidos verifica-se que existe considerável correlação entre variáveis ($R^2=0.861501$). E os valores do *Log Likelihood* (-467.125), *valor de Akaike info criterion* (948.251) e o *Schwarz criterion* (967.646) também podem ser considerados bastante razoáveis.

REGRESSION DIAGNOSTICS			
MULTICOLLINEARITY CONDITION NUMBER		113.491	
TEST ON NORMALITY OF ERRORS			
TEST	DF	VALUE	PROB
Jarque-Bera	2	9.822666	0.0073627
DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	6	3.664269	0.7220015
Koenker-Bassett test	6	7.41766	0.2839419
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	27	44.54083	0.0181691

Figura 18 – Relatório de Regressão Clássica 2

No **diagnóstico de regressão**, verifica-se que o número de condição de multicolinearidade é de **113.491**, um valor muito superior a 20, o que demonstra problemas no modelo, por outro lado a probabilidade de Jarque-Bera é **0.073627**, indicando que não se trata de uma distribuição normal. Quanto ao diagnóstico de heterocedasticidade, os três testes mostram probabilidades de **0.7220015**, **0.2839419** e **0.0181691** respectivamente, estes resultados mostram alguma heterogeneidade diminuindo a probabilidade ao longo dos três testes, o que poderá indicar um aumento da heterocedasticidade, ou seja, aumenta a dispersão dos dados em torno da recta.

A análise dos resíduos é uma forma de verificar a correlação espacial, assim após a regressão clássica, obteve, através do *Índice Global de Moran* o valor dos resíduos (I) e procedeu-se à sua análise (Figura 19), no que respeita à sua **autocorrelação espacial**.

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
FOR WEIGHT MATRIX : matriz_vizinha.gal (row-standardized weights)			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.624497	8.3934892	0.0000000
Lagrange Multiplier (lag)	1	6.7009503	0.0096362
Robust LM (lag)	1	0.1654686	0.6841708
Lagrange Multiplier (error)	1	61.8794780	0.0000000
Robust LM (error)	1	55.3439964	0.0000000
Lagrange Multiplier (SARMA)	2	62.0449467	0.0000000

Figura 19 – Diagnóstico da dependência espacial, valor dos resíduos

Da análise da figura 19, verifica-se que os resíduos se correlacionam ($I=0.624497$), como se trata de um valor bastante elevado, conclui-se que o modelo de regressão simples (OLS) não é o mais indicado para testar a localização geográfica da espécie *Salix salviifolia*.

3.4.2 Modelos de Regressão Espacial (LAG e ERROR)

Para analisar a hipótese de que o processo espacial presente é estacionário, ou seja, não existe tendência e de que pode ser capturado por apenas um parâmetro aplicaram-se modelos com efeitos espaciais globais, nomeadamente *Spatial LAG Model* e *Spatial Error Model*.

REGRESSION					
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION					
Data set	Join_BH+SALIX				
Spatial Weight	matriz_vizinha.gal				
Dependent Variable	OCORRENCIA	Number of Observations:	118		
Mean dependent var	59.5	Number of Variables	8		
S.D. dependent var	34.0624	Degrees of Freedom	110		
Lag coeff. (Rho)	0.107437				
R-squared	0.869654	Log likelihood	-463.623		
Sq. Correlation	-	Akaike info criterion	943.246		
Sigma-square	151.234	Schwarz criterion	965.412		
S.E of regression	12.2977				
Variable	Coefficient	Std. Error	z-value	Probability	
W_OCORRENCIA	0.1074375	0.03760704	2.856845	0.0042789	
CONSTANT	88.96518	43.77633	2.032267	0.0421265	
SALISALV	9.807746	0.4292944	22.8462	0.0000000	
TMIN	-0.338901	1.100055	-0.3080766	0.7580242	
TMAX	-1.089911	1.205951	-0.9037767	0.3661138	
CVP	-64.0693	88.94643	-0.7203134	0.4713319	
TMA	-0.5791514	1.379782	-0.4197411	0.6746746	
PMA	-0.00886678	0.005115474	-1.733325	0.0830378	
REGRESSION DIAGNOSTICS					
DIAGNOSTICS FOR HETEROSKEDASTICITY					
RANDOM COEFFICIENTS					
TEST	DF	VALUE	PROB		
Breusch-Pagan test	6	4.107307	0.6621574		
DIAGNOSTICS FOR SPATIAL DEPENDENCE					
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : matriz_vizinha.gal					
TEST	DF	VALUE	PROB		
Likelihood Ratio Test	1	7.004672	0.0081297		
COEFFICIENTS VARIANCE MATRIX					
CONSTANT	SALISALV	TMIN	TMAX	CVP	TMA
1916.367445	0.032728	18.415489	-35.478645	-2764.227022	-21.123581
0.032728	0.184294	0.000279	-0.010719	-2.050193	0.023791
18.415489	0.000279	1.210120	-0.013753	-30.962887	-1.210677
-35.478645	-0.010719	-0.013753	1.454319	10.381883	-0.203721
-2764.227022	-2.050193	-30.962887	10.381883	7911.466716	20.396171
-21.123581	0.023791	-1.210677	-0.203721	20.396171	1.903799
-0.186294	0.000228	-0.002448	0.003154	0.227025	0.003253
-0.255329	-0.006813	-0.003614	0.006002	0.065008	0.005974

Figura 20 – Spatial LAG Model

Neste **modelo LAG** podemos verificar a presença de um coeficiente *Rho* ($\rho = 0.107437$) que traduz a dependência espacial inerente aos dados da amostra, medindo a influência média das observações por observações vizinhas, que neste caso é baixa.

REGRESSION				
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION				
Data set	: Join BH+SALIX			
Spatial Weight	: matriz vizinha.gal			
Dependent Variable	: OCORRENCIA	Number of Observations	: 118	
Mean dependent var	: 59.500000	Number of Variables	: 7	
S.D. dependent var	: 34.062443	Degree of Freedom	: 111	
Lag coeff. (Lambda)	: 0.824215			
R-squared	: 0.960520	R-squared (BUSE)	: -399.481187	
Sq. Correlation	: -	Log likelihood	: 812.962	
Sigma-square	: 45.806983	Akaike info criterion	: 832.357166	
S.E of regression	: 6.76809	Schwarz criterion	: -	

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	53.06065	22.43108	2.365497	0.0180059
SALISALV	8.150136	0.3499722	23.28795	0.0000000
TMIN	0.2656248	0.5796852	0.4582224	0.6467927
TMAX	-0.3645771	0.5996914	-0.6079412	0.5432264
CVP	18.09663	42.35354	0.4272754	0.6691788
TMA	-0.4173754	0.7011949	-0.5952345	0.5516866
PMA	-0.002452705	0.002418787	-1.014023	0.3105719
LAMBDA	0.8242146	0.03947225	20.88086	0.0000000

REGRESSION DIAGNOSTICS				
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST	DF	VALUE	PROB	
Breusch-Pagan test	6	3.810025	0.7023670	

DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : matriz vizinha.gal				
TEST	DF	VALUE	PROB	
Likelihood Ratio Test	1	135.2884	0.0000000	

COEFFICIENTS VARIANCE MATRIX						
CONSTANT	SALISALV	TMIN	TMAX	CVP	TMA	
503.153341	0.377221	5.480552	-9.317128	-679.081553	-7.232100	
0.377221	0.122481	0.003095	-0.018651	-1.666864	0.000214	
5.480552	0.003095	0.336035	-0.030050	-7.669456	-0.343705	
-9.317128	-0.018651	-0.030050	0.359630	3.002635	0.003087	
-679.081553	-1.666864	-7.669456	3.002635	1793.822275	7.881971	
-7.232100	0.000214	-0.343705	0.003087	7.881971	0.491674	
-0.045396	-0.000068	-0.000597	0.000798	0.053290	0.000877	
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

Figura 21 – Spatial Error Model

Para este **modelo ERROR** podemos verificar a presença de um coeficiente *Lambda* ($\lambda = 0,824215$), cujo valor indica que **existe uma boa dependência** espacial inerente aos dados da amostra.

3.4.3 Comparação de Modelos de Regressão Espacial

Tabela 1 – Comparação dos valores de cada um dos modelos

Modelo	R ²	Log Likelihood	Akaike Criterion	Schwarz Criterion
OLS	0,861	-467,125	948,251	967,646
LAG	0,869	-463,623	943,246	965,412
ERROR	0,960	-399.481	812,962	832,357

Da análise da tabela 1, conclui-se que o modelo, de entre os testados, que poderá ser melhor é o *Spatial Error Model*, pelo facto de ter um o coeficiente de determinação maior ($R^2=0,960$),

assim como o *Log Likelihood* também ser maior (-399.481) que nos outros dois modelos, e por outro lado os valores dos critérios *Akaike* e *Schwarz* são menores, com valores de 812,962 e 832,357 respectivamente.

Ainda que este seja o modelo que melhor consegue explicar a dependência da ocorrência da espécie *Salix salviifolia* no território de Portugal Continental face às variáveis independentes testadas (frequência relativa da espécie *Salix salviifolia*; temperatura média do mês mais quente do ano; temperatura média do mês mais frio do ano; coeficiente de variação da precipitação, temperatura média anual; precipitação média anual), dever-se-á ter algumas reservas quanto à sua total fiabilidade e testar outros tipos de análise, como por exemplo reduzir a área de estudo, analisando apenas as bacias hidrográficas com maior número de ocorrência da espécie.

4. Métodos de Interpolação

O objectivo é compreender se a distribuição espacial da espécie *Salix slaviifolia* em Portugal Continental com base nos inventários da DQA, poderá ser ou não justificada com as suas exigências ecológicas (condições ambientais e climáticas de cada bacia hidrográfica principal). Para o efeito serão confrontados os dados ambientais e climáticos dos 404 pontos de amostragem da DQA (2004-2006) para todo o território de Portugal Continental).

4.1 Tratamento de dados

Em termos de dados, foram seleccionados dados climáticos e ambientais com maior interesse ecológico para estabelecer correlações com os dados de ocorrência da espécie *Salix salviifolia* em Portugal Continental.

E quanto aos dados ambientais, a informação disponível pelo Atlas do ambiente que está representada à escala de 1:1.000.000, portanto, a grande escala, contudo estamos a trabalhar a uma escala do país, e para colmatar este inconveniente, temos informação mais detalhada em cada ponto de amostragem, e uma vez que temos 404 pontos, facilita o processo de interpolação Kriging no *ArcGis Versão 9.3* [ESRI Inc., 1998-2009].

4.2 Análise Exploratória dos Dados

Nesta etapa, pretende-se fazer o estudo das variáveis ambientais e climáticas através da Geoestatística. Ou seja, estudar a variabilidade das variáveis no espaço, fazendo estimativas de valores desconhecidos em determinados pontos do espaço e a estimação da distribuição de valores médios no espaço, tendo em consideração a localização geográfica e a dependência espacial.

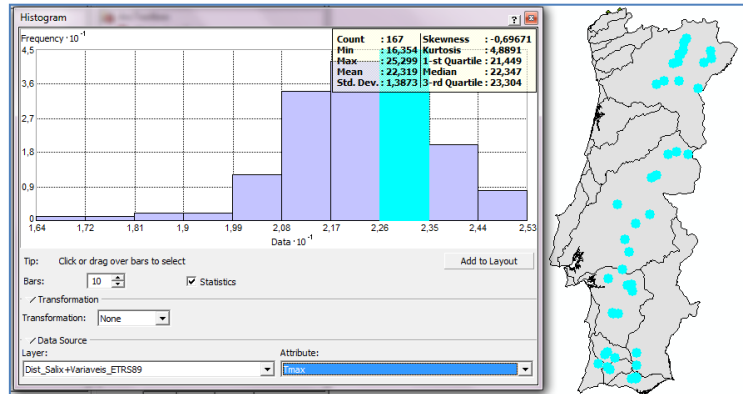
A análise exploratória de dados permite ter um melhor conhecimento do fenómeno antes de serem elaboradas as superfícies. Existem uma série de ferramentas, predominantemente gráficas que suportam este tipo de análise, conduzindo a uma estimação mais enriquecida e completa. Apresentam-se seguidamente os histogramas e algumas medidas estatísticas da dispersão e forma da distribuição das variáveis ambientais e climáticas de maior relevância em termos ecológicos para a distribuição da espécie *Salix salviifolia* no território de Portugal Continental. Neste sentido, o território estará dividido pelas principais bacias hidrográficas, uma vez que estamos a estudar uma espécie rípicola, esta divisão é importante para compreender quais as bacias hidrográficas que apresentam melhores condições ecológicas para a ocorrência desta espécie.

4.2.1 Análise de Histogramas

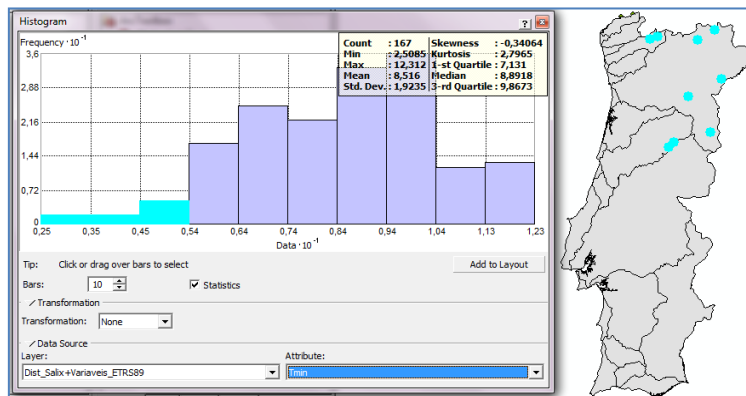
Neste passo é feito a **análise de algumas estatísticas básicas dos atributos** a serem estudados: temperatura média do mês mais quente do ano, temperatura média anual e precipitação média anual, regime hidrológico (permanente e temporário), frequência relativa da espécie *Salix salviifolia* e o coeficiente de variação da precipitação.

O **histograma** é uma ferramenta estatística, sob a forma gráfica, que fornece uma **descrição da frequência da distribuição dos dados**. A média e a mediana são medidas do centro da série. A

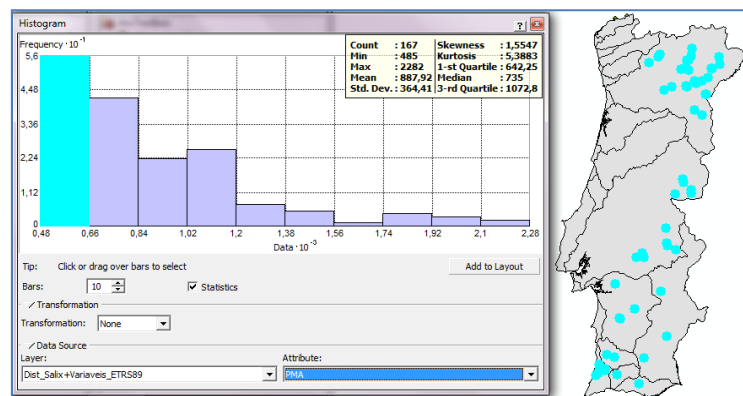
média é o quociente da soma de n números reais pelo seu número, a **mediana** divide a série em duas partes iguais, depois dos seus valores terem sido dispostos por ordem crescente. Da mesma forma o primeiro e o terceiro quartis correspondem, respectivamente, a uma proporção acumulada de 25% e 75% dos valores ordenados da série. Em relação à dispersão dos valores na série, o desvio padrão (raiz quadrada da variância), dá-nos a ideia da dispersão em torno da média. A simetria da distribuição pode ser captada pela média e mediana, ou seja, o valor da média superior ao da mediana indica uma distribuição assimétrica positiva e vice-versa, o que se verifica em ambos os casos.



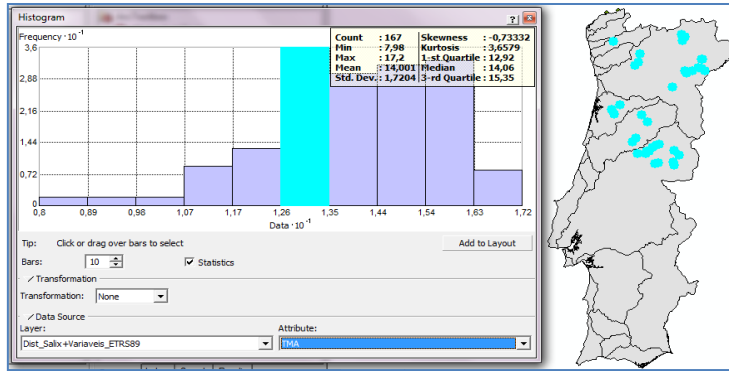
a) Temperatura Média do Mês mais Quente do Ano



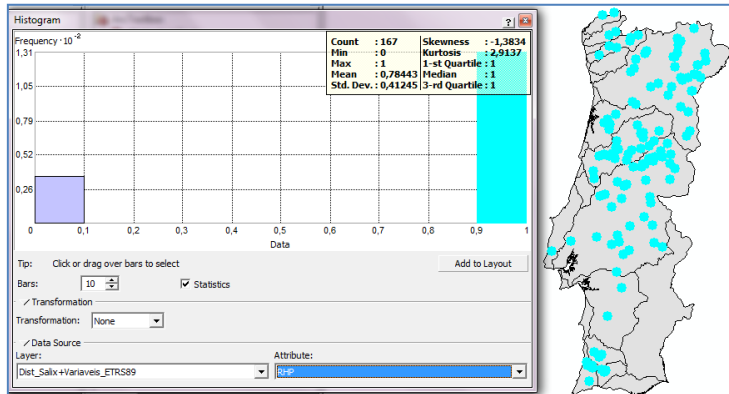
b) Temperatura Média do Mês mais Frio do Ano



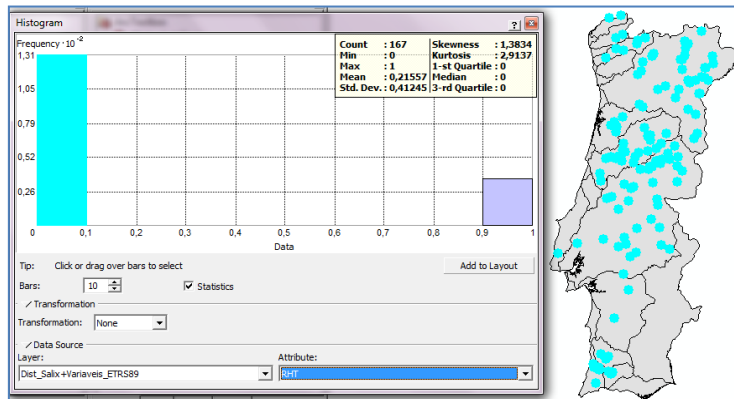
c) Precipitação Média Anual



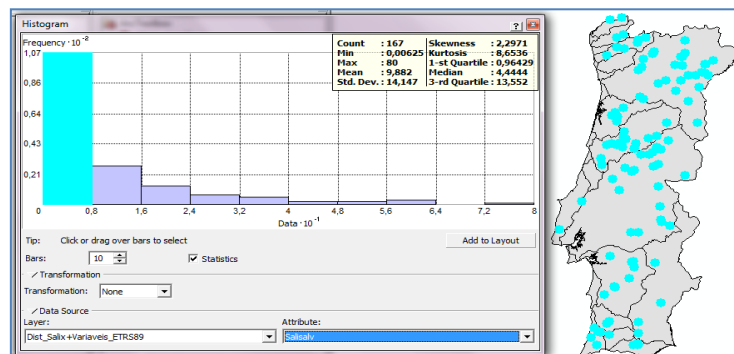
d) Temperatura Média Anual



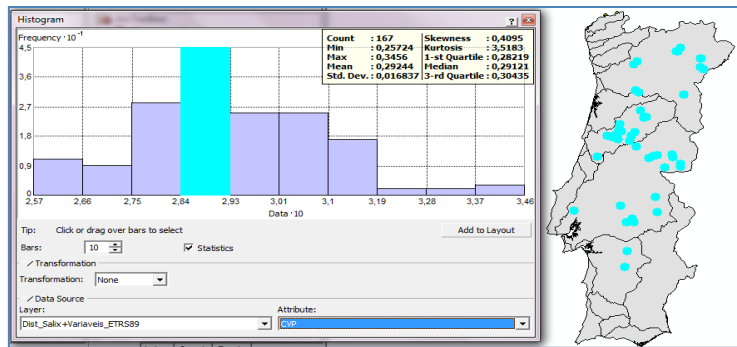
e) Regime Hidrológico Permanente



f) Regime Hidrológico Temporário



g) Frequência relativa da Espécie *Salix salviifolia*



h) Coeficiente de Variação da Precipitação

Figura 22 – Histogramas de todas a variáveis dependentes neste estudo.

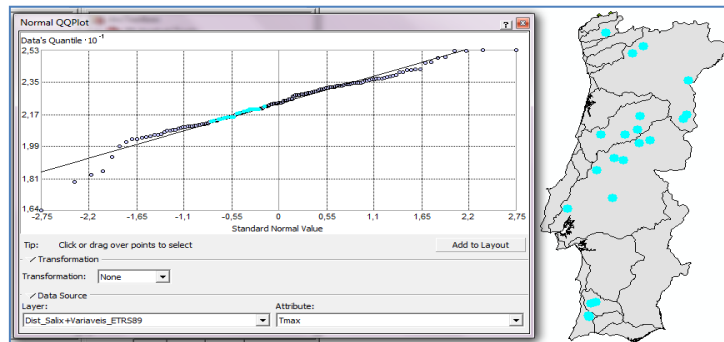
Nos histogramas da Figura 22, podem-se observar os valores máximos e mínimos e média de todas as variáveis dependentes que estão em análise neste estudo.

Como já foi referido anteriormente trata-se de uma **distribuição assimétrica positiva, o que é confirmado pelo** valor do **coeficiente Skewness** que é superior a zero, em todos os histogramas da Figura 22, logo tratam-se de distribuições assimétricas positivas. Quanto mais se aproximar de uma distribuição normal melhores serão os resultados da estimação, e mais facilmente se podem considerar os erros médios relativos, no caso do **kriging**, como intervalos de confiança.

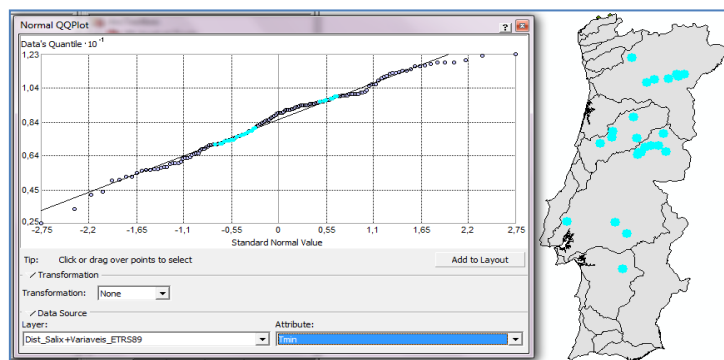
Estes **histogramas estão associados aos mapas**, assim sendo é possível selecionar os dados de uma determinada bacia hidrográfica e respectivos pontos de amostragem, valores das variáveis ambientais e climáticas e visualiza-los na respectiva bacia hidrográfica, como é possível observar individualmente em cada histograma da figura 22.

4.2.2 Avaliação da normalidade dos dados

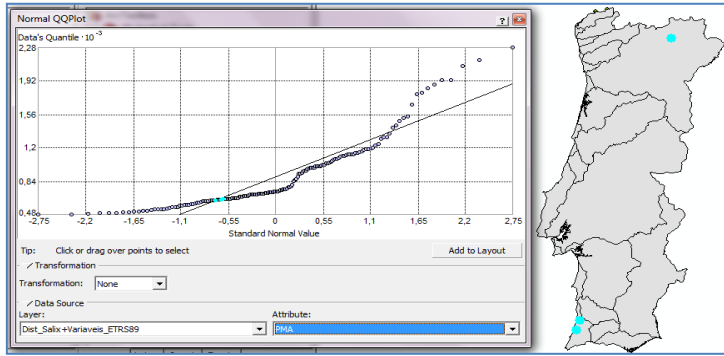
Através do *plot* da normal dos dados é possível avaliar a normalidade dos dados, ou seja, a distribuição dos valores relativamente ao valor médio. Assim pode-se verificar, por exemplo relativamente à Temperatura Média do Mês mais Quente do Ano (alínea a), figura 23), que a distribuição dos dados é normal visto o gráfico seguir a linha diagonal.



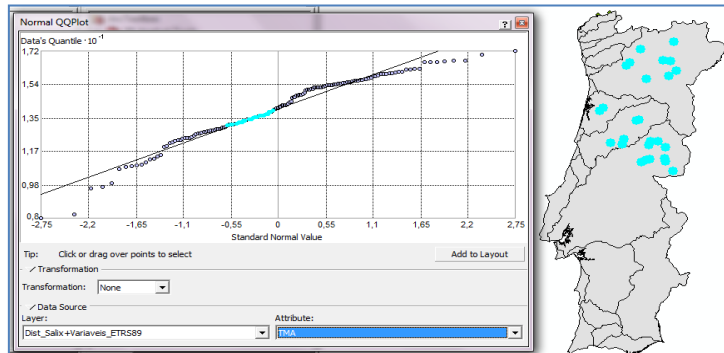
a) Temperatura Média do Mês mais Quente do Ano



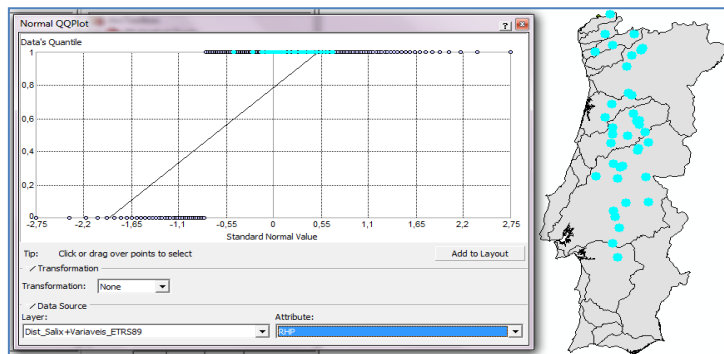
b) Temperatura Média do Mês mais Frio do Ano



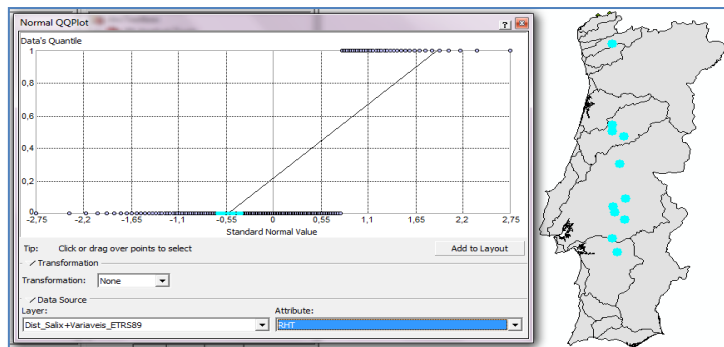
c) Precipitação Média Anual



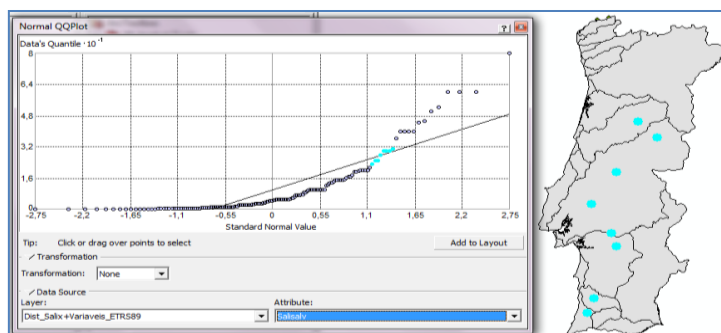
d) Temperatura Média Anual



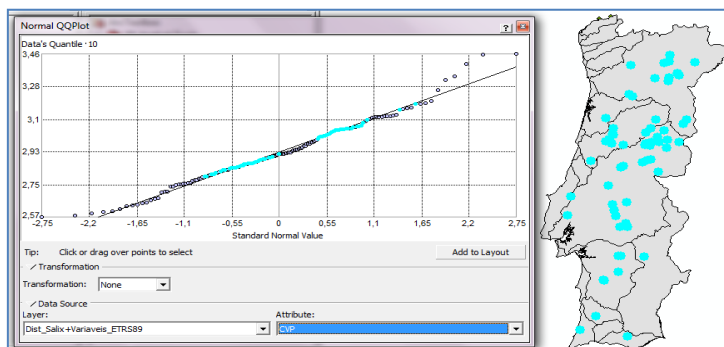
e) Regime Hidrológico Permanente



f) Regime Hidrológico Temporário



g) Frequência relativa da Espécie *Salix salviifolia*



h) Coeficiente de Variação da Precipitação

Figura 23 – QQ Plot da normal dos dados de todas as variáveis dependente em análise. Nota: pela mesma ordem da análise do histograma (ponto 4.2.1)

Similarmente ao que se passa com os histogramas, os **gráficos QQ Plot** têm facilidade em ser **associados ao mapa**, podendo-se visualizar e verificar qual a bacia hidrográfica de um possível **outlier**, como se verifica nos dados de todas as variáveis dependentes (figura 23), no entanto para esta análise foram assinalados em cada mapa, a distribuição dos pontos que melhor se ajustam relativamente ao valor médio.

4.2.3 Modelação da tendência espacial

Uma das **hipóteses** fundamentais subjacentes à **modelação do variograma** é a **ausência de tendência espacial**, ou seja, não pode existir autocorrelação devendo o valor médio, ser constante em toda a área de estudo. Para avaliar a presença de tendência espacial, embora apenas por análise visual, foram gerados oito gráficos tridimensionais para cada uma das variáveis dependentes em análise, que melhor podem explicar a distribuição da espécie *Salix salviifolia* e a sua respectiva localização nas bacias hidrográficas principais do território de Portugal Continental. Em que a ocorrência da espécie *Salix salviifolia* se encontra no plano horizontal e a variação dos valores na dimensão vertical. (Figura 24, alíneas a,b,c,d)

Para melhor interpretar os dados da *Trend Analysis* que foi feita a todas as variáveis dependentes em estudo, na figura 24.1., apresenta-se uma breve descrição dos pontos essenciais para a análise deste tipo de gráficos

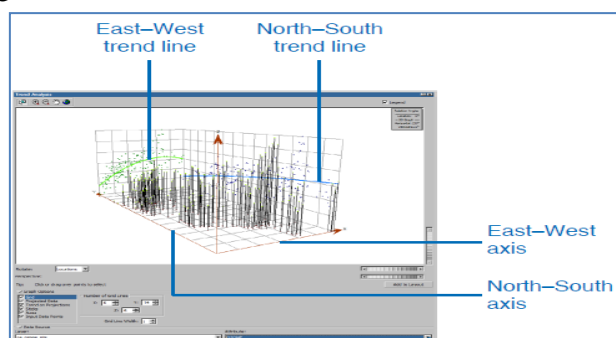
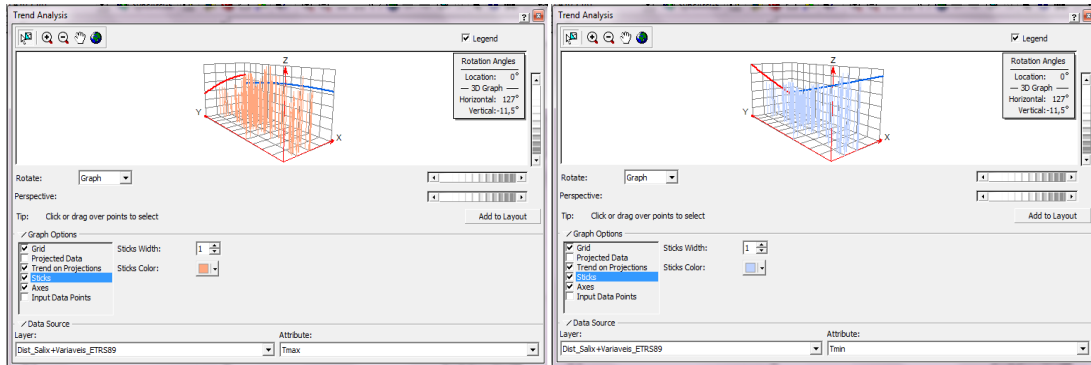


Figura 24.1- Imagem explicativa do gráfico Trend Analysis.

Cada linha vertical do plano de coordenadas 3D, da análise de tendência representa a localização e valor (em altura) de cada um dos pontos em análise. Estes pontos são projectados perpendicularmente aos planos Este-Oeste e Norte-Sul, do sistema de coordenadas, como é possível observar na figura 24.1.

Da análise dos gráficos (Figura 24), de seguida procede-se a uma breve análise da tendência dos pontos de cada variável dependente no plano:



a) Temperatura Média do Mês mais Quente do Ano e Temperatura Média do Mês mais Frio do Ano

Temperatura Média do Mês mais Quente do Ano

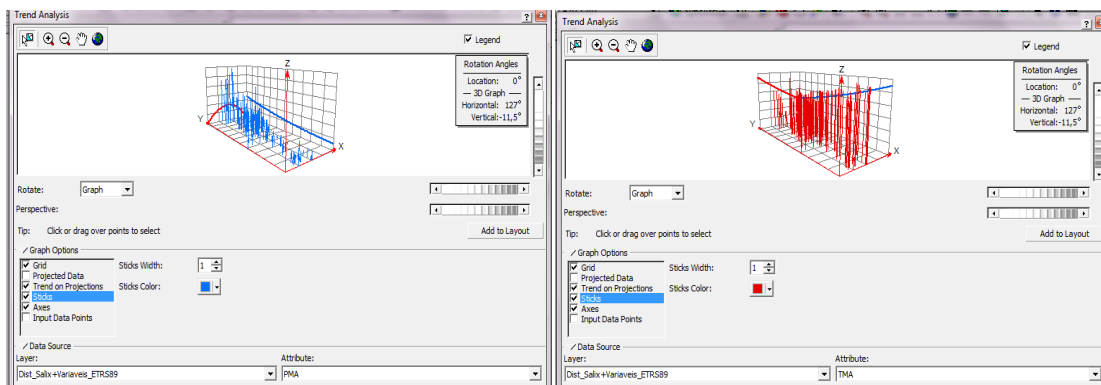
Para esta variável constata-se que na direcção E-W a tendência no território de Portugal Continental é dos valores subirem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores subirem suavemente.

Temperatura Média do Mês mais Frio do Ano

Para esta variável constata-se que na direcção E-W a tendência no território de Portugal Continental é dos valores baixarem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores subirem



b) Precipitação Média Anual e Temperatura Média Anual

Precipitação Média Anual

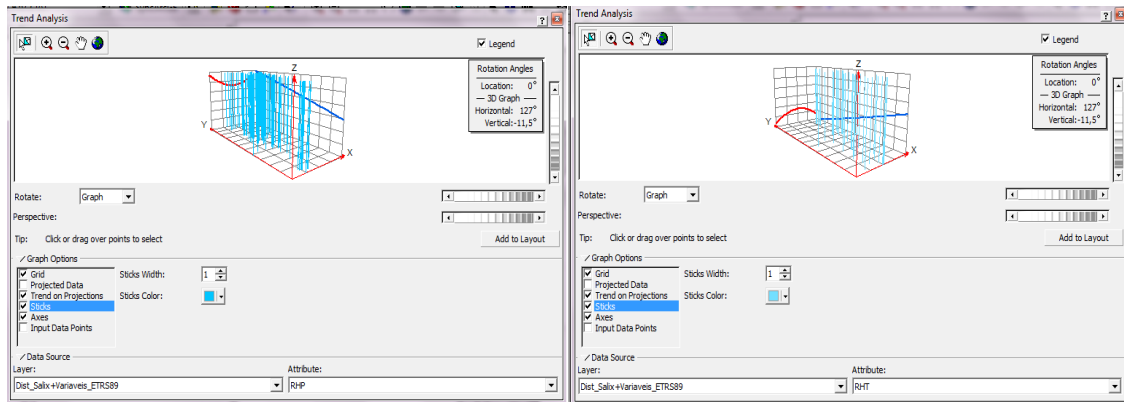
Para esta variável constata-se que na direcção E-W a tendência no território de Portugal Continental é dos valores subirem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores descерem.

Temperatura Média Anual

Para este variável constacta-se que na direcção E-W a tendência no território de Portugal Continental é de valores baixarem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores subirem.



c) Regime Hidrológico Permanente e Regime Hidrológico Temporário

Regime Hidrológico Permanente

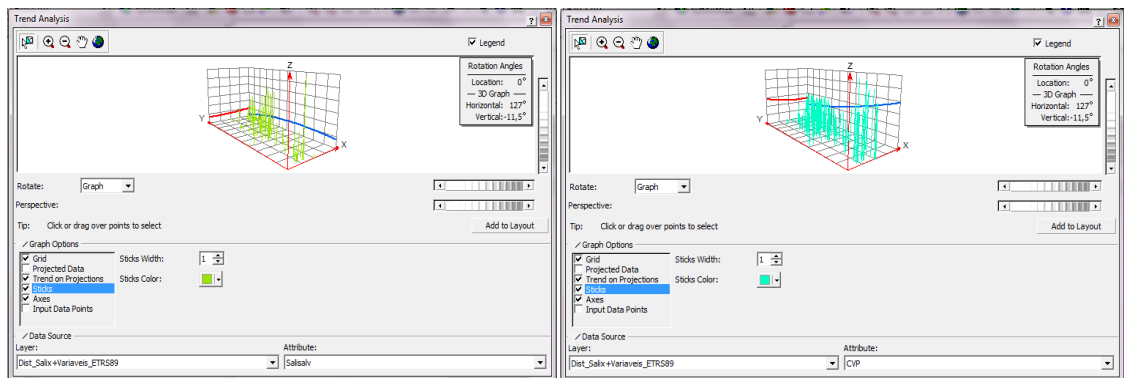
Para este variável constacta-se que na direcção E-W a tendência no território de Portugal Continental é de valores baixarem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos baixarem.

Regime Hidrológico Temporário

Para esta variável constacta-se que na direcção E-W a tendência no território de Portugal Continental é de valores subirem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores subirem.



d) Frequência relativa da Espécie *Salix salviifolia* e Coeficiente de Variação da Precipitação

Frequência relativa da Espécie *Salix salviifolia*

Para este variável constacta-se que na direcção E-W a tendência no território de Portugal Continental é de valores serem mais ou menos constantes, mas com uma muito suave tendência para subir.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores descenderem suavemente.

Coeficiente de Variação da Precipitação

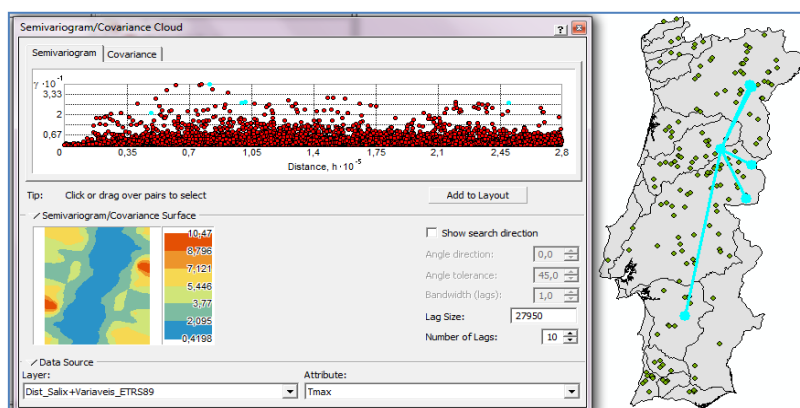
Para esta variável constacta-se que na direcção E-W a tendência no território de Portugal Continental é de valores descenderem.

Quanto à direcção N-S, a tendência desta variável no território de Portugal Continental é dos valores subirem.

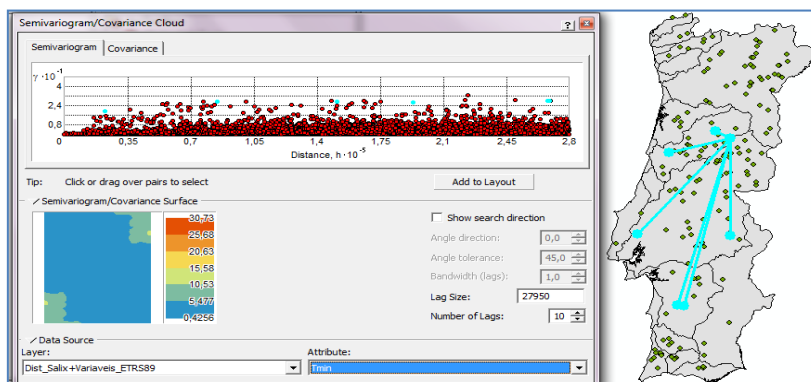
Figura 24 – *Trend Analysis* dos dados de todas as variáveis dependente em análise e respectiva interpretação de resultados. Nota: pela mesma ordem da análise do histograma (ponto 4.2.1), em grupos de duas variáveis.

4.2.3 Semivariograma

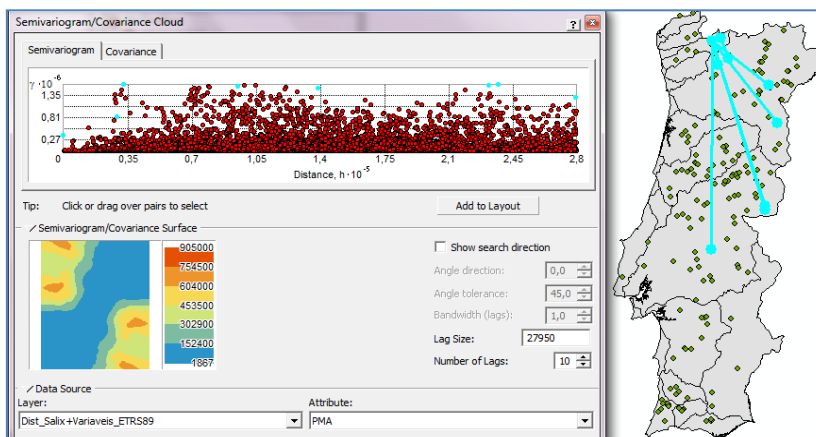
O semivariograma corresponde à nuvem de pontos em análise, ou seja, aos 404 pontos de amostragem dos inventários da DQA para todo o território de Portugal Continental. O gráfico (figura 26) mostra para cada distância (eixo xx) o quadrado da diferença de cada par de observações (eixo yy). Para definir o semivariograma, mediu-se a maior distância entre os pontos de amostragem (559km) e dividiu-se essa distância a metade para se definir o *Lag Size* e o número de *Lag's* (onde a multiplicação dos dois é menor que a metade da distância máxima). Assim sendo, o *Lag Size* estabelecido foi de 27950 metros.



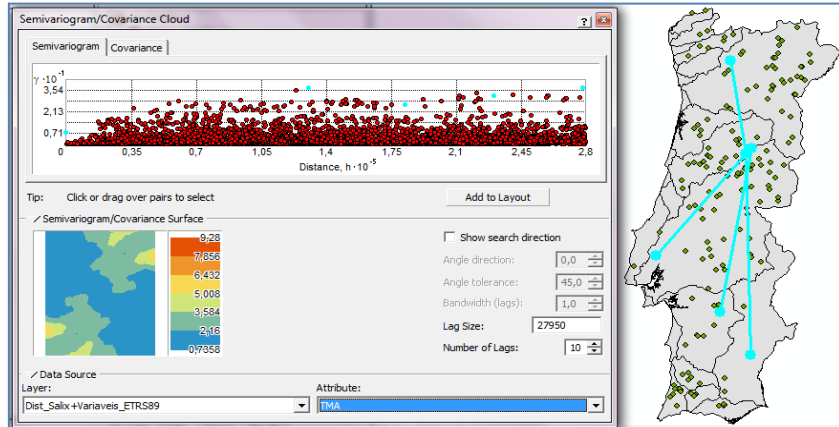
a) Temperatura Média do Mês mais Quente do Ano



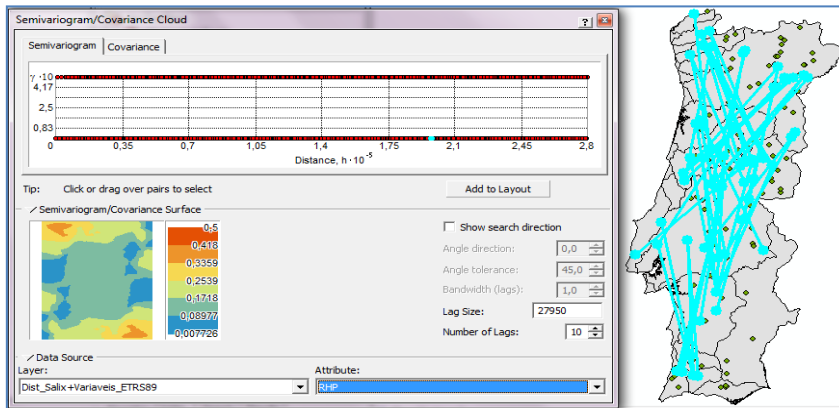
b) Temperatura Média do Mês mais Frio do Ano



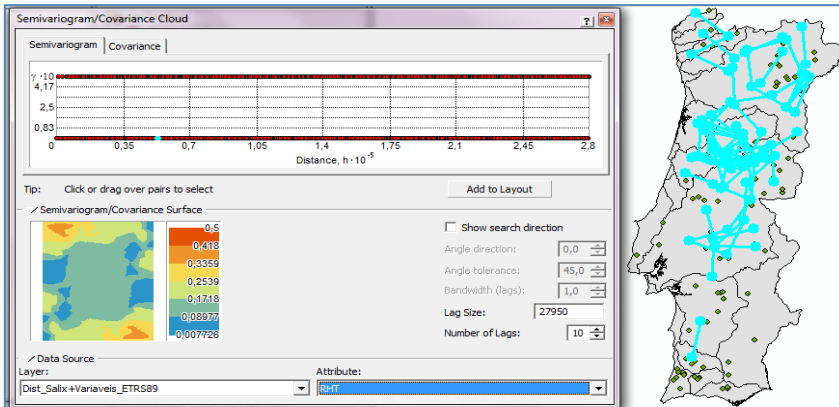
c) Precipitação Média Anual



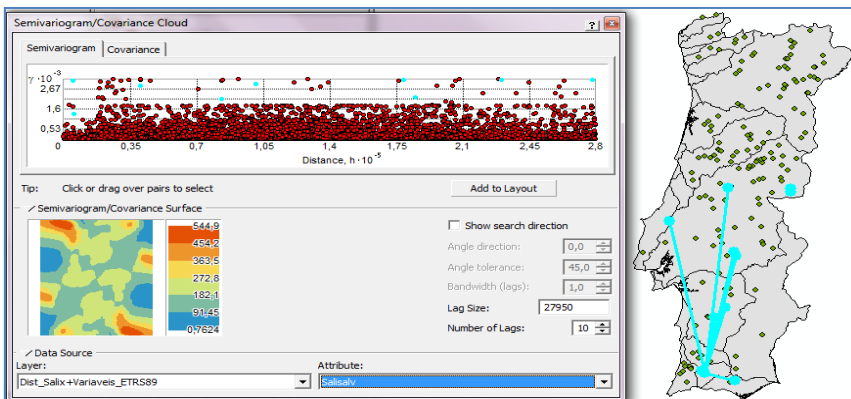
d) Temperatura Média Anual



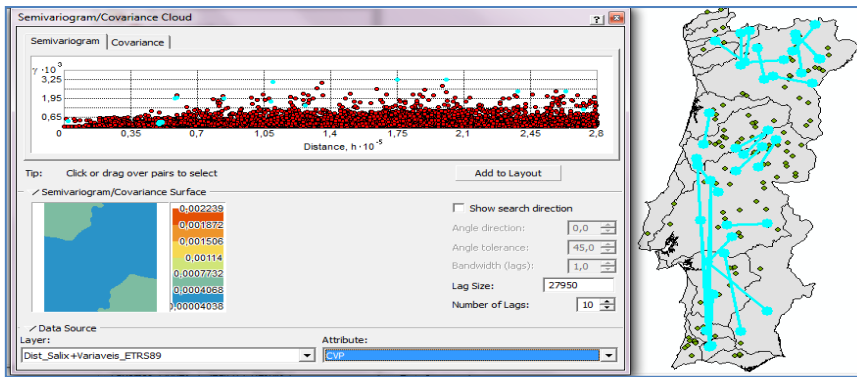
e) Regime Hidrológico Permanente



f) Regime Hidrológico Temporário



g) Frequência relativa da Espécie *Salix salviifolia*

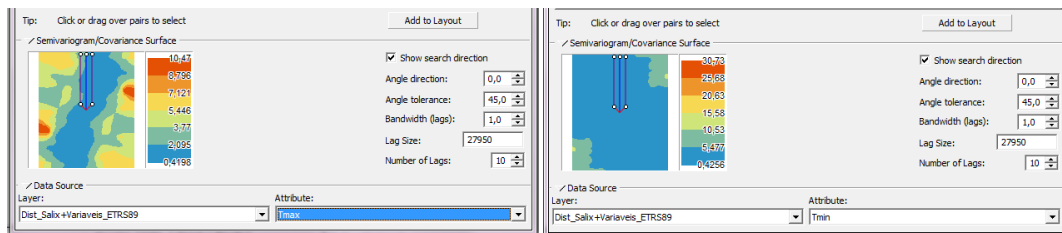


h) Coeficiente de Variação da Precipitação

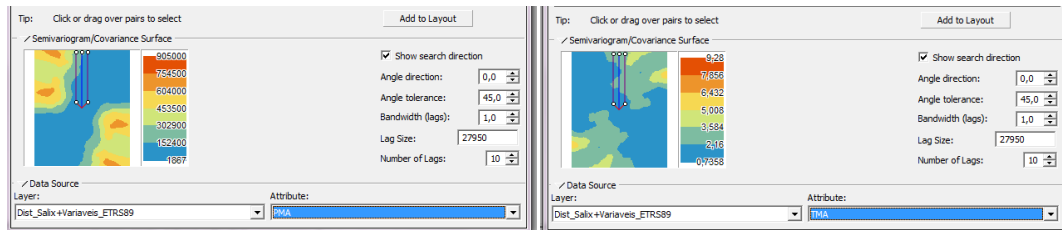
Figura 26 – Semivariogramas dos dados de todas as variáveis dependente em análise. Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

Relativamente ao **ajuste do semivariograma experimental (Figura 26)**, usando os mesmos parâmetros do semivariograma e selecionando diferentes pontos no mesmo e identificando os seus correspondentes no mapa, pode-se verificar possíveis outliers. Isto significa que a diferença quadrada da variável estudada para pares que distam entre si essa distância, é muito superior do que para outros pares, sugerindo um outlier ou um local de “não-estacionaridade”.

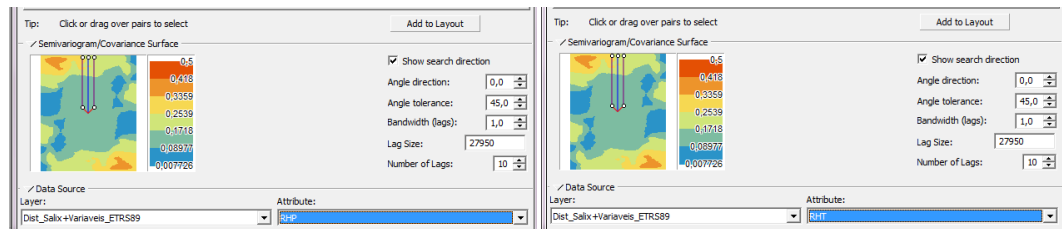
Em relação aos **efeitos direcionais**, usando a superfície do semivariograma é mostrado a média das diferenças quadradas para todos os pares de pontos amostrados que verificam a combinação distância/direcção estabelecida. Assimetrias na superfície sugerem efeitos direcionais, o qual se pode verificar nas diferentes variáveis, de acordo com **figura 27**.



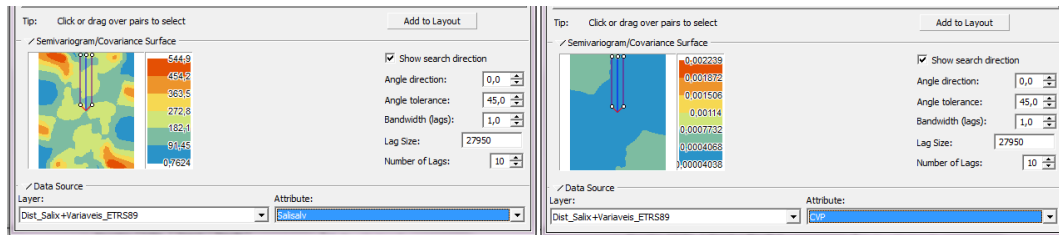
a) Temperatura Média do Mês mais Quente do Ano e Temperatura Média do Mês mais Frio do Ano



b) Precipitação Média Anual e Temperatura Média Anual



c) Regime Hidrológico Permanente e Regime Hidrológico Temporário



d) Frequência relativa da Espécie *Salix salviifolia* e Coeficiente de Variação da Precipitação

Figura 27 – Efeitos direccionais dos dados de todas as variáveis dependente em análise. Nota: pela mesma ordem na análise do histograma (ponto 4.2.1).

4.3 Interpolação Espacial

A **interpolação espacial** é baseada na informação espacial do **semivariograma** e implementada na análise geoestatística. Assim foi gerada uma superfície interpolada mantendo os parâmetros por defeito, contudo tendo em conta a avaliação efectuada anteriormente (semivariogramas, efeitos direccionais, entre outros).

4.3.1 Inverso da distância ponderada (IDW)

É um método de interpolação onde se estima valores onde não existem pontos de ocorrência da espécie *Salix salviifolia*, utilizando uma média ponderada dos pontos de ocorrência circundantes, em função da distância. Esta relação é inversamente proporcional, uma vez que **quanto maior a distância menor o peso** que tem na média, e **quanto menor a distância maior o peso**. É considerado o método mais simples de interpolação, pois atribui o maior peso ao ponto mais próximo, diminuindo esse peso com o aumento da distância.

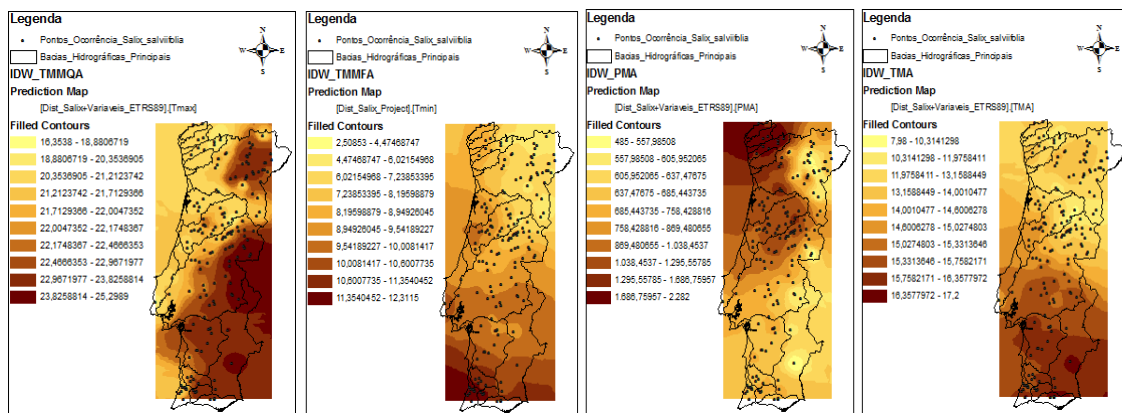


Figura 28.1 – Mapas de estimação de todas as variáveis dependente em análise, por método de interpolação IDW.

Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

Da análise das superfícies de interpolação geradas na **figura 28.1** conclui-se que para as Bacia Hidrográfica (BH) principais os valores das variáveis são os seguintes:

Temperatura Média do Mês mais Quente do Ano (TMMQA)

A Noroeste da BH do Douro, Noroeste da BH do Tejo e BH's do Sado e Guadiana, é onde existe maior TMMQA.

Temperatura Média do Mês mais Frio do Ano (TMMFA)

Toda a BH Douro, a Norte da BH do Mondego e Norte da Bh do Tejo, é onde existe maior TMMFA.

Precipitação Média Anual

Valores mais altos nas BH's do Minho, Lima, Cávado, Ave e na zona litoral da BH do Douro.

Temperatura Média Anual

Valores mais altos nas BH'S do Sado, Guadiana, Várias Algarve e a Sul da BH do Tejo.

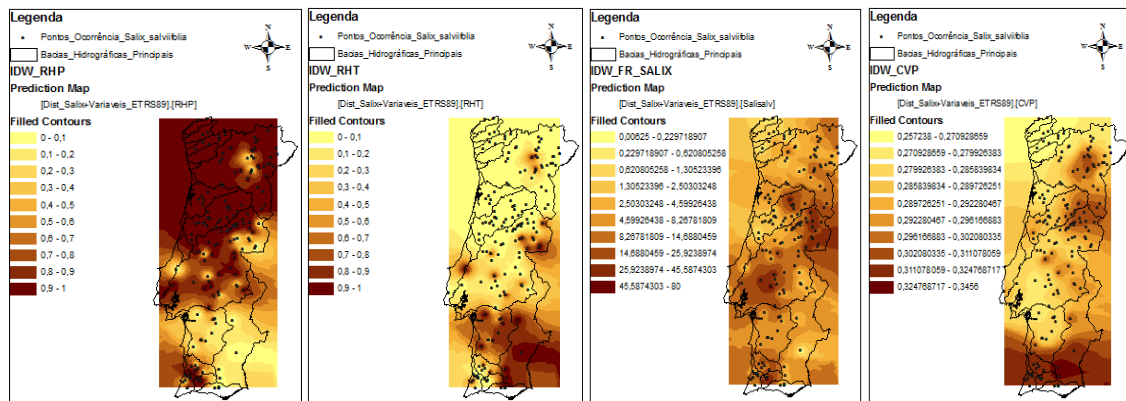


Figura 28.2 – Mapas de estimação de todas as variáveis dependente em análise, por método de interpolação *IDW*.
Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

Da análise das superfícies de interpolação geradas na **figura 28.2** conclui-se que para as Bacia Hidrográfica (BH) principais os valores das variáveis são os seguintes:

Regime Hidrológico Permanente

Está mais presente nas BH's do Minho, Lima, Ave, Cávado, Douro, Vouga, Mondego, pequena zona a Norte da BH do Tejo, Mira e Arade.

Regime Hidrológico Temporário

Mais presente nas restantes bacias hidrográficas.

Frequência relativa da Espécie *Salix salviifolia*

Maior frequência da espécie nas BH's do Tejo, Mondego e Arade.

Coefficiente de Variação da Precipitação

Valores mais altos a Oeste da BH's do Douro e Mondego, Noroeste BH do Tejo, zona Sul da BH do Guadiana e BH's do Mira e Arade.

4.3.2 Kriging

Kriging é um método de regressão usado em geoestatística para **aproximar ou interpolar dados**. Pode ser entendido como uma **predição linear** que parte do princípio que pontos próximos no espaço tendem a ter valores mais parecidos do que pontos mais afastados. O *kriging* assume que os dados recolhidos de uma determinada população se encontram correlacionados no espaço.

É **baseado numa função contínua que explica o comportamento de uma variável nas diversas direções de um espaço geográfico**, e permite associar a variabilidade da estimativa com base na distância que existe entre um par de pontos, pelo uso de um **semivariograma**, o qual permite verificar o nível de similitude que existe entre pontos à medida que estes se afastam.

O método usado foi o **kriging Ordinário**, onde as médias locais não são necessariamente próximas da média da população (valor constante mas desconhecido) usando-se apenas os pontos vizinhos para a estimação. É o método mais usado em problemas ambientais.

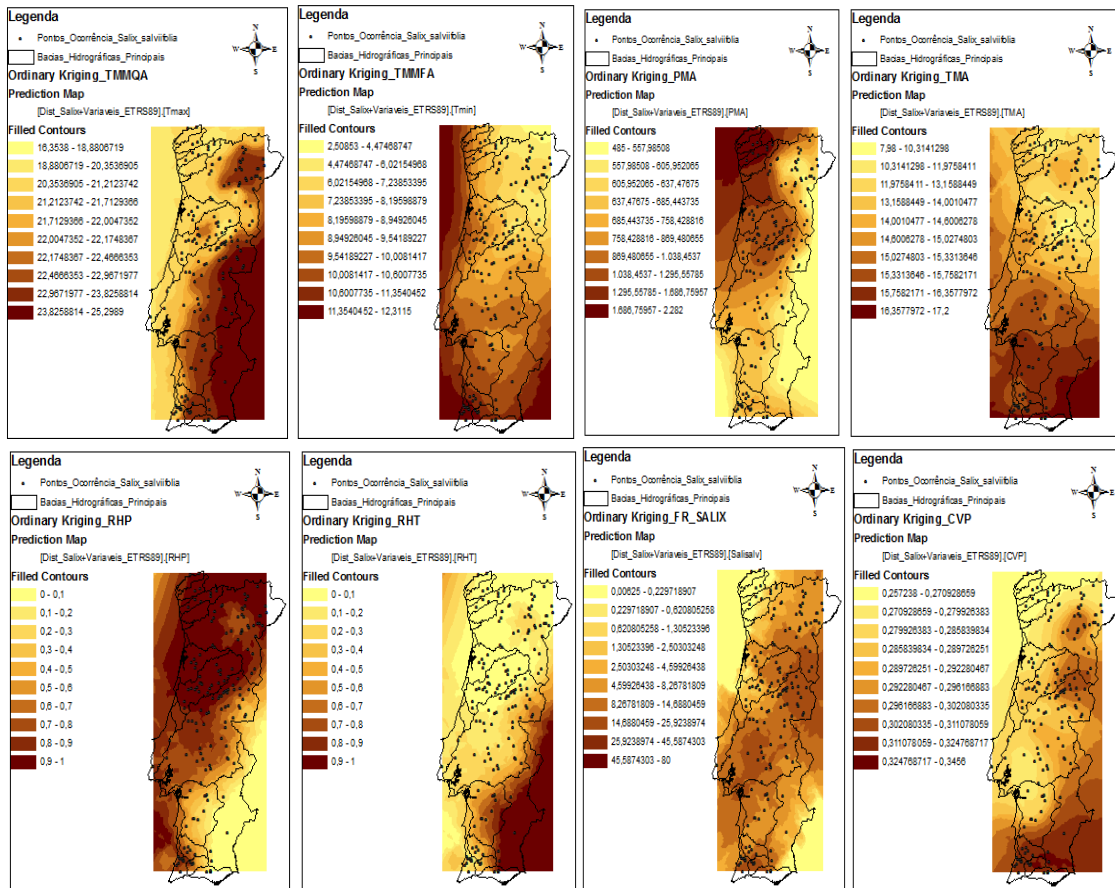
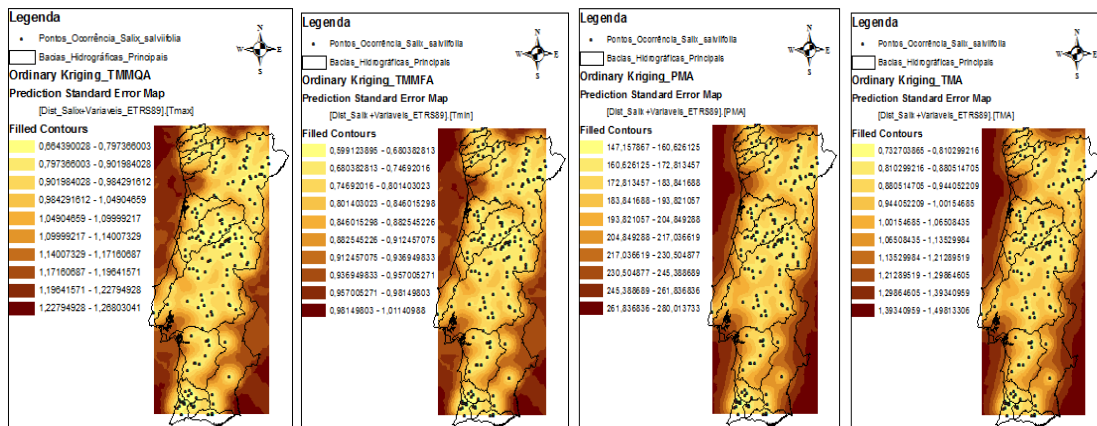


Figura 29 – Mapas de estimação de todas as variáveis dependente em análise, por método de interpolação Kriging. Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

Da análise das superfícies de interpolação geradas conclui-se que são consistentes com as do método anterior, embora que mais suavizadas. Neste método a função semivariograma é ajustada aos pontos de ocorrência da espécie *Salix salviifolia*, através do modelo esférico. Também se gerou uma superfície de previsão do erro padrão (figura 30), observando-se, que o erro aumenta das cores mais claras para as mais escuras, ou seja, o aumento da distância aos pontos de ocorrência da espécie *Salix salviifolia*.



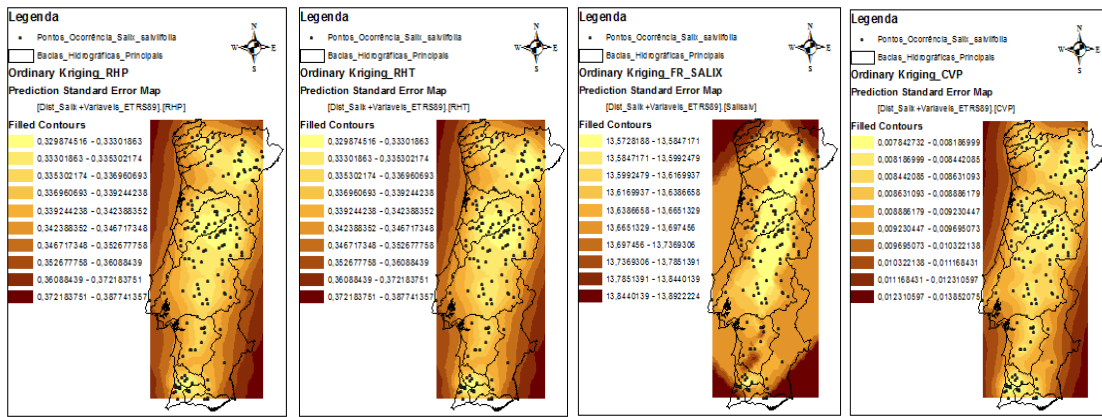


Figura 30 – Mapa de estimação do **erro padrão** de todas as variáveis dependente em análise.

Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

Mas como a análise visual não é um método suficiente para avaliar a qualidade da interpolação, procedeu-se à quantificação do ajustamento.

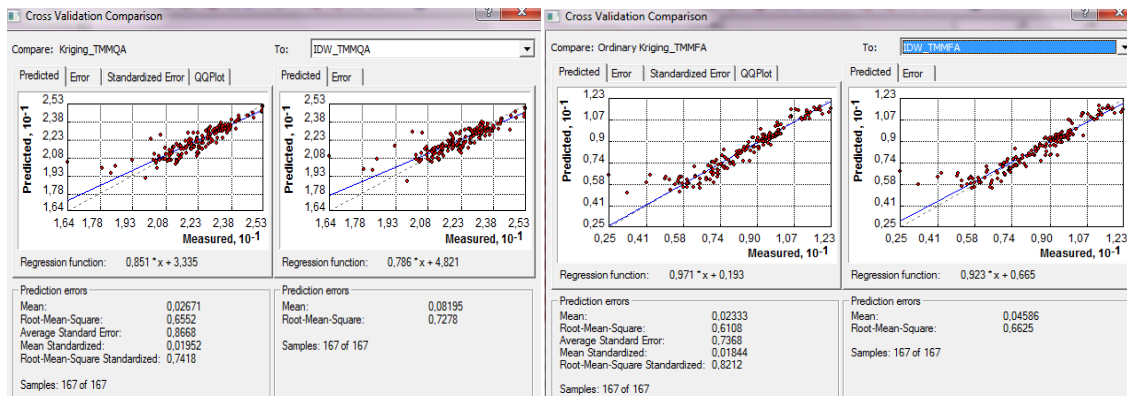
4.3.2 Avaliação da qualidade da interpolação

Os erros da estimação foram obtidos por validação cruzada (*cross-validation*). A qualidade da interpolação pode ser avaliada através dos erros associados à estimação, a interpolação é melhor quando os erros apresentarem os seguintes valores:

- O Erro Médio estimado deve ser o mais próximo possível de zero (**Mean**);
- O Erro Médio Quadrático (EMQ) deve ser o mais baixo possível (**Root-Mean-Square**);
- A Média dos Erros Normalizada deve ser o mais baixa possível (**Average Standard Error**);
- O Erro Médio Quadrático Normalizado deve ser o mais próximo possível de 1 (**Root-Mean-Square standardized**).

Interessa que os erros apresentem as características descritas, pois tal significa que a diferença, entre **valores reais e os valores estimados**, obtida por validação cruzada **é o mais baixa possível**. Logicamente, os erros são maiores nos locais com menor densidade da espécie *Salix salviifolia* pelo que a análise dos valores nestas zonas terá de ser mais cuidada, pois estamos perante uma área de estudo de grande dimensão (Território de Portugal Continental).

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging* (figura 31, alíneas a, b,c,d), de seguida apresentam-se os resultados verificados para cada variável dependente e as conclusões sobre a sua estimação com base nos pressupostos de qualidade da interpolação referindo anteriormente.



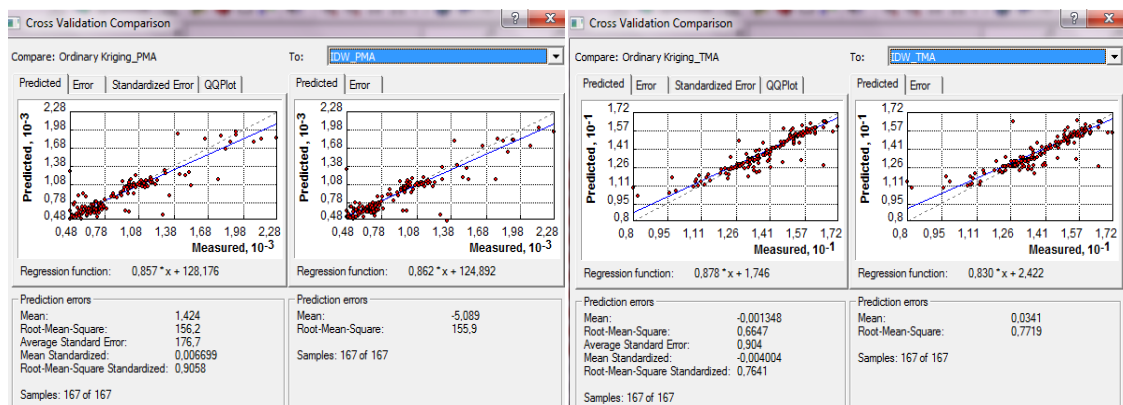
a) Temperatura Média do Mês mais Quente do Ano e Temperatura Média do Mês mais Frio do Ano

Temperatura Média do Mês mais Quente do Ano (TMMQA)

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimacão tem qualidade em termos da TMMQA. Também se constata que como o erro médio quadrático (**Root-Mean-Square**) é superior à média dos erros normalizados (**Average Standard Error**), então existe variabilidade da estimacão subestimada.

Temperatura Média do Mês mais Frio do Ano (TMMFA)

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimacão tem qualidade em termos da TMMF. Também se constata que como o erro médio quadrático (**Root-Mean-Square**) é superior à média dos erros normalizados (**Average Standard Error**), então existe variabilidade da estimacão subestimada.



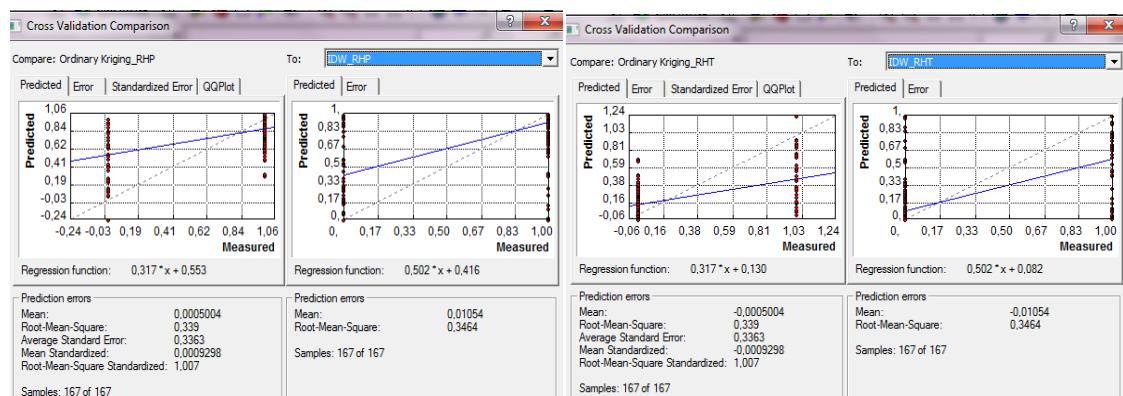
b) Precipitação Média Anual e Temperatura Média Anual

Precipitação Média Anual (PMA)

Relativamente à estimacão desta variável, verifica-se anormalidade nos valores obtidos por *cross-validation*, ou seja, valores altos para **Mean**, **Root-Mean-Square**, **Average Standard Error**, sugerindo que a qualidade desta estimacão é fraca.

Temperatura Média Anual (TMA)

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimacão tem qualidade. Também se constata que o erro médio quadrático (**Root-Mean-Square**) é superior à média dos erros normalizados (**Average Standard Error**), então existe variabilidade da estimacão subestimada.



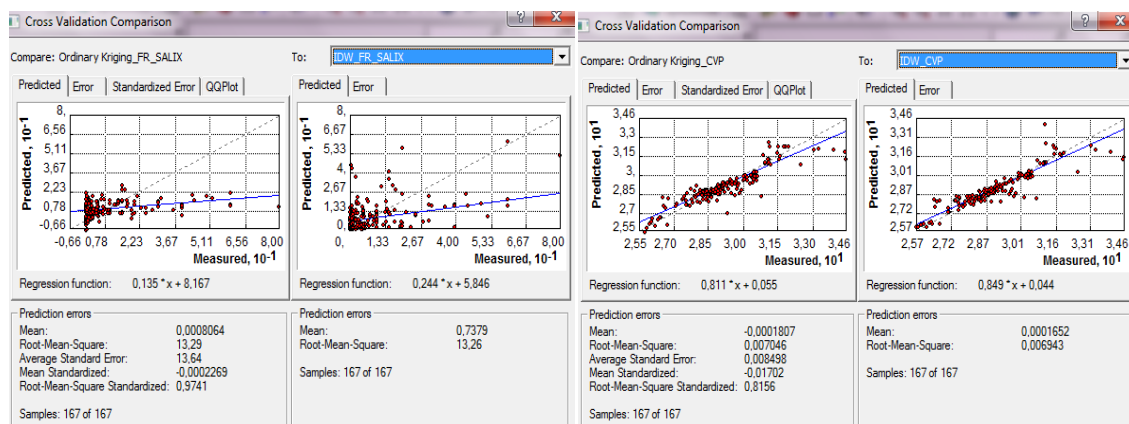
c) Regime Hidrológico Permanente e Regime Hidrológico Temporário

Regime Hidrológico Permanente (RHP)

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimação tem qualidade. Também se constata que como o erro médio quadrático (*Root-Mean-Square*) é superior à média dos erros normalizados (*Average Standard Error*), então existe variabilidade da estimação subestimada.

Regime Hidrológico Temporário (RHT)

Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimação tem qualidade. Também se constata que como o erro médio quadrático (*Root-Mean-Square*) é superior à média dos erros normalizados (*Average Standard Error*), então existe variabilidade da estimação subestimada.



d) Frequência relativa da Espécie *Salix salviifolia* e Coeficiente de Variação da Precipitação

Frequência relativa da Espécie *Salix salviifolia* (FR_SALIX)

Relativamente à estimação desta variável, verifica-se anomalias nos valores obtidos por *cross-validation*, ou seja, valores altos para, *Root-Mean-Square*, *Average Standard Error*, sugerindo que a qualidade desta estimação é fraca.

Coeficiente de Variação da Precipitação (CVP)

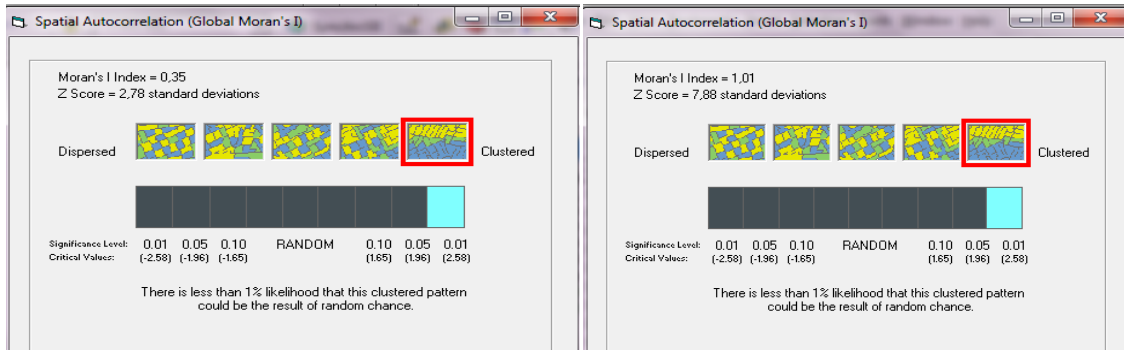
Da análise da validação cruzada efectuada entre o método *IDW* e *Kriging*, verifica-se que para esta variável cumpre de forma aceitável todos os valores dos pressupostos da qualidade da interpolação anteriormente mencionados, concluindo-se que a estimação tem qualidade em termos da CVP. Também se constata que como o erro médio quadrático (*Root-Mean-Square*) é superior à média dos erros normalizada (*Average Standard Error*), então existe variabilidade da estimação subestimada.

Figura 31 – *Cross-Validation* de todas as variáveis dependente em análise e respectiva interpretação de resultados. Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

4.3.3 Avaliação da Autocorrelação Espacial

De modo a saber se o padrão de amostras das variáveis dependentes são espacialmente autocorrelacionadas, usou-se o *Índice de Moran* para encontrar **hotspots locais**.

Da análise da figura 32, de seguida apresentam-se os resultados quanto à existência ou não de autocorrelação espacial para cada variável dependente:



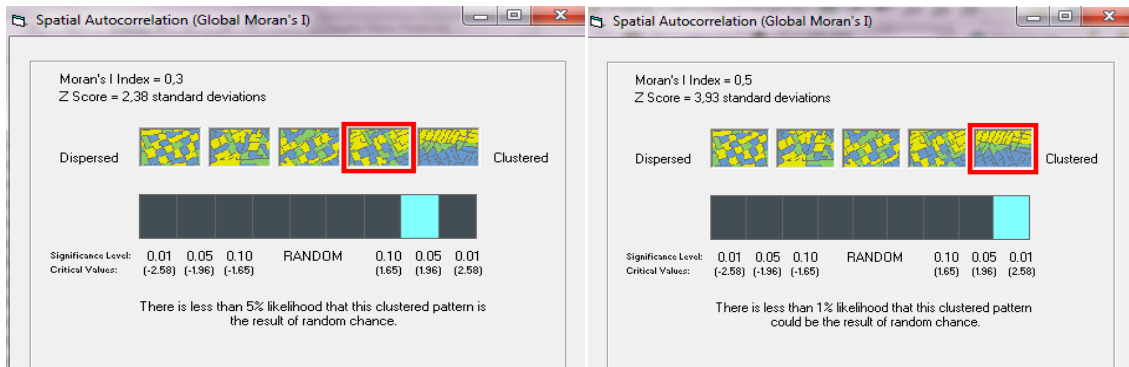
a) Temperatura Média do Mês mais Quente do Ano e Temperatura Média do Mês mais Frio do Ano

Temperatura Média do Mês mais Quente do Ano (TMMQA)

Para TMMQA existe autocorrelação espacial em forma de aglomerados.

Temperatura Média do Mês mais Frio do Ano (TMMFA)

Para TMMFA existe autocorrelação espacial em forma de aglomerados.



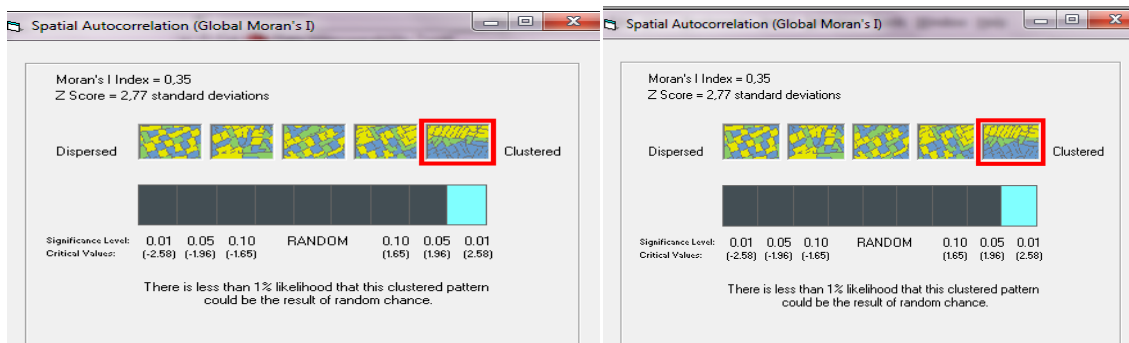
b) Precipitação Média Anual e Temperatura Média Anual

Precipitação Média Anual (PMA)

Para PMA existe autocorrelação espacial.

Temperatura Média Anual (TMA)

Para TMA existe autocorrelação espacial em forma de aglomerados.



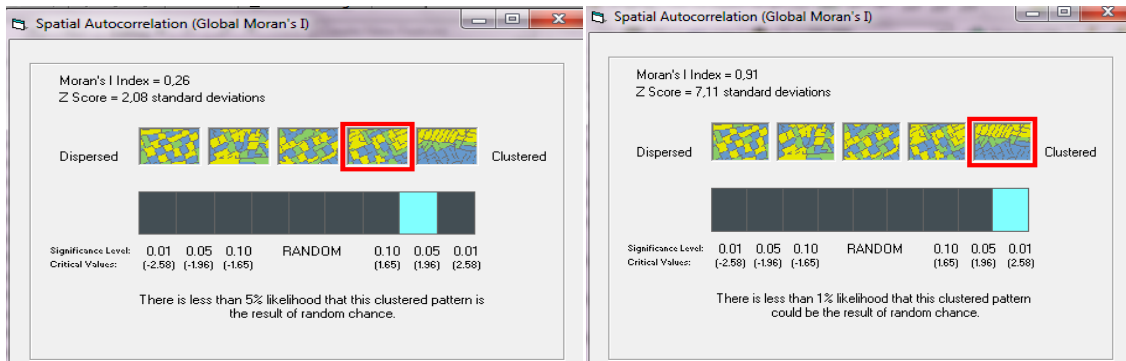
c) Regime Hidrológico Permanente e Regime Hidrológico Temporário

Regime Hidrológico Permanente (RHP)

Para RHP existe autocorrelação espacial em forma de aglomerados.

Regime Hidrológico Temporário (RHT)

Para RHT existe autocorrelação espacial em forma de aglomerados.



d) Frequência relativa da Espécie *Salix salviifolia* e Coeficiente de Variação da Precipitação

Frequência relativa da Espécie *Salix salviifolia* (FR_SALIX)

Para FR_SALIX existe autocorrelação espacial.

Coeficiente de Variação da Precipitação (CVP)

Para CVP existe autocorrelação espacial em forma de aglomerados.

Figura 32 – Autocorrelação de todas as variáveis dependente em análise e respectiva interpretação de resultados.

Nota: pela mesma ordem da análise do histograma (ponto 4.2.1).

De uma forma geral todas as variáveis apresentam autocorrelação espacial em forma de aglomerados, com excepção do das variáveis PMA e FR_SALIX, que apresentam uma autocorrelação espacial com um nível de significância mais baixo. Para melhor visualizar os *cluster* e *outliers*, fez-se uma última análise recorrendo ao *Índice Local de Moran*. (figura 33), determindo igualmente para todas as variáveis dependentes em estudo neste trabalho.

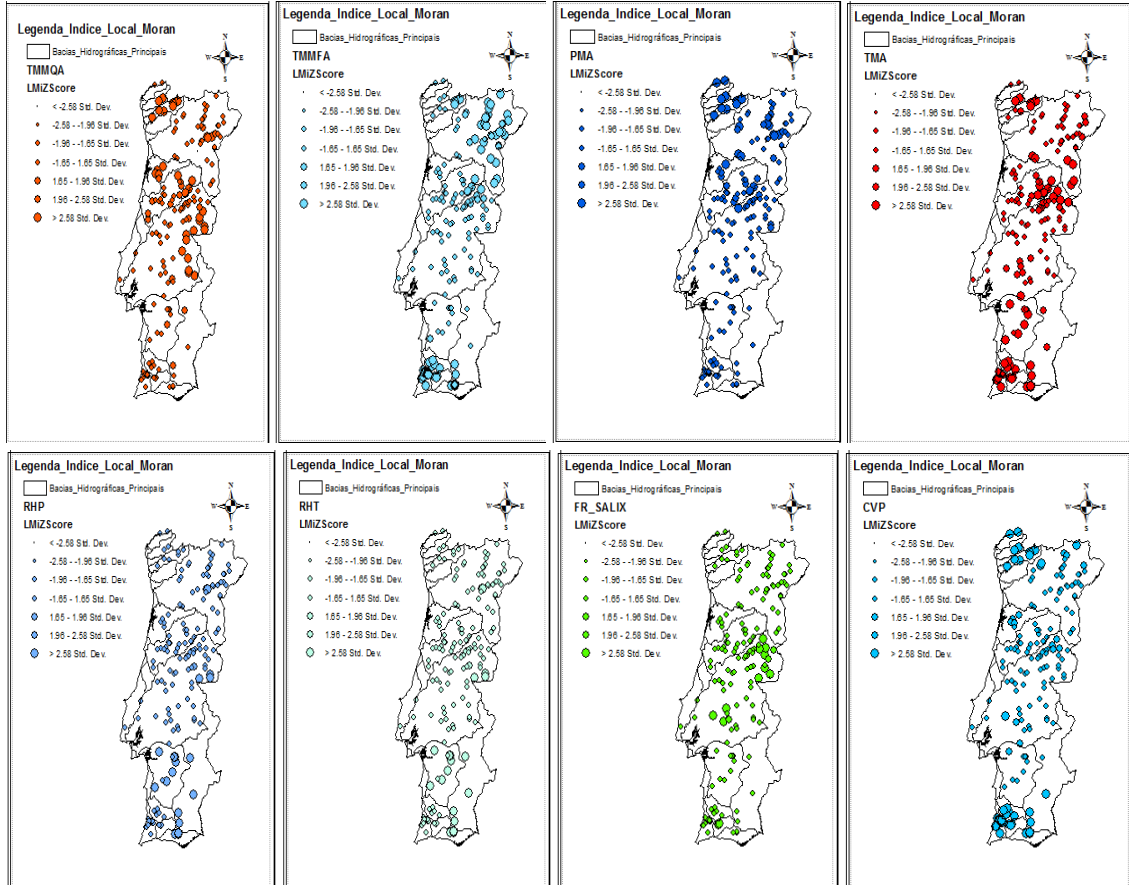


Figura 33 – Mapeamento de *Clusters* e *Outliers* para todas as variáveis dependente em análise.

5. Análise de Redes

O objectivo deste capítulo, após a compreensão da distribuição espacial da ocorrência da espécie *Salix salviifolia* no território de Portugal Continental (divisão por bacias hidrográficas principais) com base nos inventários da DQA, é definir zonas para repositório de material vegetal que posteriormente será utilizado em obras de requalificação ambiental. Neste sentido com base na informação dos pontos de ocorrência da espécie, pretende-se assinalar a um nível de maior detalhe as áreas homogéneas dentro da rede hidrográfica, definindo uma distância aceitável no perfil transversal e longitudinal da galeria ripícola, tendo em conta que estas zonas já têm boas condições ecológicas para a reprodução da espécie.

5.1 Tratamento de dados

Com a informação disponível dos pontos de ocorrência da espécie *Salix salviifolia* e depois da determinação de *cluster* com base na informação da frequência relativa desta espécie, para se perceber, dentro dos pontos de ocorrência, quais as zonas cuja população deste indivíduos é maior e em que zonas das bacias hidrográficas principais se encontram, e por fim, chegando a um nível mais detalhado, foi introduzida a rede hidrográfica para todo o território de Portugal Continental.

Ao acrescentar a rede hidrográfica de todo o território de Portugal Continental, foi criada uma nova *Network Dataset*, que se trata de uma colecção de elementos de rede topologicamente conectados (arestas, junções e curvas) que normalmente são utilizados para representar redes de estradas. Cada elemento de rede está associado com um conjunto de atributos de rede e os conjuntos de dados de rede são normalmente utilizados para modelar sistemas de fluxo sem direcção. Assim quando se executa uma análise usando *ArcGIS Network Analyst*, no *ArcGis Versão 9.3* [ESRI Inc., 1998-2009], a análise é feita sempre a um conjunto de dados de rede.

5.2 Área de Influência

Pelos objectivos já expostos no início deste capítulo, determinou-se as áreas de influência dos pontos de ocorrência da espécie *Salix salviifolia*, que têm as mesmas condições ecológicas favoráveis para otimizar a reprodução e até a transferência de material vegetal para outras áreas com condições semelhantes. Para o efeito, foi necessário recorrer ao algoritmo de análise de redes: *Service Area*. Com este algoritmo, determinou-se a área de influência para cada ponto de ocorrência com maior frequência da espécie, dentro da rede hidrográfica e considerando o perfil transversal e longitudinal da galeria ripícola, foram seleccionadas duas distâncias que se julgou serem as mais coerentes para este caso de estudo: 1000 metros e 3000 metros. Assim, definiram-se *buffers* em torno do perfil transversal e longitudinal da galeria ripícola, tendo em consideração os pontos de ocorrência da espécie, rede hidrográfica, e outra informação adicional como clusters para identificar as zonas das bacias hidrográficas cuja frequência relativa da espécie é mais elevada. Dado as grandes dimensões da área de estudo, foi escolhida a bacia hidrográfica do rio Tejo, para esplanar as áreas de influência para servirão de repositório de material vegetal, como se pode verificar pela **figura 34**.

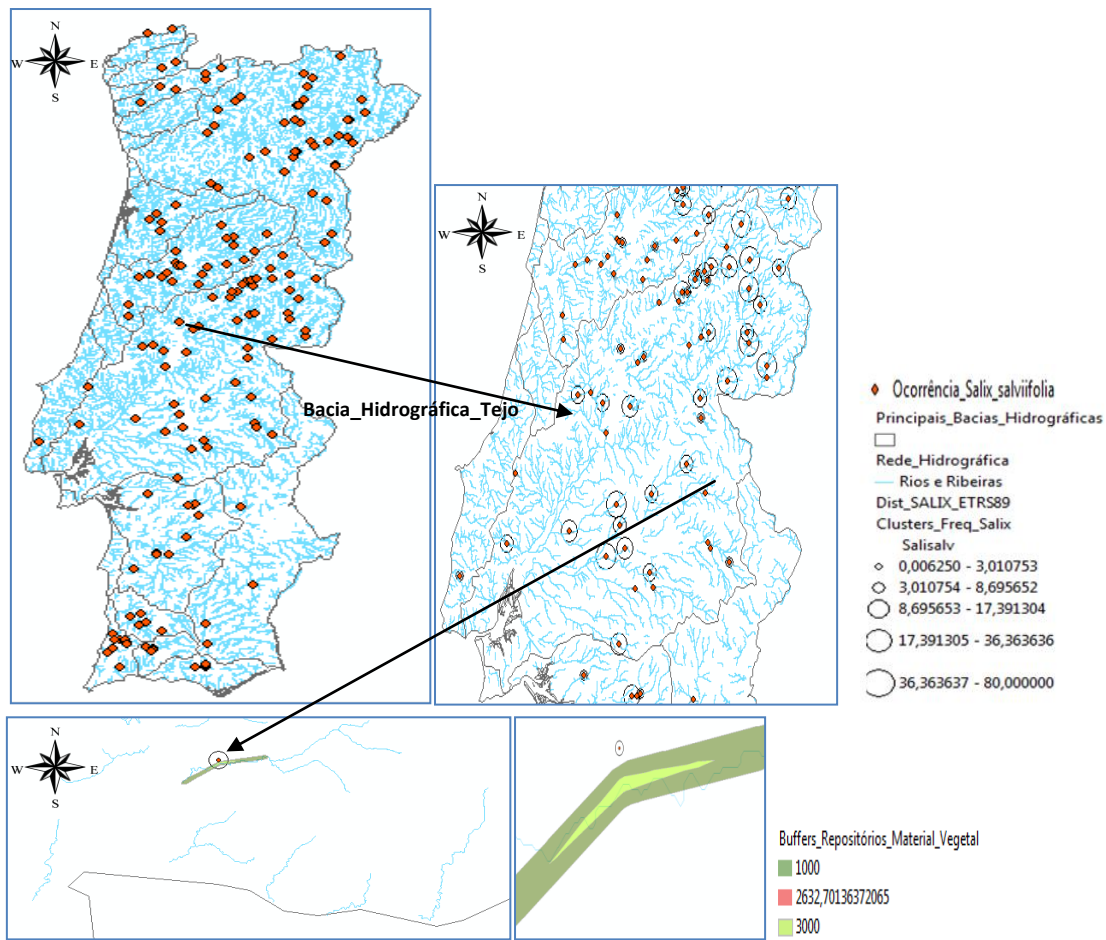


Figura 34 – Buffer 1000 m e 3000m em torno dos pontos de ocorrência da espécie *Salix salviifolia* (e.g de repositório de material vegetal na bacia hidrográfica do rio Tejo).

Bibliografia

ArcGis Desktop 9.3 Help: acedido em

<http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Tutorials>

Baldaia, S. (2012): “Estudo e Análise da Distribuição Espacial de Incêncios”. Trabalho realizado no âmbito da cadeira de análise espacial de sistemas de informação geográfica. Mestrado em SIG-Tecnologias e Aplicações. Faculdade de Ciências da Universidade de Lisboa.

Borges, S. (2013): Capítulo 2 do relatório de desenvolvimento- “Delimitação de Regiões de Proveniência para Espécies Ribeirinhas”. Instituto Superior de Agronomia, texto não publicado.

Câmara G.; Carvalho, M. S.; Cruz, O. G.; Correa, V. (2004): “Análise Espacial de Áreas. Análise Espacial de Dados Geográficos”, Brasília, EMBRAPA.

Catita, C. (2010): Apontamentos das Aulas de Análise Espacial de Informação Geográfica, texto não publicado.

Druck, S., Carvalho, M.S., Câmara, G., Monteiro, A.V.M. (2004): "Análise Espacial de Dados Geográficos", Brasília, EMBRAPA.

Paiva, J. e Rodriguez, A. e Correia, V. (2004): “Métodos Computacionais para Analisar Padrões de Pontos Espaciais”.

Fontes de dados:

- A informação disponibilizada na página do Instituto da Água, I.P. (INAG) refere-se aos dados recolhidos nas campanhas de amostragem que decorreram entre 2003 e 2006 no âmbito da implementação da Directiva Quadro da Água em Portugal. Estes dados poderão ser utilizados

livremente, sendo, contudo, necessário referir a origem dos dados. A informação encontra-se compilada num ficheiro WinZip® que contém 5 ficheiros Microsoft Office Excel® correspondentes às características gerais das estações de amostragem e aos dados recolhidos para cada elemento de qualidade biológica (fitobentos – diatomáceas, macrófitos, invertebrados bentónicos e fauna piscícola)

- Informação bioclimática produzida em formato GRID (ESRI) por Monteiro-Henriques (2010) e disponibilizada através do website: http://home.isa.utl.pt/~tmh/aboutme/Informacao_geografica.html.
- Informação Ambiental e climática do Atlas do Ambiente a uma escala original de 1:1.000.000

Software utilizado:

- *ArcGis Versão 9.3* [ESRI Inc., 1998-2009]
- *OpenGeoDa Versão 1.4.1* [GeoDa Center e filiados (R-Geo)]