

# INTEGRAÇÃO DE REDES NEURONAIS ARTIFICIAIS E SIG PARA A MODELAÇÃO DE HABITATS POTENCIAIS NO PARQUE NATURAL DAS SERRAS DE AIRE E CANDEEIROS

César Capinha<sup>1</sup>; Raquel Melo<sup>2</sup>; António Flor<sup>3</sup>

## Resumo

A Geografia e mais propriamente a Biogeografia têm desempenhado um papel de extrema importância na preservação da biodiversidade. Uma das principais ferramentas na definição de estratégias com esse fim tem sido a modelação de *habitats* potenciais.

Este trabalho apresenta uma metodologia de obtenção de habitats potenciais através da integração de modelos espaciais representativos de vários factores biofísicos com registos de ocorrência de três espécies de flora (variáveis dependentes): *Thymus villosus* L. sub. *villosus*; *Teucrium chamaedrys* L. e *Silene longicilia* (Brot.) Otth. A área de estudo corresponde ao Parque Natural das Serras de Aire e Candeeiros (PNSAC), tendo sido utilizada uma resolução mínima considerada de elevado detalhe (900 m<sup>2</sup>).

Para a caracterização biofísica da área foram modelados diversos factores ecológicos (variáveis independentes) como a radiação solar potencial, intensidade de vento, profundidade do horizonte A do solo, geologia do substrato, drenagem acumulada, e uso do solo. Como modelo estatístico de integração foram utilizadas regressões não paramétricas obtidas a partir de redes neuronais artificiais. Este é um método de utilização recente na área cujos valores de desempenho se têm revelado superiores à média em vários trabalhos.

A avaliação dos modelos foi realizada efectuando uma validação cruzada *K-fold*, em que as amostras iniciais foram particionadas em K amostras sendo apenas uma delas utilizada para validação e o processo repetido K vezes, permitiu atingir valores de desempenho entre os 74% e os 85%.

**Palavras-chave:** Modelação, habitats, redes neuronais artificiais, flora, factores ecológicos

---

<sup>1,2</sup> Departamento de Geografia da Faculdade de Letras da Universidade de Lisboa, Alameda da Universidade, 1600 – 214 Lisboa, Tel: +351 217910000,

e-mail: <sup>1</sup>cesarcapinha@hotmail.com; <sup>2</sup>raquel\_m@netcabo.pt

<sup>3</sup> Parque Natural das Serras de Aire e Candeeiros, e-mail: serradeaire@hotmail.com

## **I - Introdução**

Actualmente, na comunidade científica, é consenso geral que a biodiversidade está sob ameaça. Uma das principais causas, responsável pelo aumento percentual das espécies em risco, e também apontada como sendo a de maior impacte, é a destruição de *habitats*.

Contudo, a geografia não tem estado alheia à luta pela conservação da biodiversidade e uso sustentável dos recursos naturais. Sendo no espaço físico que se manifestam muitos dos impactes causados sobre os seres vivos, o conhecimento e o estudo biogeográfico têm desempenhado um papel de extrema importância na sua conservação. Este papel, desempenhado pela biogeografia, levou a uma definição recente do conceito de Biogeografia da Conservação (Whittaker *et al.*, 2005). Uma das principais ferramentas desta ciência em conjunto com a Macroecologia prende-se com a realização de modelos preditivos, com recurso a informação espacializada, da distribuição biogeográfica ou de *habitats* potenciais. A criação destes modelos preditivos de ocorrência ou adequação a uma determinada espécie ou comunidade permite, por um lado, identificar locais não inventariados onde estas possam ocorrer e, por outro, identificar áreas com condições favoráveis à sua existência, ainda que esta não se verifique. A utilidade destes resultados é óbvia, pois a obtenção de locais de adequabilidade potencial proporciona um conhecimento mais aprofundado da importância ecológica da área em estudo, tornando-se uma mais valia na planificação de estratégias de conservação, estudos de avaliação de impactos, planos de ordenamento, acções de re-introdução, entre outros.

De um modo simples, este tipo de modelos pode ser interpretado como a análise multivariada de dados espacializados, passíveis de esclarecer a relação existente entre a variável dependente (como por exemplo, presença de espécie ou comunidade), com uma ou mais variáveis independentes representativas de factores físicos ou bióticos com influência na sua distribuição.

Até muito recentemente, grande parte da realização destes modelos preditivos era aplicada a extensas áreas, com contrastes elevados e níveis de detalhe reduzidos. Actualmente, um dos desafios vigentes prende-se com a aplicação destes modelos a áreas cada vez mais reduzidas e com escalas de modelação mais detalhadas.

### **1 – Caracterização da área de estudo**

A área de estudo corresponde ao Parque Natural das Serras de Aire e Candeeiros (PNSAC) (39° 30' N; 8° 50' W). Os seus limites encerram, aproximadamente, uma área total de 389 km<sup>2</sup> nos quais se integram a totalidade da Serra dos Candeeiros, a Oeste, o Planalto de Santo

António, ao centro, uma pequena parte do Planalto de São Mamede, a Norte, e a Serra de Aire, a Nordeste. O substrato é predominantemente calcário sendo escassa a disponibilidade de água à superfície. Em termos climáticos, a área do PNSAC situa-se numa zona de transição entre o domínio Atlântico e o Mediterrâneo. Fitogeograficamente, encontram-se três situações clímax distintas: cercais, azinhais e sobreirais (Flor, 2005) correspondendo, grande parte da vegetação natural do parque, às suas etapas de degradação.

O PNSAC é uma área protegida, detentora de uma grande diversidade de fauna e flora, e de um elevado número de espécies com estatuto de protecção legal. Não obstante a sua importância ecológica, tem-se confrontado com diversas ameaças à sustentabilidade destas espécies, na sua maioria relacionadas com a extracção de inertes, a implantação de infra-estruturas de vários géneros ou a expansão urbana.

## **2 – Espécies em estudo**

Para a realização dos modelos de *habitat* potencial foram seleccionadas três espécies de flora existentes na área do PNSAC: *Thymus villosus* L. sub. *villosus*; *Teucrium chamaedrys* L. e *Silene longicilia* (Brot.) Otth.

A escolha destas espécies baseou-se em dois pressupostos fundamentais: um primeiro, ligado a um maior interesse na conservação ou conhecimento acerca da espécie, e outro, de carácter metodológico, relacionado com a existência de um número significativo de dados, indicadores da sua presença nesta área. Assim, esses dados de presença de cada espécie foram, posteriormente, completados com dados que indicam a sua ausência, obtidos através de levantamentos no terreno. Para a sua recolha foi utilizado um GPS, equipado com correcção diferencial WASS, que permite auferir um erro posicional máximo de, sensivelmente, 3 metros. Para a espécie *Thymus villosus* sub. *villosus* foram registados 23 locais de presença e 20 de ausência, para o *Teucrium chamaedrys* 22 locais de presença e 36 de ausência e para a *Silene longicilia* 95 locais de presença e 15 de ausência.

## **II - Introdução aos modelos independentes**

A realização de modelos de *habitat* potencial requer, na quase totalidade das abordagens possíveis, a existência de factores considerados explicativos da distribuição da espécie ou comunidade que se está a analisar. É a partir destes factores que a modelação vai interagir com os dados de presença da espécie, de modo a calibrar os resultados.

Guisan & Zimmermann (2000), citando Austin (1980 e 1985), referem a distinção entre os recursos e os factores directos e indirectos. Como recursos são designadas as variáveis que

representam elementos consumíveis pelas espécies, como os nutrientes, a água ou a luz para as plantas. Os factores directos são também variáveis biofísicas de importância fisiológica, mas que não são passíveis de ser consumidos, como o pH ou a intensidade do vento. Por último, consideram-se ainda as variáveis indirectas, que não representam directamente uma importância ecológica directa para as espécies ou comunidades, como é o caso do declive, da exposição e da altitude.

Do ponto de vista metodológico, para a sua aplicação em sistemas de informação geográfica, são apenas considerados dois grupos de variáveis. O primeiro grupo reúne as variáveis que determinam directamente a ocorrência ou não de determinada espécie, ou seja, os factores ecológicos. Os segundos correspondem a factores indirectos que apresentam, regra geral, uma maior precisão devido à exactidão dos seus processos de criação, correspondendo muitas vezes a uma derivação directa de modelos digitais do terreno (como o declive, a posição topográfica e a exposição), no entanto, não explicam de uma forma directa a ocorrência das espécies. Para a realização deste trabalho optou-se por escolher variáveis, na sua maioria, pertencentes ao primeiro tipo por diversas razões: a inclusão de factores ecológicos possibilita um melhor conhecimento da relação dos objectos de estudo (espécies) com o meio, os resultados obtidos são mais realistas e de interpretação directa e, por último, a construção destes factores em sistemas de informação geográfica é, em si mesmo, um objectivo para este trabalho, possibilitando a divulgação de metodologias para a realização de representações espaciais de alguns factores ecológicos mais importantes, alguns deles ainda pouco explorados nesta área. Esta escolha acarreta consigo a necessidade de que os modelos realizados possuam um rigoroso controlo do erro, de modo a evitar a sua propagação nos resultados finais. Além do reduzido erro, a sua definição espacial deverá ainda ser elevada, uma vez que a área a trabalhar (PNSAC) é de reduzidas dimensões no contexto de um trabalho deste tipo. Não obstante a primazia dada a estes factores, optou-se, também, por incluir dois outros modelos indirectos, representativos de factores ecológicos considerados importantes e, para os quais, não se conseguiu realizar modelos directos, ou com reduzidos valores de erro.

Odum (2004), refere os principais factores físicos limitantes para a regulação de populações, sobretudo produtores: a temperatura, a luz, a água, os gases atmosféricos, os nutrientes, as correntes e pressões, o solo e o fogo. A representação espacial de alguns destes factores é difícil, sobretudo para aqueles que apresentam uma relação difusa com as variáveis espaciais mais facilmente modeladas, como a altitude ou o declive. É neste contexto que os gases atmosféricos e os nutrientes são praticamente excluídos dos trabalhos de modelação de

*habitats*. O fogo não será também contemplado e, embora este tenha exercido uma forte influência sobre várias áreas do PNSAC, a construção de modelos de *habitat* potencial pressupõe que as populações se encontram em equilíbrio com o ecossistema, não sendo, desta forma, apropriado para populações que se encontrem em dinâmicas de crescimento ou regressão a curto prazo.

A água, embora constitua um factor limitante de elevada importância, é aqui representada por um factor indirecto, o modelo de fluxo acumulado de água. O desconhecimento de metodologias adequadas de modelação espacial directa desta variável, dentro das necessidades deste trabalho, fez com que se utilizasse este índice em sua substituição. Para tal, há ainda que considerar a elevada permeabilidade da maioria do substrato do PNSAC. Assim, reconhece-se, à partida, a possibilidade de imprecisão quantitativa deste modelo, no entanto, considerando-se uma permeabilidade uniforme pela área, pressupõe-se que a hierarquização espacial traduz a variabilidade relativa existente entre as células.

Nas correntes e pressões, para ecossistemas terrestres, o elemento mais significativo corresponde ao vento e à intensidade com que fustiga o terreno. Assim, para representar este factor, foi considerado um modelo de intensidade do vento. A temperatura e a luz, correlacionadas em termos de distribuição espacial, encontram-se traduzidas por um modelo de radiação solar potencial. Por fim o solo. Para a área de estudo, uma das melhores características encontradas para a sua representação, foi o seu grau de desenvolvimento (ou disponibilidade) que foi traduzido por um modelo de profundidade do horizonte A. A somar a esta caracterização, foi também considerada a inclusão do factor indirecto “substrato geológico”, pela importância que tem nos vários atributos dos solos. Além dos factores citados por Odum (2004), foi ainda considerado um modelo de uso do solo, pela implicação que os diversos usos têm na variação espacial, ao nível do solo, de muitos destes factores.

### **1 - Criação do modelo digital do terreno**

Muitas das variáveis utilizadas neste trabalho são derivadas de um modelo digital do terreno (*MDT*) da área do Parque, pelo que a sua correcta criação reveste-se de extrema importância, diminuindo a propagação de erros e permitindo obter resultados mais precisos, como tem sido posto em evidência por vários autores (Felicísimo, 1994 e Reis, 1996).

A criação do modelo digital do terreno foi efectuada com base na informação altimétrica vectorial de curvas de nível e pontos cotados, das cartas militares de Portugal, à escala 1: 25 000. Com esta informação, procedeu-se à criação de um modelo de rede triangular irregular (*TIN*), com uma margem envolvente à área do PNSAC.

## 2 – Modelo de radiação solar potencial

A construção de modelos de interação da radiação solar com a superfície terrestre assenta na geometria entre uma determinada posição do Sol e o modelo digital do terreno da área de interesse. Segundo Oke (1996), as principais medidas angulares que definem esta relação, são as seguintes:

$\phi$  – Latitude do local;  $Z$  – Ângulo Sol – Zénite;  $\delta$  – Declinação solar;  $h$  – Ângulo horário;  $\zeta$  – Ângulo do azimute solar

Considerando estas componentes, e através da lei do co-seno da iluminação e da equivalência trigonométrica,

$$\cos Z = \sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos h \quad (0.1)$$

pode-se calcular a quantidade de radiação ( $S$ ) que atinge um determinado local num determinado período de tempo, obtida através da fórmula

$$S = I_0 \cdot \cos Z \quad (0.2)$$

onde  $I_0$  representa a constante solar de  $1367 \text{ kWh/m}^2$ .

O valor desta radiação deve, ainda, entrar em conta com os efeitos de atenuação da atmosfera causados pela presença de partículas em suspensão e gases, e pela distância que os raios solares têm que percorrer, sendo esta distância variável de acordo com a altura do Sol em relação ao horizonte.

Por último, o valor obtido é válido apenas para uma superfície plana, não entrando em conta com a geometria do terreno.

A relação dos raios solares com a morfologia do terreno é dada pelo ângulo de incidência ( $\hat{\Theta}$ ) entre a normal da vertente e os raios solares, sendo traduzida pela equação

$$\cos \hat{\Theta} = \cos \hat{\beta} \cdot \cos Z + \sin \hat{\beta} \cdot \sin Z \cdot \cos(\Omega - \hat{\Omega}) \quad (0.3)$$

em que:

$Z$  – Ângulo zenital;  $\hat{\beta}$  – Ângulo da vertente (declive);  $\Omega$  – Ângulo do azimute solar;  $\hat{\Omega}$  – Ângulo azimutal da vertente (exposição);  $\hat{\Theta}$  – Ângulo de incidência entre a radiação e a normal à vertente.

O valor da radiação solar potencial para cada vertente é, então, dado pelo produto do ângulo de incidência ( $\cos \hat{\Theta}$ ) com a radiação potencial na normal do plano horizontal ( $S$ ).

No entanto, esta relação representa unicamente o efeito de sombra (quando ocorre) por auto-ocultamento, ou seja, o efeito de sombra gerado pela própria célula, não considerando o ocultamento com origem em unidades de relevo adjacentes (Felicísimo, 1994). A obtenção

das áreas de sombra pode ser facilmente realizada através de programas SIG que contenham módulos de cálculo de modelos de sombra, como é o caso do *Software* livre *SAGA GIS* (Conrad, 2003), que foi aqui utilizado.

### ***Cálculo do modelo***

A operação de cálculo do valor de radiação potencial, para o PNSAC, é realizada para cada uma das células que compõem o modelo de terreno, sendo ainda necessário integrar o modelo de declives e um modelo de exposição das vertentes, de modo a relacioná-la com a morfologia do terreno, como dado na equação do ângulo de incidência. No entanto, tal como evidenciado no ponto anterior, a quantidade de radiação solar para um local varia em termos temporais com o ângulo horário e com a declinação do Sol, ou seja, uma representação correcta da quantidade total anual de radiação potencial recebida teria que compreender a radiação de todos os dias do ano, por unidade temporal diária (por exemplo, hora a hora). Desta forma, quanto mais detalhada esta resolução temporal, mais preciso será o valor obtido, devido à variação do ângulo de incidência com a forma do terreno. Na impossibilidade técnica da realização de um número tão elevado de modelos e, tendo em conta que o interesse maior deste derivado se prende com a variação da distribuição espacial da radiação no terreno e não com os valores totais anuais, realizou-se a construção do modelo para o que podemos chamar de “dia médio”, ou seja, um dia de equinócio com valores horários acumulados da radiação potencial recebida nas 24 horas.

O cálculo foi realizado com recurso à programação de uma rotina no *Software ILWIS*, que aplica as equações anteriores ao valor de cada célula para cada meia – hora. Posteriormente, foram retirados os valores de células que os modelos automáticos do *SAGA GIS* (Conrad, 2003) indicam como estando à sombra naquele período de tempo. O modelo final corresponde ao somatório das 48 meias horas, convertidos para  $\text{kJ/m}^2$ .

Por se tratar de um modelo derivado de equações de carácter exacto, não se considerou necessário recorrer à sua validação.

O mapa resultante encontra-se na figura 1.

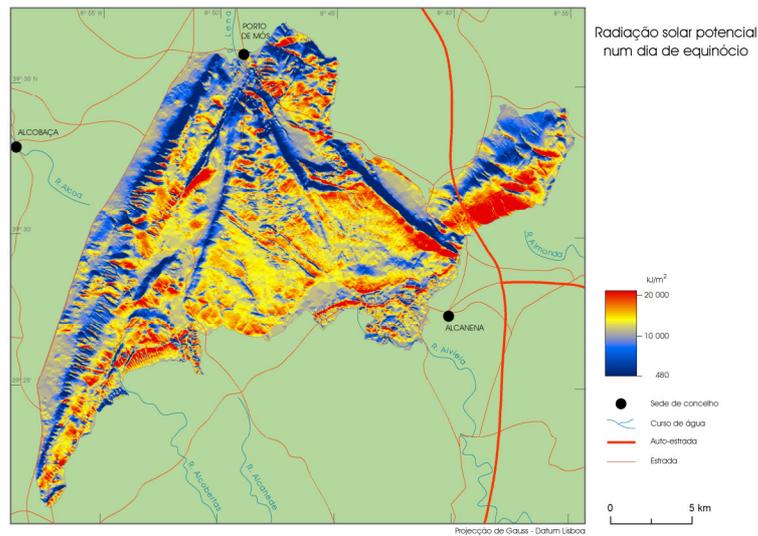


Figura 1 – Modelo da radiação solar potencial num dia de equinócio.

### 3 – Modelo de profundidade do horizonte A do solo

A formação do horizonte A do solo é um processo complexo onde interagem diversas variáveis com relações difusas. Como a modelação espacial da profundidade do horizonte A não dispõe de algoritmos específicos de cálculo, recorreu-se à utilização de dois métodos estatísticos, e um com base em conhecimento experimental. Cada uma destas abordagens foi utilizada para a realização de um modelo, sendo por fim utilizado o que proporcionou melhores resultados. Os modelos estatísticos escolhidos correspondem à regressão linear múltipla e à probabilidade *bayesiana*, esta última utilizada no sentido da probabilidade de cada célula pertencer a uma determinada classe de profundidade. Foi também realizado um outro, por combinação linear ponderada, com valores atribuídos numa base experimental. Estes métodos foram considerados, conceptualmente, os mais adequados em função do pretendido e da informação disponível.

Para a realização que qualquer um destes modelos foi necessário recolher, inicialmente, uma amostra de profundidades deste horizonte na área em estudo. No total foram efectuados 78 levantamentos, recorrendo a um método de amostragem aleatória estratificada, considerando a variação das diversas variáveis categóricas adiante descritas. A realização deste tipo de amostragem para variáveis contínuas é, obviamente, impraticável, tendo-se neste caso optado por efectuar levantamentos em áreas o mais contrastadas possível.

Do total de pontos amostrados foram extraídos, aleatoriamente, 12 ( $\approx 15\%$ ) para validação dos modelos, sendo estes calibrados apenas com os restantes 66 ( $\approx 85\%$ ).

A escolha das variáveis independentes para o cálculo de um modelo deste género requer, à partida, algum conhecimento dos factores que influenciam o desenvolvimento do solo. As

variáveis consideradas baseiam-se no trabalho de Costa (2004), que refere como principais factores de formação do solo: o clima, os organismos, a rocha-mãe, o relevo e o tempo.

O factor climático não é considerado, uma vez que se admite que a área de estudo não apresenta contrastes suficientemente significativos para o processo de formação do solo. Também o factor tempo não foi considerado por não ser espacialmente representável, pelo menos considerando um modelo estático, como se pretende. Deste modo foram apenas considerados os restantes factores. Os organismos (neste caso apenas macroscópicos) são representados por um índice de vegetação, a rocha-mãe pela geologia do substrato e o relevo por alguns modelos representativos da sua complexidade. A geologia foi obtida a partir da vectorização das Cartas Geológicas de Portugal 26-B, 26-D, 27-A e 27-C, à escala 1: 50 000. No que respeita aos índices topográficos foram considerados o declive, o índice CTI (*Compound Topographical Index*) e o índice STI (*Sediment Transport Index*) (Hengl *et al.*, 2003), assim como um modelo de áreas planas classificadas em topo, patamar e base. Este modelo foi obtido através de um algoritmo desenvolvido para o presente trabalho, e tendo como base a comparação de valores de altitude das últimas células, consideradas planas, com as células limítrofes. Por fim, como representação da influência biológica, considerou-se um índice NDVI (*Normalized Difference Vegetation Index*) obtido a partir de bandas LANDSAT.

### ***Modelação com recurso à regressão linear múltipla***

O modelo de regressão linear múltipla apresenta, à partida, algumas limitações para o presente estudo. Com efeito, a integração de variáveis categóricas neste modelo é limitada, sendo apenas passível de realizar através da sua simplificação para variáveis binomiais. Assim, a integração da informação foi realizada através de uma regressão passo a passo “forward”, pois, com este processo, o modelo além de recusar as variáveis que não contribuem para melhorar significativamente as previsões da variável dependente, comunica também o contributo de cada uma das aceites.

A aplicação do modelo atingiu um coeficiente de determinação ( $R^2$ ) de 70,9%. Neste modelo apenas foram consideradas as variáveis com um grau de confiança de 95% ( $p < 0,05$ ) e, por conseguinte, somente integraram o modelo o “CTI” ( $p=0,0$ ), as “Bases” ( $p < 0,0$ ), o “Declive” ( $p= 0,03$ ) e os “Tops” ( $p= 0,048$ ). Entre os resultados do modelo constam, também, os valores dos parâmetros que expressam a equação da regressão 1.4

$$\text{Horizonte A} = 13,8 + 13,9 \times \text{Bases} - 5,21 \times \text{Tops} - 0,219 \times \text{Declive} + 1,16 \times \text{CTI} \quad (0.4)$$

Esta equação foi executada utilizando os mapas como variáveis. O mapa gerado é visível na figura 2.



**Figura 2 – Modelação da profundidade do horizonte A do solo com recurso à regressão linear múltipla.**

### ***Modelação com recurso à análise ponderada de factores***

Este método assenta, essencialmente, na atribuição de valores a cada classe dentro de cada variável e, por fim, entre as várias variáveis. O modelo resultante terá os seus valores mais elevados para os locais considerados com maior grau de desenvolvimento. Estes quantitativos foram atribuídos, tendo por base algum conhecimento de especialistas, e os valores de coeficiente de correlação de Pearson, entre os valores de profundidade amostrados e as diversas variáveis consideradas. Para a execução deste modelo foram apenas consideradas as variáveis contínuas que apresentaram coeficientes de correlação considerados significativos (declives e CTI) e as restantes variáveis categóricas iniciais (geologia e áreas planas). Estes modelos foram escalonados de acordo com a interpretação de vários resultados estatísticos entre estes e as profundidades medidas no terreno e por fim agregados e re-escalonados entre si para a obtenção do modelo final.

O modelo resultante foi reescalonado linearmente, atribuindo-se o valor máximo registado de profundidade (40cm) ao valor máximo do escalonamento, e o mesmo para o valor inferior (1cm).

### ***Modelação com recurso à probabilidade bayesiana***

A metodologia de implementação com recurso à probabilidade *bayesiana* teve por base o trabalho de Reis *et al.* (2003), e compreende quatro fases antes da integração: cálculo das probabilidades à priori de encontrar cada classe numa variável, probabilidade conjunta de encontrar as classes das várias variáveis e a probabilidade condicionada, ou seja, a

probabilidade de encontrar cada um dos eventos (classe de profundidade de horizonte A) associado a cada classe de cada variável e o cálculo da probabilidade de encontrar cada um dos eventos na área de estudo.

A utilização deste método requer, à partida, que todos os temas integrantes sejam categóricos, incluindo a variável dependente. Com base na interpretação de um histograma de pontos e nos objectivos a que o mapa final se destina, foram efectuados três mapas matriciais de eventos, cada um contendo uma das seguintes classes: 0 – 7 cm; >7 – 19 cm; >19 – 40 cm.

Tal como na combinação ponderada, também os modelos considerados para integrar esta análise probabilística, foram os que apresentaram coeficientes de correlação considerados significativos com a variável dependente (declives e CTI) e restantes variáveis categóricas iniciais (geologia e áreas planas).

Uma vez criados todos os temas necessários, foi então efectuada a integração da informação para cada um dos três temas, através da fórmula 1.5, adaptada da teoria da probabilidade de *Bayes*.

$$P_{\text{Classe}} = \frac{\prod p_{xi} \cdot c_{xi}}{p_y^{n-1} \cdot C_{jxi}} \quad (0.5)$$

Sendo:  $p_{xi}$  a probabilidade à priori de cada variável independente;  $c_{xi}$  a probabilidade condicionada de cada variável independente;  $p_y$  a probabilidade de encontrar cada um dos eventos na área de estudo; e  $C_{jxi}$  a probabilidade conjunta.

Os temas com probabilidades de 0 a 100 foram congregados num único, que atribui o valor da classe com maior valor de probabilidade de ocorrer.

### ***Avaliação e validação dos modelos***

Para avaliação do desempenho de cada um destes modelos foi então aplicada uma validação cruzada com os 15% de pontos amostrados não incluídos na calibração dos modelos. Para esta avaliação foram calculadas, para os modelos contínuos, várias estatísticas básicas dos valores de erro, como o seu desvio-padrão e o erro médio, de acordo com a fórmula 1.6

$$E.M. = \sum |X_i - Y_i| / n \quad (0.6)$$

Para o modelo categórico, derivado da probabilidade de *Bayes*, foi efectuada uma comparação directa entre as classificações obtidas e os valores dos pontos de validação.

**Quadro I – Erro médio e estatísticas sumárias (cm) dos valores de erro do modelo originado por regressão linear múltipla e ponderação de factores**

	<i>n</i>	E. M.	Desvio-padrão	Mínimo	Máximo	Mediana
Regressão linear	12	4,2	2,9	0	8,2	5
Ponderação de factores	12	6,5	5,1	1	18,44	4,9

**Quadro II - Percentagens de validação das classes obtidas com modelação por probabilidades de Bayes**

<i>n</i>	% de classe correcta	% de classe incorrecta	% de classe incorrecta que pertencem à classe imediatamente seguinte	% de classe incorrecta que não pertencem à classe imediatamente seguinte
12	66,6	33,3	75,3	24,7

O modelo que, à partida, apresentou os melhores resultados foi o derivado da regressão linear múltipla. O erro médio apresentado foi menor que o do método de modelação por ponderação de factores (-2,3 cm), o desvio-padrão foi menor (-2,2 cm) e o erro máximo foi também bastante inferior (-10,2 cm). O modelo derivado das probabilidades de Bayes também não atingiu um grau de desempenho muito elevado (66%), muito possivelmente devido à necessidade de desagregação da variável dependente em três sub-amostras.

Considerando-se estes resultados, a modelação de *habitats* irá integrar o mapa obtido por regressão linear múltipla.

#### **4 - Modelo da geologia do substrato**

A geologia do substrato, surge também como uma das variáveis que melhor pode ajudar à diferenciação das propriedades do solo, complementando o modelo anterior. A sua obtenção foi efectuada a partir da digitalização das cartas geológicas de Portugal 26-B, 26-D, 27-A e 27-C, à escala 1: 50 000. Apesar desta escala ser menor que a restante informação integrada neste trabalho, o peso que empiricamente esta variável aparenta representar e o facto de muitas das formações geológicas aqui presentes apresentarem áreas de grandes dimensões originando erros cartográficos consideráveis apenas nos seus limites, fez com que a sua integração como modelo independente fosse considerada adequada.

As classes de geologia das formações foram amplamente simplificadas, tendo-se optado por manter unicamente a referência à designação mais geral de cada classe indicada nas cartas.

Assim foram individualizadas as seguintes classes: Calcários; Calcários e margas; Detritos; Aluviões; Arenitos; Rochas vulcânicas e Grés.

## 5 - Modelação da intensidade do vento

A modelação, quer da intensidade do vento quer da sua direcção à escala local, tem sido uma das áreas menos exploradas e menos acessíveis no âmbito da modelação espacial, particularmente no que diz respeito à sua integração com Sistemas de Informação Geográfica. Apesar da existência de diversos modelos exteriores aos SIG, a especificidade da área de estudo, dotada de uma topografia complexa, e o elevado detalhe requerido, fez recair as possibilidades de escolha em sistemas ainda pouco desenvolvidos e que requerem um elevado esforço de configuração.

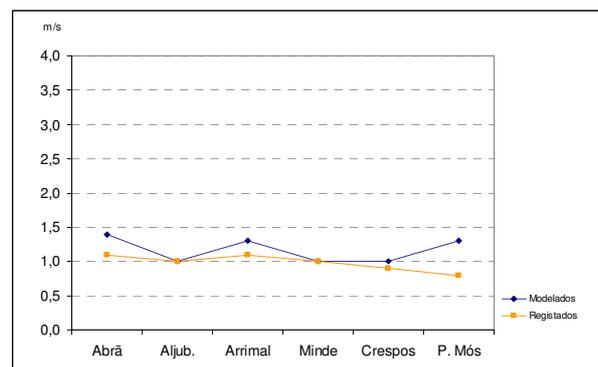
Após uma pesquisa acerca das possibilidades de modelação e da sua adequação à área de estudo, optou-se pela utilização do algoritmo em desenvolvimento pelo *United States Forests Service (USFS)*, denominado *WindWizard*, sob o regime de utilizadores de teste (*beta testers*). Este ambiente de modelação assenta nos princípios da dinâmica de fluidos, pelo que aplica as equações de *Navier-Stokes*, para descrever o fluxo do vento ao longo das variações do terreno. A turbulência, sendo a principal debilidade das equações referidas, é modelada com a aproximação  $k-\epsilon$  (Butler *et al.*, 2006).

A aplicação destas equações implementadas no *WindWizard* computa células tridimensionais de dimensão variável sobre o modelo de altitudes, interpretando para cada uma o fluxo e a intensidade, de acordo com os valores das células vizinhas. A sua aplicação sobre a área de estudo requer vários parâmetros de entrada, sendo necessário fornecer, logo à partida, um modelo digital do terreno em estrutura matricial, pelo que se forneceu um modelo matricial com resolução de célula de 30 metros. Neste modelo digital do terreno foi incluída uma área envolvente de 5 km em redor dos limites do PNSAC, de modo a entrar em conta com o efeito da orografia circundante. Relativamente à representação do terreno é ainda necessário indicar a sua rugosidade dominante. A rugosidade corresponde ao nível de atrito e tipo de obstáculos que o fluxo de ar encontra no seu percurso. Assim, e considerando o domínio arbustivo existente na área de estudo, foi indicada uma rugosidade média adequada a este tipo de cobertura. Os parâmetros da velocidade e orientação do fluxo de vento, também necessários como parâmetro de entrada, foram obtidos com base nas aero-sondagens de Lisboa-Portela. Apesar do afastamento geográfico entre os dois locais, o *WindWizard* assenta a sua modelação em valores registados em altitude, de preferência ligeiramente acima do ponto mais elevado que se encontre na área, que é de 678 m na Serra de Aire. Assim, as leituras

foram efectuadas aos 925 hPa que, com a redução da atracção gravitacional da terra, corresponde aproximadamente aos 800 metros de altitude. As leituras realizadas a 925 hPa, ainda que com algum afastamento da área de estudo, mantêm-se fiáveis, pois à altitude que este nível de pressão se reporta não são sentidos, para esta área, os efeitos da orografia que caracterizam as medições das estações meteorológicas à superfície. Assim, foram utilizados os registos gratuitos das aero-sondagens de Lisboa-Portela, compreendendo um período disponível de 7 anos, de 1999 a 2005, com interrupções. Os valores de velocidade e orientação aos 925 hPa foram agrupados e separados por octante de proveniência, tendo sido calculada a frequência de cada um deles ao longo do ano, bem como a sua velocidade média. Na modelação foi então introduzido o valor central de cada octante, em graus, e a velocidade média correspondente, tendo sido efectuado, a *posteriori*, a ponderação final dos modelos.

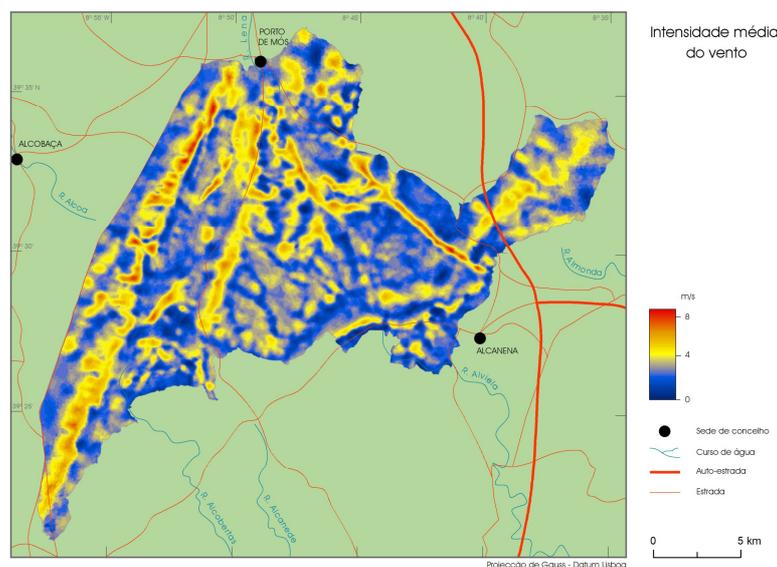
### **Validação do modelo**

A validação deste modelo está condicionada pelos dados das estações meteorológicas existentes na área. Dentro dos limites do PNSAC existem duas estações meteorológicas automáticas, uma localizada a Norte de Minde e a outra no Arrimal, junto à Lagoa Pequena. Para a validação, além destas, foram ainda consideradas outras quatro estações, situadas fora dos limites do PNSAC, mas consideravelmente afastadas dos limites do modelo de altitudes que serviu de base à modelação do vento. As quatro estações automáticas exteriores ao PNSAC são as de Abrã, Porto de Mós, Aljubarrota e Crespos. Os registos destas estações não vão além de 3-4 anos, o que de certo modo impossibilita a avaliação do modelo total. Assim, e tendo em conta este factor limitante, optou-se por avaliar um modelo por estação que entra em conta unicamente com o período para o qual esta possui valores registados, ou seja, procedeu-se à criação de um modelo para cada estação com os valores em altitude (Lisboa-Portela) ponderados pelo período de tempo que a estação meteorológica disponibiliza. Após a construção ponderada de cada um destes modelos, procedeu-se à comparação dos seus valores com os valores registados nas estações, visível na figura 3.



**Figura 3 – Comparação entre os valores de intensidade do vento modelados e registados.**

Como se pode verificar, os resultados modelados aproximam-se bastante dos valores registados pelas estações, sendo a diferença máxima de 0,6 m/s na estação de Porto de Mós. Com uma amplitude de variação total de 8,7 m/s, um erro máximo registado na ordem dos 7% é considerado bastante adequado para o fim a que se propõe o modelo. Nota-se, no entanto, uma tendência para a modelação de valores mais elevados que, efectivamente, os registados. Tal facto poderá dever-se à existência de barreiras à progressão do vento, que não estejam contempladas no modelo digital do terreno, como habitações ou vegetação de grande porte. O modelo resultante encontra-se na figura 4.



**Figura 4 – Modelo da intensidade média do vento.**

## 6 - Modelo de fluxo acumulado de água

Como referido na introdução deste capítulo, a variável água é traduzida apenas por um modelo indirecto, derivado do modelo digital do terreno. Este modelo corresponde à representação quantitativa do fluxo acumulado, que passa por cada uma das células, como resultado da drenagem das células a montante.

Este modelo traduz o somatório do número de células das quais recebe fluxo, acrescido do seu próprio valor (1). Desta forma é possível hierarquizar as células de acordo com a maior ou menor quantidade de água recebida. Esta hierarquização foi realizada sem que se tivesse efectuado a remoção de áreas que funcionam como “poço”, ou seja, áreas que não drenam para o exterior, resultantes da sua natureza cársica e da existência de uma elevada quantidade de depressões. O cálculo foi realizado directamente com recurso ao módulo de hidroprocessamento do *Software ILWIS*. Após o seu cálculo, foi necessário efectuado em

escalonamento de 0 a 100 dos seus valores originais, devido aos valores de acumulação extremamente elevados que estes modelos por vezes atingem.

## **7 - Modelo do uso do solo**

A informação utilizada para este processo corresponde às bandas espectrais captadas pelo sensor *Enhanced Thematic Mapper plus* (ETM+) do satélite Landsat 7, para o ano de 2001, e com uma resolução espacial de 30 metros.

Antes de se proceder a qualquer processo de extracção de informação útil, foi necessário realizar um conjunto de procedimentos com vista à correcção de eventuais erros decorrentes do processo de aquisição de informação. Estes procedimentos incluem a georreferenciação ou correcção geométrica e a correcção radiométrica da imagem.

Para a obtenção deste modelo, optou-se por realizar uma classificação supervisionada, a qual se baseia num conhecimento prévio da área em estudo. Este conhecimento prévio resultou de levantamentos do uso de solo na área de estudo. No total foram efectuados 46 levantamentos, correspondentes a diferentes usos do solo, dos quais foram seleccionados, com base num método de amostragem aleatória estratificada, 27 pontos para a realização do modelo e 19 para a validação do mesmo.

No presente trabalho, as áreas de teste necessárias para este método foram delimitadas como base na agregação das bandas espectrais 3, 2 e 1, da qual resulta uma composição colorida correspondente à cor real, e nos 27 pontos de amostragem destinados à realização do modelo que, por conseguinte, correspondem a diferentes usos do solo. A individualização das classes que irão constar no mapa foi realizada com base em observações no terreno, nas quais se constatou 7 classes predominantes: misto de rocha e solo nu com vegetação dispersa, terreno agrícola, vegetação herbácea, vegetação arbustiva, vegetação arbórea, pedreira, e em informação já existente, no caso da área edificada. Dadas as dificuldades que surgiram na distinção entre a vegetação herbácea e a arbustiva, optou-se por integrar estes dois usos na mesma classe. Estas áreas de teste são, então, definidas como um ficheiro vectorial de assinaturas, ao qual se atribui o nome de cada classe, sendo este ficheiro de assinaturas processado sobre as 7 bandas.

Para a realização da classificação utilizou-se o classificador de máxima verosimilhança, por ser um dos métodos de classificação mais utilizados e considerado um dos que apresenta melhores resultados (por ex. Misáková (2007)).

Este método pressupõe o cálculo do vector médio e da matriz de co-variância das respostas espectrais de cada classe. Ao assumir que os níveis radiométricos de cada amostra se

encontram próximos da distribuição normal (Gaussiana), dada através do vector médio da classe, a atribuição do pixel é feita à classe segundo a qual ele terá mais probabilidade de pertencer. O mapa resultante pode ser observado na figura 5.

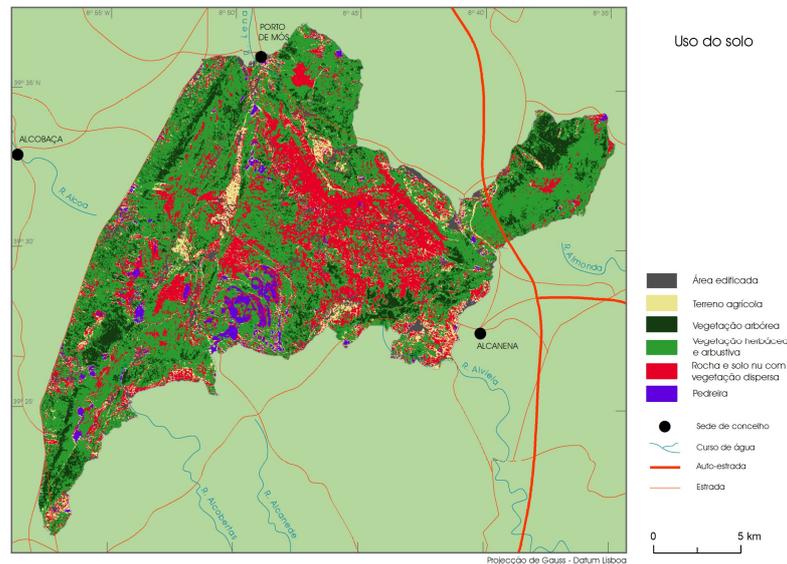


Figura 5 – Mapa do uso do solo.

### **Validação do modelo**

Para a validação do modelo realizou-se uma sobreposição entre o mapa originado anteriormente e os pontos amostrados no terreno que se encontravam destinados a este fim. Assim, foram calculadas as percentagens relativas aos pontos cujas classes foram correctamente atribuídas e, conseqüentemente, foram também calculadas as percentagens de carácter erróneo. A precisão obtida foi de 89,5%.

### **III - Introdução aos métodos de modelação de *habitats* potenciais**

Definido em termos gerais, um modelo de *habitat* potencial representa a relação entre as espécies e os factores que se considera controlarem a sua existência. A avaliação desta relação pode ser efectuada de várias formas e deve ter em consideração, sobretudo, o tipo de dados existente e o grau de conhecimento acerca das relações da espécie com o meio.

No que diz respeito aos dados são, sobretudo, os de carácter dependente que ditam a natureza da formulação a utilizar. Estes podem-se referir a levantamentos de presença-ausência, abundância, ou taxas demográficas. O tipo mais comum de dados relaciona-se com a presença-ausência, sendo muitas vezes pautado apenas por dados de presença, obtidos, por exemplo, em herbários. Esta condição dos dados impossibilita a utilização de muitos dos métodos estatísticos utilizados para modelação de *habitats*, aspecto em relação ao qual este

trabalho se precaveu, através do levantamento de diversas localizações onde nunca se verificou a ocorrência das espécies de interesse.

O tipo de relação existente entre a espécie e os factores do meio também deve ser tido em conta. As relações mais conhecidas são as lineares, que correspondem, por exemplo, a situações de aumento do número de indivíduos proporcionalmente ao aumento da disponibilidade de água. Esta relação, na maioria das vezes, tende a ser válida apenas para intervalos de variação curtos. Assim, tal como descrito pela lei da tolerância de Shelford, o contínuo aumento da disponibilidade de água faria atingir um limiar a partir do qual esta inverteria a tendência de crescimento, originando uma relação em forma de “sino” (relação Gaussiana). Apesar de se possuir algum conhecimento sobre este tipo de relações estabelecidas entre as diversas espécies e alguns factores, este não é de todo completo. McCune (2006) refere que, apesar do papel central destas relações na modelação de *habitats* potenciais, o nosso conhecimento relativamente a elas é ainda primitivo, sobretudo quando são considerados vários factores em simultâneo.

O método de integração da informação considerando os aspectos anteriormente descritos pode, ainda assim, assumir diversas formas. A sua escolha, segundo Guisan e Zimmerman, (2000), corresponde à escolha de um algoritmo apropriado ao contexto da modelação e para previsão de um tipo particular de variável de resposta, de estimação de coeficientes. Os mesmos autores referem sete categorias de modelos estatísticos possíveis de utilizar, os quais, na nossa opinião, deverão ainda ser acompanhados de métodos baseados em conhecimento de especialistas através de ponderações directas. As categorias compreendem: regressões lineares e não lineares paramétricas e não paramétricas; métodos de classificação; modelos de medição de similaridade; técnicas de ordenação; modelos de probabilidade de Bayes; redes neuronais e por último uma categoria que conjuga várias destas técnicas. A primeira categoria compreende os métodos mais comumente utilizados, como a regressão linear múltipla ou a regressão logística. Paralelamente a estas, ocorre também a regressão aditiva generalizada (*GAM*), um método não paramétrico surgido recentemente nesta área, com resultados geralmente superiores aos modelos anteriores (Segurado e Araújo, 2004). Os métodos de classificação compreendem técnicas como as árvores de decisão ou técnicas de análise de semelhança. Os métodos de medição de similaridade, baseiam-se na construção de índices de semelhança dos dados de presença existentes e, como tal, dispensam dados de ausência. Também as técnicas de ordenação utilizam somente os dados de presença e compreendem modelos, como a análise factorial do nicho ecológico (*ENFA*), que é utilizada, por exemplo, em modelos como o *BIOMAPPER*. A quinta categoria compreende a probabilidade de Bayes,

já descrita na modelação da profundidade do horizonte A do solo. As redes neuronais, incluídos na penúltima categoria, correspondem a modelos não paramétricos, conceptualmente semelhantes à regressão aditiva generalizada, e que podem ser utilizadas tanto em classificação como para regressão.

Segurado e Araújo (2004), que efectuam uma comparação do desempenho de alguns destes métodos para modelação de habitats, referem ainda os interpoladores espaciais, onde se pode referir a interpolação linear simples ou a *krigagem*. Estes autores, utilizando um volume elevado de dados de presença e ausência de espécies, efectuaram uma comparação das performances de cada uma destas técnicas. Os resultados obtidos neste trabalho indicam um melhor desempenho geral das redes neuronais. Guisan e Zimmerman (2000) referem que a utilização de redes neuronais é uma área promissora na modelação de *habitats*, apesar do ainda reduzido número de exemplos existentes. Inversamente, as técnicas com resultados inferiores, corresponderam às que utilizam apenas dados de presença, ou seja, os índices de semelhança e as técnicas de ordenação. Entre estes níveis de desempenho situaram-se as restantes técnicas, com destaque pela positiva para a GAM.

Considerando os resultados obtidos por estes vários trabalhos, optou-se pela utilização das redes neuronais como método estatístico de integração. Este método, pela sua natureza não paramétrica, apresenta a vantagem de não assumir nenhuma forma para a relação entre as espécies e os factores considerados, o que se coaduna com o grau de conhecimento das espécies deste trabalho.

### **1 - Introdução às redes neuronais artificiais como método de regressão não paramétrico**

As redes neuronais artificiais podem ser entendidas como modelos estatísticos adaptativos, baseados na estrutura de funcionamento do cérebro (Abdi, 2003). Dizem-se adaptativos porque conseguem aprender a estimar os parâmetros de uma população, utilizando um pequeno número de exemplares desta. A sua estrutura é basicamente constituída por unidades simples, designadas de neurónios, por analogia com as unidades cerebrais, sendo que estas unidades estão ligadas por um conjunto de arcos com ponderações associadas. O processo de aprendizagem corresponde, de um modo geral, à simples modificação dos valores de ponderação atribuídos a cada arco que liga os neurónios.

As redes neuronais artificiais estruturam comumente as suas unidades em camadas (*layers*). A primeira é a camada de entrada (*input layer*) e a última a de saída (*output layer*). As camadas intermédias, quando existentes, são designadas de camadas escondidas (*hidden layers*). A informação a ser analisada pela rede dá, então, entrada pelos neurónios da camada

de entrada e é propagada através dos arcos para a camada seguinte. Posteriormente, o resultado deste processamento é propagado para os neurónios da camada seguinte, e assim por diante até que seja atingida a camada de saída. Este processamento em rede tem como objectivo apreender as associações existentes entre os padrões das variáveis consideradas não dependentes (de entrada) e as dependentes (de saída). Este processo de aprendizagem é realizado através da alteração dos valores de ponderação (por atenuação ou amplificação), atribuídos aos arcos que ligam cada neurónio, em função da performance obtida.

A aprendizagem, tal como já foi referido, é definida pelo valor de ponderação atribuído a cada arco, podendo as suas alterações de valor serem separadas em processos de aprendizagem supervisionada e não supervisionada. A aprendizagem supervisionada, adoptada neste estudo, ocorre quando os valores de saída são conhecidos e utilizados, e se entra em linha de conta com os sinais de erro derivados da ponderação.

O modelo utilizado para efectuar esta investigação apresenta uma estrutura de múltipla camada perceptrão (*multilayer perceptron*) e, para o efeito, recorreu-se ao *software Tiberius Data Mining*. Esta estrutura, sendo uma das mais populares, é constituída por uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Com excepção da camada de entrada, todas as outras são constituídas por neurónios que apresentam capacidade de aprendizagem. Estas redes são do tipo progressivo, ou seja, o fluxo processa-se unicamente no sentido da saída de uma camada, para a entrada da camada seguinte, sem que ocorram processos de retrocesso. A camada de entrada integra as variáveis consideradas condicionantes dos resultados da camada de saída. Assim, a camada de entrada vai corresponder às variáveis independentes e a camada de saída à presença ou ausência das espécies. Deste modo, é ainda necessário configurar o número de camadas escondidas e o número de neurónios presente em cada uma destas camadas.

Após a configuração da rede, é indicado o tipo de aprendizagem que esta vai efectuar. O sistema *Tiberius* utiliza a retropropagação do erro, correspondendo este processo a um método de aprendizagem por correcção do mesmo. A forma de aprendizagem realiza-se em duas fases: inicialmente é emitido um valor de teste, no sentido directo (das camadas de entrada para a de saída), em que as ligações das camadas apresentam um valor fixo. Posteriormente, ao atingir a camada de saída, é gerado um sinal de erro através da subtracção dos valores chegados com os encontrados na camada de saída. Por conseguinte, este sinal de erro obtido é propagado de volta através dos mesmos neurónios, sendo os pesos das ligações ajustados de forma a aproximar, o mais possível, os valores obtidos com os da camada de saída.

Assim, a utilização de redes neuronais é, sobretudo, indicada nos casos em que não é conhecida a natureza exacta das relações existentes entre os *inputs* e os *outputs* ou quando os sistemas são demasiado complexos, para que se conheçam relações directas ou lineares, como acontece nas relações entre as espécies consideradas e os restantes factores.

## **2 – Modelação do *habitat* potencial para as espécies de estudo**

A realização dos modelos de habitat potencial assenta, inicialmente, na intersecção entre os pontos de presença e ausência de determinada espécie, e os valores obtidos, através dos modelos espaciais das variáveis independentes, nessa mesma localização.

Os resultados adquiridos nesta operação são agrupados numa tabela única, que passa a conter a indicação de presença ou ausência, e correspondentes valores das variáveis consideradas para a construção dos modelos de *habitat* potencial.

Antes de se proceder à construção da rede neuronal artificial é necessário converter as variáveis categóricas em variáveis binomiais e proceder à sua simplificação. O elevado número de classes que estas apresentam, torna necessária a sua simplificação, de forma a reduzir o número de variáveis integrantes da regressão. Este processo permite, também, reduzir a influência de classes para as quais não se conhecem registos acerca da existência ou não da espécie.

A natureza heurística das redes neuronais artificiais faz com que os seus resultados finais apresentem ligeiras diferenças. Este aspecto implica a realização de mais de um modelo, sendo o final obtido através de uma média aritmética simples. Assim, foram realizados cinco modelos preliminares, com vista à obtenção de cada um dos modelos finais.

O método de aprendizagem escolhido, para o treino das redes, corresponde à tentativa de redução do erro quadrático médio, sendo cada um dos cinco modelos treinados por um período de cerca de 1 hora. As equações que traduzem as regressões são aplicadas, directamente, aos modelos de variáveis independentes. Os seus valores resultantes vão, tendencialmente, aproximar-se do valor de presença (1) ou de ausência (0). No entanto, o intervalo de variação pode percorrer toda a amplitude que separa estes valores ou até ultrapassá-la. Para uniformizar os valores finais, tornou-se necessário proceder a um escalonamento linear simples, baseado nos valores máximo e mínimo modelados para um intervalo de probabilidade de ocorrência de 0 a 100. Os modelos resultantes são visíveis nas figuras 6, 7 e 8.

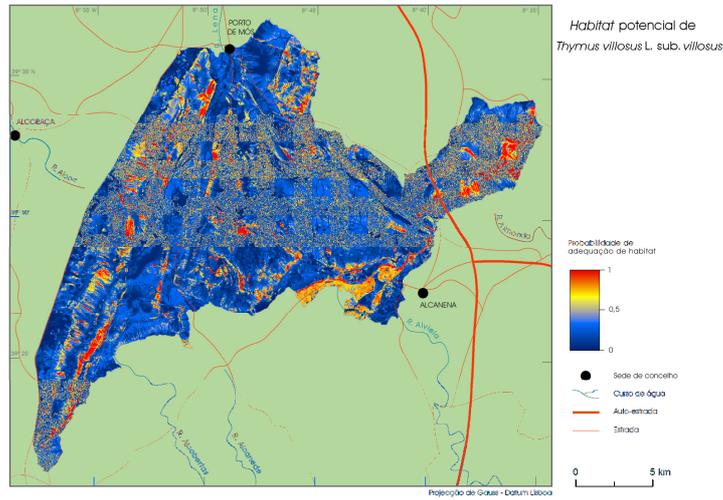


Figura 6 – Habitat potencial de *Thymus villosus* L. sub. *villosus*.

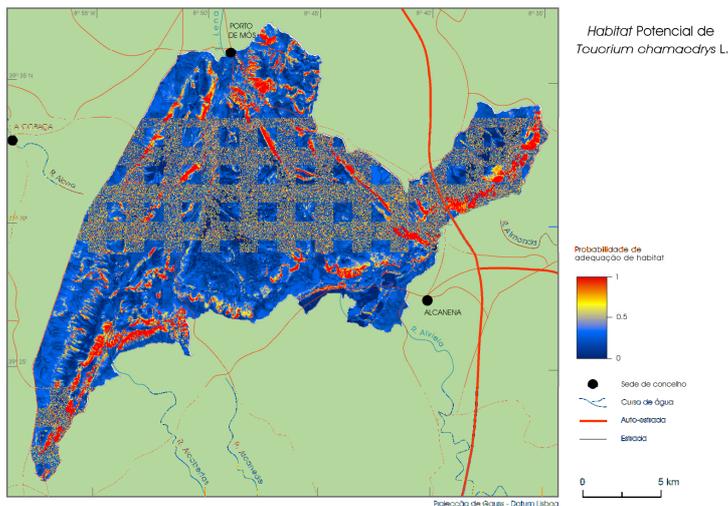


Figura 7 – Habitat potencial de *Teucrium chamaedrys* L.

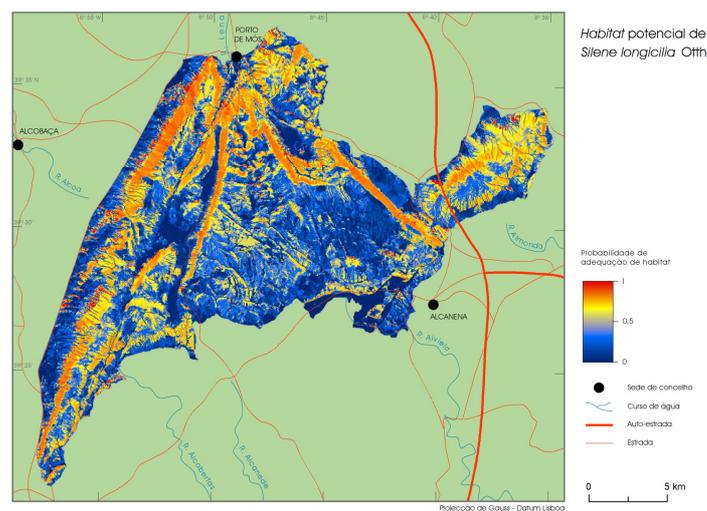


Figura 8 – Habitat potencial de *Silene longicilia* Otth.

### Validação dos modelos

A validação destes modelos foi realizada de forma a minimizar a necessidade de se extrair dados da amostra que serve de calibração à regressão. Para isso foi efectuada uma validação

cruzada *K-fold*. Este método de validação consiste na retenção, para efeitos de validação, de um reduzido número de registos de presenças aleatórias, sendo o modelo calibrado com os restantes. Os valores obtidos pela regressão são, então, comparados com os valores extraídos para validação. O processo é repetido várias vezes, com sub-amostras distintas, sendo o valor de validação final obtido através da média aritmética simples das diferentes comparações.

Para a criação de cada uma das sub-amostras de validação foram extraídos quatro registos. O quadro III mostra os valores estatísticos simples do erro obtido de cada um dos modelos finais para cada um dos cinco modelos preliminares, que lhes deram origem. Estes valores de erro correspondem à diferença entre o valor de probabilidade modelada, e o valor de presença (1) ou de ausência (0) efectivamente registados.

**Quadro III – Erro médio e estatísticas sumárias dos valores de erro para os três modelos de *habitat* potencial**

	<i>n</i>	E. M.	Desvio-padrão	Mínimo	Máximo	Mediana
<i>Thymus villosus</i>	20	0,26	0,12	0	0,96	0,22
<i>Teucrium chamaedrys</i>	20	0,21	0,09	0	0,78	0,21
<i>Silene longicilia</i>	20	0,15	0,08	0	0,86	0,17

### ***Interpretação dos resultados***

Através da análise aos valores de validação obtidos verifica-se uma maior prestação do modelo representativo do *habitat* potencial da espécie *Silene longicilia*, atingindo uma precisão média de 85% (100 – E.M.x100). O modelo seguinte, em termos de desempenho, corresponde ao realizado para a espécie *Teucrium chamaedrys*, com 79 % e, por fim, o correspondente ao *Thymus villosus* sub. *villosus* com 74%. Os valores de desempenho verificados aparentam estar relacionados com o número de registos utilizados para a calibração das redes neuronais. O modelo da *Silene longicilia*, para o qual foi incluído o maior número de registos de presença-ausência (110), apresentou o melhor desempenho, tendo este diminuído, progressivamente, nos outros dois modelos, de acordo com a redução do número destes registos.

### **Conclusões**

A realização deste trabalho permitiu, sobretudo, identificar uma metodologia válida para a obtenção de *habitats* potenciais. Com resultados de desempenho entre os 74% e os 85% pode-se confirmar, em grande parte, a aplicabilidade das redes neuronais artificiais em trabalhos

desta natureza. É de referir o aumento do desempenho com o respectivo acréscimo de número de pontos amostrados, o que evidencia a elevada importância de levantamentos de campo exaustivos.

Identificou-se que os softwares de SIG, de um modo geral, apresentam ainda muitas limitações no campo da análise estatística multivariada, tendo sido necessário recorrer a programas não direccionados, de forma directa, para a análise de informação espacial. Este factor depende-se como o mais limitante para a realização e uso deste tipo de modelos.

A modelação de muitas das propriedades biofísicas dos ecossistemas é já possível, quer por métodos determinísticos, quer por aproximação estatística. Estes modelos devem, no entanto, atingir elevados valores de precisão de forma a diminuir, o máximo possível, a multiplicação do seu erro. Esta situação foi, possivelmente, um dos principais agentes a contribuir para a diminuição do desempenho dos modelos de *habitat* potencial realizados neste trabalho.

De entre os factores considerados, são vários os que ainda não possuem metodologias de representação espacial muito desenvolvidas. Foi, no entanto, na hidrologia, uma das áreas com maior desenvolvimento em sistemas de informação geográfica, onde se verificou a maior inadequação dos seus resultados, como variável ecológica.

O elevado nível de detalhe utilizado neste trabalho (900 m<sup>2</sup>) também permitiu diminuir a generalização de informação, que pauta muitos dos trabalhos de modelação de *habitats* potenciais, e que dificulta o reconhecimento de áreas com reduzida expressão espacial. Os constantes avanços na tecnologia tornam cada vez mais evidentes as vantagens de se trabalhar a estas escalas de detalhe, outrora condicionadas pela capacidade de processamento dos sistemas.

### **Referências bibliográficas:**

ABDI, H. - Neural Networks. Pp. 1-5. *In*: LEWIS-BECK, M. ; BRYMAN, A.; FUTING, T. - (eds.). *Encyclopaedia of Social Sciences Research Methods*. Thousand Oaks (CA): Sage. 2003.

BUTLER, B. W. Et al. – *WindWizard*: a new tool for fire management decision support [Em linha]. [S.l.]: USDA Forest Service Proceedings, 2006. p. 787-796. [Consult. 26 Fev. 2007]. Disponível em: [http://www.fs.fed.us/rm/pubs/rmrs\\_p041/rmrs\\_p041\\_787\\_796.pdf](http://www.fs.fed.us/rm/pubs/rmrs_p041/rmrs_p041_787_796.pdf).

CONRAD, O. - SAGA: System for Automated Geo-scientific analysis. Saga team, Gottingen. 2003.

COSTA, J. – *Caracterização e constituição do solo*. 7ª ed. Lisboa: Fundação Calouste Gulbenkian, 2004.

FELICÍSIMO, A. M. – *Modelos digitales del terreno* [Em linha]: introducción y aplicaciones en las ciencias ambientales. Oviedo: Pentalfa, 1994. [Consult. 26 Fev. 2007]. Disponível em: <http://www.etsimo.uniovi.es/~feli/pdf/libromdt.pdf>.

FLOR, A. – *Plantas a proteger no Parque Natural das Serras de Aire e Candeeiros*. Lisboa: Instituto da Conservação da Natureza, 2005.

GUISAN, A.; ZIMMERMANN, N.E. – Predictive habitat distribution models in ecology. *Ecological modelling*. Netherlands: Elsevier. N.º 135 (2000), pp. 147-186.

HENGL, T.; GRUBER, S.; SHRESTHA, D. P. – *Digital terrain analysis in ILWIS* [Em linha]. Enschede: International Institute for Geo-Information Science & Earth Observation, 2003, actual. 3 Jul. 2003. [Consult. 26 Fev. 2007]. Disponível em: <http://www.itc.nl/personal/shrestha/DTA>.

McCUNE, B. – *Nonparametric multiplicative regression for habitat modeling* [Em linha]. [Oregon: s.n.], 2006. [Consult. 26 Fev. 2007]. Disponível em: <http://soil.scijournals.org/cgi/content/full/64/6/2046>.

MISÁKOVÁ, L. – Mapping of urban green cover using different image processing techniques on VHR satellite data. In: Proceedings of: High-Resolution Earth Imaging for Geospatial Information, ISPRS Hannover Workshop, 2007.

ODUM, E. P. – *Fundamentos de ecologia*. 7ª ed. Lisboa: Fundação Calouste Gulbenkian, 2004.

OKE, T. R. – *Boundary layer climates*. 2<sup>nd</sup> ed. London: Routledge, 1996.

REIS, E. – *Aplicação dos sistemas de informação geográfica na análise morfológica de bacias hidrográficas*. Lisboa: Universidade Técnica de Lisboa, 1996. Dissertação de Mestrado.

REIS, E. et al. – Integração de dados espaciais em SIG para avaliação da susceptibilidade à ocorrência de deslizamentos. *Finisterra*. Lisboa: Centro de Estudos Geográficos. Vol. XXXVIII, N.º 76 (2003), p. 3-34.

SEGURADO, P.; ARAÚJO, M. – An evaluation of methods for modelling species distributions. *Journal of Biogeography*. [S.l.]: Blackwell Publishing. N.º 31 (2004), p. 1555-1568.

WHITTAKER, R.; ARAÚJO, M. – Conservation biogeography: assessment and prospect. *Diversity and distributions*. [S.l.]: Blackwell Publishing. N.º 11 (2005), p. 3-23.