

Identificação automatizada de morcegos através de modelos estatísticos

Bruno Silva

Universidade de Évora e Plecotus - Estudos Ambientais, *bsilva@plecotus.com*

Gonçalo Jacinto

DMAT/ECT e CIMA-UE da Universidade de Évora, *gjcj@uevora.pt*

Paulo Infante

DMAT/ECT e CIMA-UE da Universidade de Évora, *pinfante@uevora.pt*

Sílvia Barreiro

Plecotus - Estudos Ambientais, *sbarreiro@plecotus.com*

Pedro Alves

Plecotus - Estudos Ambientais, *pjalves@plecotus.com*

Palavras-chave: regressão logística multinomial, redes neuronais, identificação de morcegos, ecolocalização

Abstract: Desenvolvimentos recentes nas metodologias de monitorização de morcegos utilizadas em Portugal, especialmente a introdução de estações de gravação automáticas, conduziram a um problema de análise devido à quantidade de dados obtida. Face a este problema propomos desenvolver um método automatizado de análise e classificação de pulsos de ecolocalização de morcegos, através do desenvolvimento de um programa de computador baseado em modelos estatísticos e utilizando uma base de dados de pulsos de morcegos gravados em Portugal.

Neste artigo apresentamos os resultados obtidos para a análise e classificação das espécies *P. pipistrellus*, *P. pygmaeus* e *M. schreibersii*. Para a análise estatística usámos modelos de regressão logística multinomial e conjuntos de redes neuronais artificiais, tendo sido obtidas taxas globais de classificações corretas superiores a 94%.

1 Introdução

Este trabalho centra-se no estudo e proteção de morcegos, animais de características extraordinárias, mas também bastante vulneráveis. São os únicos mamíferos com capacidade de voo sustentado e muitas das espécies possuem um sistema de ecolocalização bastante avançado. Os seus hábitos predominantemente noturnos associados à ecolocalização, maioritariamente através de pulsos de ultrassom, tornam difícil a sua identificação visual em voo [4]

e consequentemente o seu estudo no estado selvagem. Só após o desenvolvimento de equipamento especializado (detetores de ultrassons), bem como a descoberta de pulsos de ecolocalização específicos nas várias espécies é que o seu estudo se tornou mais fácil e de certa forma possível. No entanto existem algumas dificuldades na identificação de morcegos por pulsos de ecolocalização, já que estes apresentam uma grande variabilidade podendo levar à sobreposição das características dos pulsos em várias espécies [4], o que conduz a uma grande incerteza e subjetividade neste tipo de identificação. Em Portugal existem três espécies, *P. pipistrellus*, *P. pygmaeus* e *M. schreibersii*, cujos parâmetros de ecolocalização se sobrepoem com alguma frequência [8] dificultando a sua correta identificação. Estas espécies apresentam diferentes estatutos de conservação: *P. pipistrellus* e *P. pygmaeus* são considerados "pouco preocupantes", ou seja, as suas populações são amplas e abundantemente distribuídas, enquanto a espécie *M. schreibersii* é considerada "vulnerável", ou seja, apresenta alguns fatores de risco [2]. Devido a estas circunstâncias a sua correta identificação por intermédio de detetores de ultrassom é uma necessidade premente para um melhor conhecimento da abundância e distribuição destas espécies, especialmente no que respeita à espécie *M. schreibersii*.

Muitos trabalhos têm sido apresentados em relação à identificação de morcegos com base em modelos estatísticos. Por exemplo, em [6] compararam os resultados de uma rede neuronal e de análise discriminante (Discriminant Function Analysis, DFA) para identificar 12 espécies de morcegos da Grã-Bretanha enquanto que em [9] compararam conjuntos de redes neurais (Ensembles Neural Networks, ENN), análise discriminante e máquinas de vetor de suporte para a classificação de 14 espécies de morcegos. As gravações utilizadas nestes dois trabalhos são as mesmas e foram obtidas no terreno, sendo que apenas os pulsos de melhor qualidade foram analisados. Mais recentemente, em [10] utilizaram ENN para classificar 34 espécies de morcegos europeus, tendo as gravações para este estudo sido obtidas a partir da base de dados EchoBank.

Tendo em conta a possível variação dos parâmetros de ecolocalização de uma mesma espécie em diferentes áreas geográficas [5], assim como a evidência de que métodos estatísticos podem suplantam operadores humanos em classificação de espécies de morcegos através dos pulsos de ecolocalização, definimos três objetivos para este trabalho. O primeiro objetivo foi a obtenção de gravações de pulsos de morcegos junto à saída de abrigos conhecidos ou gravações após captura e libertação da mão para garantir a correta identificação dos pulsos utilizados, dando início à criação da primeira base de dados de gravações de morcegos obtidas em Portugal. Estas gravações foram efetuadas no terreno, com o ruído e os problemas que daí advém, já que a base de dados obtida não é limpa artificialmente. O segundo objetivo foi o desenvolvimento de um método automático para detetar, isolar e medir os pulsos presentes nas gravações de forma a eliminar o enviesamento humano

e analisar de forma rápida um grande número de gravações. O terceiro objetivo foi a criação de modelos de regressão logística multinomial e de redes neuronais artificiais para classificar os pulsos das três espécies em questão. Este trabalho está organizado da seguinte forma. Na Secção 2 apresentamos sucintamente os modelos estatísticos usados. Na Secção 3 apresentamos os métodos utilizados para obter as gravações e analisar os pulsos obtidos. Na Secção 4 apresentamos os resultados obtidos e terminamos com algumas considerações finais na Secção 5.

2 Metodologia Estatística

Optámos por duas abordagens distintas: modelos de regressão logística multinomial e redes neuronais artificiais. Esta decisão deve-se à complementaridade entre as duas metodologias. A utilização de redes neuronais artificiais permite lidar com grandes conjuntos de dados, mesmo na presença de colinearidade ou de relações não-lineares entre variáveis. Contudo, perante bases de dados pequenas, o que ocorre quando as espécies de morcegos são raras ou difíceis de capturar, não conseguem ajustar-se corretamente. Nestes cenários, os modelos de regressão logística podem ser usadas para prever o resultado da classificação. Além disso, a possibilidade de interpretar os parâmetros obtidos pode abrir novas perspectivas sobre a importância de alguns parâmetros particulares na classificação das espécies de morcegos. A análise estatística foi implementada no software R project, versão 2.15.3 [7].

2.1 Regressão logística multinomial

Para classificar as espécies de morcegos de acordo com os parâmetros obtidos nas gravações dos pulsos, ajustámos um modelo de regressão logística multinomial, cuja variável resposta foi codificada da seguinte forma: 0 – *M. schreibersii*, 1 – *P. pipistrellus* e 2 – *P. pygmaeus*. As duas funções *logit* são então dadas por:

$$g_j(\mathbf{x}) = \ln \left(\frac{P(Y = j|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p, \quad j = 1, 2;$$

onde p representa o número de covariáveis e $\mathbf{x}$ (um vetor linha) é o conjunto de variáveis explicativas associadas às observações, com $x_0 = 1$, e onde $\boldsymbol{\beta}$ representa o vetor dos coeficientes das p covariáveis. A probabilidade condicional de ocorrência de cada espécie de morcego dado o vetor de covariáveis pode escrever-se como $P(Y = i|\mathbf{x}) = \frac{e^{g_i(\mathbf{x})}}{\sum_{k=0}^2 e^{g_k(\mathbf{x})}}$, $i = 0, 1, 2$, com $g_0(\mathbf{x}) = 0$.

Para ajustar o modelo multinomial seguimos a metodologia recomendada por Hosmer e Lemeshow [3], que basicamente usa a mesma metodologia usada nos modelos binomiais: (1) para o modelo inicial foram selecionadas todas as covariáveis que se revelaram significativas na fase univariada (valor

$p < 0,20$); (2) a partir deste modelo foram eliminadas sucessivamente, e por ordem decrescente dos valores p de Wald, todas as covariáveis não significativas (valor $p > 0,10$); (3) verificámos se alguma(s) das covariáveis que não foram incluídas no modelo inicial se mostra(m) agora significativa(s) em presença das que estão no modelo, caso em que são adicionadas ao modelo. Para verificar a linearidade das covariáveis contínuas (que no nosso caso são todas), testar a existência de interações, bem como para a análise de resíduos para pesquisa de observações influentes ou outliers, como sugerido por [3] (pag. 277), usámos o pressuposto da independência das alternativas irrelevantes, onde modelos logísticos binomiais separados foram ajustados e os pressupostos verificados nestes modelos. As inferências finais são, contudo, baseadas nos coeficientes e nos erros padrão estimados do modelo multinomial. A qualidade do ajustamento foi verificada pelo teste de Hosmer e Lemeshow aplicado aos modelos binomiais, e a capacidade discriminativa do modelo foi avaliada pela área sob a curva ROC (Area Under Curve, AUC). Além disso obteve-se o valor de R^2 de MacFadden do modelo multinomial. O modelo final foi validado usando a técnica de validação de bootstrap.

2.2 Redes neurais artificiais

Para este trabalho utilizámos uma classe particular de redes neurais denominada perceptrão multicamadas sem realimentação. Doravante a menção a rede neuronal irá sempre referir-se a esta classe particular. As redes foram criadas com uma camada escondida e com função de ativação logística, quer na camada escondida, quer na camada de saída. A função composta calculada por este tipo de rede pode ser expressa por:

$$\mathbf{o}(\mathbf{x}) = f_0 \left(w_{0k} + \sum_i w_{ik} * f_o \left(w_{oi} + \sum_p x_p w_{pi} \right) \right)$$

onde $\mathbf{o}$ é o vetor de saída (de dimensão k), k o número de neurónios na camada de saída, i o número de neurónios na camada oculta, p o número de covariáveis, f_0 a função de ativação, $\mathbf{x}$ o vetor das covariáveis, w_{pi} os pesos dos neurónios da camada oculta, w_{ik} os pesos dos neurónios da camada de saída, w_{0i} as constantes da camada oculta e w_{0k} as constantes da camada de saída. Estas redes foram treinadas com um algoritmo de treino supervisionado, o GRPROP [1], utilizando a soma dos quadrados dos erros para ajustar os pesos da rede durante o treino. Foram utilizados conjuntos de redes em vez de uma única rede de forma a reduzir a variabilidade das previsões. Foram treinadas 21 redes e utilizado o conjunto das cinco com menor erro para a classificação, sendo o resultado final do conjunto a média dos resultados das cinco redes. A base de dados foi dividida, sendo que 75% das observações foram apresentadas às redes durante o treino e as restantes 25% apenas foram utilizadas para validação após o treino. Os valores de

cada parâmetro foram divididos pelo seu máximo absoluto para mapear os valores de entrada na rede entre 0 e 1 e facilitar a convergência. Para decidir qual o número de neurónios que deveriam ser incluídos na camada escondida foram testados os desempenhos de diversos conjuntos de redes. Foram testados conjuntos com 15, 20, 25 e 30 neurónios e selecionado aquele que apresentou menor erro nas observações de validação.

3 Métodos

3.1 Amostragem

As gravações foram efetuadas em 21 locais distintos no centro e sul de Portugal, maioritariamente à porta de abrigos conhecidos ou após libertação de morcegos capturados. Em ambos os casos houve uma identificação morfológica da espécie, prévia à obtenção das gravações. Apenas no caso da espécie *P. pipistrellus* foram obtidas algumas gravações em pleno voo, sem identificação morfológica prévia, mas apenas em situações em que a sua identificação fosse inequívoca. As gravações visaram apenas a obtenção de pulsos de navegação já que estes são um dos tipos de pulsos mais característicos das espécies de morcegos e são emitidos continuamente durante o voo, sendo por isso bastante indicados para a classificação. As gravações foram efetuadas utilizando detetores Pettersson D240x (Pettersson Elektronik AB) em modo de tempo expandido, acoplados a um gravador digital (Archos Gmini 402). Na Tabela 1 são indicados o número de gravações efetuadas para cada espécie, o número de pulsos de ecolocalização individuais contidos nessas gravações e também o número de localizações distintas, e respetivas condições, onde essas gravações foram obtidas. O número mínimo de gravações foi calculado de forma a que o erro máximo da estimativa da frequência de máxima energia dos pulsos fosse inferior a 5%, com uma confiança de 95%. O número de gravações obtidas para estas espécies superou esse número mínimo.

Tabela 1: Número de gravações e respetivo número de pulsos de cada espécie, número de diferentes locais de obtenção de gravações e respetivas condições de gravação (Lm – libertação da mão; Ea – entrada de abrigo; Vl – voo livre).

Espécie	# gravações	# pulsos	# locais	Condições de gravação
<i>P. pipistrellus</i>	65	313	13	7 Lm, 21 Ea, 37 Vl
<i>P. pygmaeus</i>	118	642	8	22 Lm, 96 Ea
<i>M. schreibersii</i>	150	701	6	116 Lm, 34 Ea

3.2 Análise das gravações

Para analisar e medir as gravações foi desenvolvido um método automático, implementado usando a linguagem R, que permite importar gravações, aplicar um filtro para limpar ruído indesejado, detetar os pulsos de morcego e

isolá-los. Os pulsos isolados são então sujeitos a análise tendo sido medidos 19 parâmetros. O procedimento é completamente automático retirando assim qualquer subjetividade da análise. Houve algumas gravações que estavam demasiado contaminadas com ruído ou em que os pulsos tinham muito pouca intensidade e que tiveram de ser manualmente descartadas, visto que o programa desenvolvido em R não as conseguiu analisar. Nas gravações onde este problema não ocorreu, não foram eliminados quaisquer pulsos medidos, por mais fracos que fossem, de modo a aumentar a variabilidade dos dados de base para a modelação estatística. A medição dos parâmetros incidiu somente nos domínios espectral, através da Transformada rápida de Fourier aplicada a todo o pulso isolado, e temporal. Os parâmetros medidos e a sua descrição podem ser observados na Tabela 2.

Tabela 2: Descrição dos parâmetros analisados.

Parâmetro	Descrição do parâmetro
fmaxe	Frequência de máxima energia
dur	Duração do pulso
fini15	Freq. inicial estandardizada a -15dB da intensidade de fmaxe
fini10	Freq. inicial estandardizada a -10dB da intensidade de fmaxe
fini5	Freq. inicial estandardizada a -5dB da intensidade de fmaxe
ffin15	Freq. final estandardizada a -15dB da intensidade de fmaxe
ffin10	Freq. final estandardizada a -10dB da intensidade de fmaxe
ffin5	Freq. final estandardizada a -5dB da intensidade de fmaxe
bw15	Diferença de frequência entre fini15 e ffin15
bw10	Diferença de frequência entre fini10 e ffin10
bw5	Diferença de frequência entre fini5 e ffin5
posFMAXE	Posição temporal de fmaxe
ptFMAXE	posFMAXE como proporção da duração do pulso
bwffini15	Diferença da frequência fmaxe e fini15
bwffini10	Diferença da frequência fmaxe e fini10
bwffini5	Diferença da frequência fmaxe e fini5
bwffin5	Diferença da frequência fmaxe e ffin5
bwffin10	Diferença da frequência fmaxe e ffin10
bwffin15	Diferença da frequência fmaxe e ffin15

4 Resultados e discussão

Para o modelo logístico multinomial o subconjunto inicial de covariáveis foi composto pelas covariáveis *dur*, *posFMAXE*, *ffin15*, *bw10*, *bwffini15* e *bwffin15*, e foram utilizadas as medianas de todos os pulsos de uma dada gravação. Só se escolheram estas do conjunto mais vasto de 19 covariáveis extraídas, devido a problemas de colinearidade entre elas. Todas estas covariáveis são contínuas. De seguida, procedemos ao método de seleção de variáveis do modelo multinomial referido na Secção 2.

Usando os dois modelos binomiais subjacentes para avaliar a qualidade do ajustamento, verificou-se para o modelo (0 – *M. schreibersii*, 1 – *P. pipistrellus*) uma boa qualidade de ajustamento, com uma AUC de 0,99 e um valor p para o teste de Hosmer e Lemeshow de 0,99 ($\chi^2_{(8)} = 0,86$). O modelo (0 – *M. schreibersii*, 2 – *P. pygmaeus*) também apresenta uma razoável qualidade de ajustamento aos dados, com uma AUC de 0,96 e um valor p do teste de Hosmer e Lemeshow de 0,09 ($\chi^2_{(6)} = 10,8$). Em ambos os modelos, o pressuposto de linearidade foi verificado para todas as covariáveis, não se verificaram interações significativas e a análise de resíduos não detetou a existência de outliers nem de observações influentes. Ambos os modelos foram validados pelo método de bootstrap. O modelo de regressão logística multinomial apresenta um bom ajustamento ao conjunto de dados, com um R^2 de McFadden de 0,75. O modelo final é apresentado na Tabela 3.

Tabela 3: Variáveis significativas do modelo logístico multinomial.

Coefficientes	Estimativas	Erro padrão	valor p	OR ($IC_{95\%}$)
1:(intercept)	111,3	26,8	<0,01	-
2:(intercept)	-22,5	7,2	<0,01	-
1:dur	-1,6	0,7	0,02	0,20 (0,05; 0,75)
2:dur	0,5	0,3	0,03	1,71 (1,06; 3,75)
1:posFMAXE	-3,2	1,1	<0,01	0,04 (0,01; 0,34)
2:posFMAXE	-0,9	0,3	<0,01	0,40 (0,21; 0,74)
1:ffin15	-1,9	0,5	<0,01	0,15 (0,06; 0,37)
2:ffin15	0,5	0,1	<0,01	1,70 (1,31; 2,19)
1:bwffin15	-3,6	1,0	<0,01	0,03 (0,01; 0,19)
2:bwffin15	-3,1	0,6	<0,01	0,04 (0,01; 0,13)

Como um dos objetivos do trabalho é a classificação das espécies de morcegos, apresentamos os resultados da classificação do modelo de regressão multinomial na Tabela 4. Como medidas de avaliação consideramos a sensibilidade, que indica a taxa de acertos na classificação de uma dada espécie, a taxa de erro, que indica a percentagem de classificações de uma dada espécie que na realidade correspondem a outra, e a taxa global de classificações corretas, que indica a percentagem de classificações que correspondem efetivamente à espécie indicada. O modelo classificou corretamente 96,9% das gravações da espécie *P. pipistrellus* errando 6,0% das classificações atribuídas a esta espécie, classificou corretamente 89,0% das gravações de *P. pygmaeus* errando 13,9% das classificações atribuídas a esta espécie e classificou corretamente 89,3% das gravações de *M. schreibersii* errando 6,9% das classificações atribuídas a esta espécie. A taxa global de classificações corretas do modelo é de 90,7%.

Uma das vantagens do modelo logístico multinomial é que permite perceber como as diferentes covariáveis influenciam a probabilidade de classificação das espécies. Na Figura 1 estão representadas as probabilidades de classi-

Tabela 4: Resultados da classificação do modelo logístico multinomial (Msch – *M. schreibersii*; Ppip – *P. pipistrellus*; Ppyg – *P. pygmaeus*); Sens – Sensibilidade; Erro – Taxa de erro). Taxa global de classificações corretas assinalada a negrito na célula inferior direita.

	Observados					
Estimados	Ppip	Ppyg	Msch	Total Est.	Sens (%)	Erro (%)
Ppip	63	4	0	67	96,9	6,0
Ppyg	1	105	16	122	89,0	13,9
Msch	1	9	134	144	89,3	6,9
Total Obs	65	118	150	333	90,7	

figuração nas diferentes espécies, para uma gravação escolhida aleatoriamente da base de dados, fazendo variar cada covariável isoladamente. Pode-se observar, por exemplo, que o aumento da covariável *bwffin15*, irá aumentar a probabilidade de uma gravação pertencer a *M. schreibersii*. Observando os valores das razões das chances na Tabela 3, também verificamos, por exemplo, que a possibilidade de ser classificado como *P. pygmaeus* aumenta 1,7 vezes por cada unidade de aumento da duração do pulso.

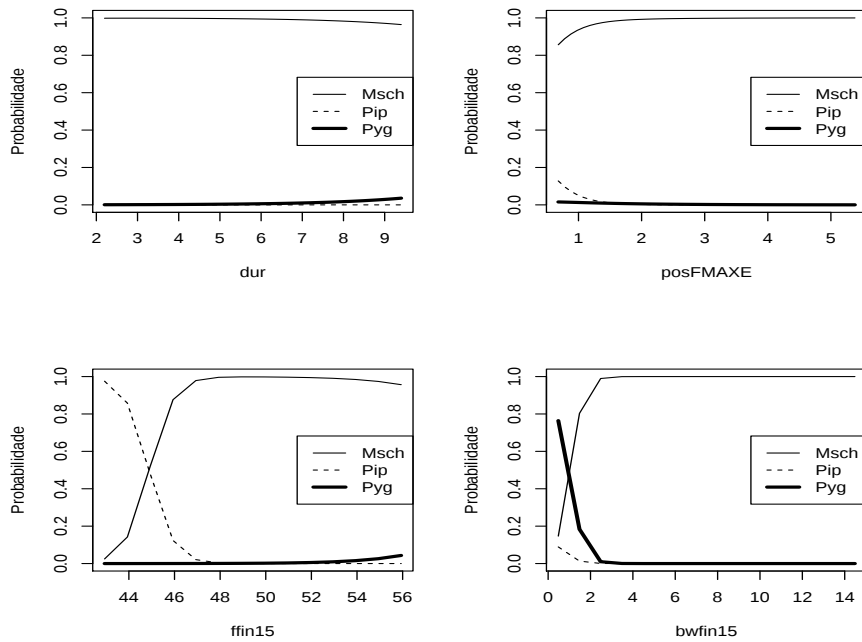


Figura 1: Influência da variação das covariáveis (mantendo as restantes fixas) na probabilidade de classificação das diferentes espécies, para uma gravação com os seguintes valores: *dur* = 3,9 ms, *posFMAXE* = 3,1 ms, *ffin15* = 50,0 kHz e *bwffin15* = 3,0 kHz.

De seguida apresentamos os resultados obtidos através das redes neuronais artificiais. Para a modelação com redes neuronais foram usados todos os

pulsos individuais medidos e todos os 19 parâmetros. Na Tabela 5 apresenta-se a soma dos quadrados dos erros das quatro arquiteturas testadas para a classificação. Observa-se que o conjunto de redes com 15 neurónios na camada escondida é o que apresenta menor erro nas observações de validação.

Tabela 5: Soma de quadrados do erro (SQE) dos dados de validação das várias arquiteturas testadas.

	Número neurónios			
	15	20	25	30
SQE	43,5	45,8	46,9	48,2

Na Tabela 6 são apresentados os resultados das classificações deste conjunto de redes com 15 neurónios na camada escondida. É possível observar que o conjunto de redes classifica corretamente 93,3% dos pulsos da espécie *P. pipistrellus* errando 4,1% das classificações atribuídas a esta espécie, classifica corretamente 91,8% dos pulsos da espécie *P. pygmaeus* errando 6,4% das classificações atribuídas a esta espécie e classifica corretamente 96,2% dos pulsos da espécie *M. schreibersii* errando 6,3% das classificações atribuídas a esta espécie. O conjunto de redes apresenta uma taxa global de classificações corretas de 94,8%. De notar que a espécie *M. schreibersii*, que tem um estatuto de conservação vulnerável, é aquela para a qual o conjunto de redes apresenta maior sensibilidade.

Tabela 6: Resultados da classificação do conjunto de redes com 15 neurónios (Msch – *M. schreibersii*; Ppip – *P. pipistrellus*; Ppyg – *P. pygmaeus*); Sens – Sensibilidade; Erro – Taxa de erro), estando assinalado a negrito a taxa global de classificações corretas.

Estimados	Observados				Sens (%) Erro (%)	
	Ppip	Ppyg	Msch	Total Est.		
Ppip	70	3	0	73	93,3	4,1
Ppyg	3	146	7	156	91,8	6,4
Msch	2	10	179	191	96,2	6,3
Total Obs	75	159	186	478	94,8	

5 Conclusão

Neste trabalho foram atingidos três objetivos principais: as fundações para a primeira base de dados de gravações de pulsos de ecolocalização de morcegos obtidas em Portugal, o desenvolvimento de um programa em R para automaticamente isolar e medir pulsos de morcego de uma gravação e a obtenção de modelos estatísticos para classificar esses pulsos através dos parâmetros medidos. Foram utilizadas duas abordagens estatísticas distintas: modelos de regressão logística multinomial e redes neuronais artificiais.

Os modelos de redes neuronais apresentaram taxas de classificações corretas bastante interessantes e superiores às obtidos com a regressão logística,

embora os resultados não sejam diretamente comparáveis uma vez que as variáveis utilizadas em ambos os modelos não são as mesmas. Contudo, nos modelos de regressão logística, a informação adicional acerca dos parâmetros relevantes, associadas a um método simples e rápido para calcular a probabilidade de uma dada gravação pertencer a uma determinada espécie, poderá trazer um contributo importante ao estudo de morcegos.

O estudo apresentado neste trabalho é parte de um projeto de investigação em curso, estando a captura/gravação de morcegos ainda a decorrer. Pretendemos estender esta análise às 25 espécies de morcegos existentes em Portugal continental e desenvolver um programa de computador que integre a análise e a classificação estatística automática de espécies de morcegos.

Agradecimentos

Ao Bat Conservation International, Inc (BCI) por ter financiado este trabalho através de uma bolsa de investigação, e à PLECOTUS – Estudos Ambientais Unip, Lda por ter fornecido todo o material técnico.

Referências

- [1] Anastasiadis, A., Magoulas, G., Vrahatis, M. (2005). New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing*, 64, 253–270.
- [2] Cabral, M., *et al.* (2006). *Livro vermelho dos vertebrados de Portugal*, 2ª edição. Lisboa, Instituto de Conservação da Natureza e Assírio & Alvim.
- [3] Hosmer, D., Lemeshow, S. (2000). *Applied Logistic Regression*. New York, Wiley.
- [4] Jones, G., Vaugh, N., Parsons, S. (1999). Acoustic identification of bats from directly sampled vocalizations. *Acta chiropterologica*, 2(2), 155-170.
- [5] Parsons, S. (1997). Search-phase echolocation calls of the New Zealand lesser short-tailed bat (*Mystacina tuberculata*) and long-tailed bat (*Chalinolobus tuberculatus*). *Canadian Journal of Zoology*, 9(75), 1487-1494.
- [6] Parsons, S., Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *The Journal of experimental biology*, 203(17), 2641-56.
- [7] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- [8] Russo, D., Jones, G. (2002). Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *Journal of Zoology*, 1(258), 91-103.
- [9] Redgwell, R., Szewczak, J., Jones, G., Parsons, S. (2009). Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. *Algorithms*, 2(3), 907-924.
- [10] Walters, C., *et al.* (2012). A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49(5), 1064-1074.